Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

Dr. Anja Lisa Riediger

born in: Rastatt, Germany

Oral examination: 30th March 2026

Complementary analysis of genomic and epigenomic features
in plasma and urinary cell-free DNA
for risk stratification in prostate cancer at first diagnosis

Referees: Prof. Dr. Oliver Stegle
Prof. Dr. Holger Sültmann

Summary

Prostate cancer (PCa) is a biologically and clinically heterogeneous disease. Current diagnostic strategies, including prostate-specific antigen (PSA) testing, imaging-based clinical staging, and histopathological assessment, are limited in their ability to fully capture this heterogeneity and to reliably support individualized risk stratification. Consequently, there is a critical need for additional molecular biomarkers and integrative diagnostic approaches that enable personalized, patient-oriented treatment decisions. Liquid biopsy analyses allow the minimally invasive assessment of tumor-derived molecules in body fluids, providing a comprehensive view of the entire tumor burden and its heterogeneity. The analysis of cell-free DNA (cfDNA), and in particular circulating tumor DNA (ctDNA), enables genomic and epigenomic tumor profiling. However, ctDNA detection in PCa is challenging due to low ctDNA shedding and a low genomic alteration burden, especially in localized disease. The aim of this PhD project was to establish a multimodal liquid biopsy framework that integrates genomic and epigenomic cfDNA analyses from matched plasma and urine samples of PCa patients at initial diagnosis. The goal was to improve ctDNA detection and to achieve a comprehensive molecular characterization of heterogeneous PCa, with a particular focus on identifying localized disease and stratifying aggressive tumors. To this end, I applied low-coverage whole-genome sequencing and cell-free methylated DNA immunoprecipitation sequencing to plasma and urinary cfDNA to assess chromosomal instability, copy number variations, cfDNA fragmentation patterns, and aberrant DNA methylation. I performed these analyses on 109 plasma samples and 102 urine samples obtained from 73 newly diagnosed PCa patients and 36 cancer-free controls. Most PCa patients presented with localized disease, while approximately one quarter had lymph node or distant metastases. In addition, I analyzed eight fresh-frozen PCa tissue samples with matched buffy coat DNA to enable comparisons with liquid biopsy data and to identify tumor-informative characteristics. Furthermore, I compared results from genomic and epigenomic cfDNA profiling between plasma and urine to identify shared and complementary characteristics across both biofluids, and between PCa patients and controls to distinguish tumor-specific patterns. I subsequently selected representative cfDNA features for ctDNA detection, with detection thresholds defined based on control samples. Finally, I assessed ctDNA positivity in relation to clinical and pathological characteristics. Fragmentation analyses revealed distinct and tumor-informative cfDNA profiles in plasma and urine, with biofluid-specific fragmentation features distinguishing PCa patients from controls. Genomic profiling identified copy number variations and increased chromosomal instability in both biofluids, providing complementary evidence of ctDNA presence. Samples with higher proportions of detectable ctDNA harbored recurrent genomic alterations consistent with those observed in PCa tissue samples and with known alterations in primary PCa reported in the literature. Epigenomic analyses identified differentially methylated regions in both plasma and urinary cfDNA that distinguished metastatic PCa patients from PCa patients without distant metastases and from cancer-free controls. Furthermore, plasma and urinary cfDNA from PCa patients showed increased methylation levels in PCa tissue-derived methylation regions, which were validated using an external PCa tissue dataset, further supporting ctDNA detection. Overall, genomic and epigenomic differences were most pronounced in advanced PCa, while cfDNA features related to DNA methylation and chromosomal instability also enabled discrimination between advanced and localized PCa. The integrated assessment of cfDNA fragmentation, copy number variations, chromosomal instability, and methylation in plasma and urine substantially increased ctDNA detection rates compared to single-parameter or single-biofluid analyses. CtDNA was detected in 45% of newly diagnosed PCa patients, including 42% of localized cases and 56% of advanced cases. Importantly, ctDNA was detectable in a considerable proportion of patients with low to intermediate PSA levels (< 10 ng/ml), a clinical scenario in which risk stratification remains challenging. These findings underscore the potential of cfDNA-based genomic

and epigenomic markers to address an unmet clinical need and to complement established diagnostics for refined risk stratification at initial diagnosis. In conclusion, this proof-of-concept study introduces the first reported multimodal liquid biopsy framework integrating genomic and epigenomic cfDNA analyses from both plasma and urine in newly diagnosed PCa patients. The results demonstrate that combining multiple cfDNA features across two biofluids enables a more comprehensive molecular characterization of PCa and improves ctDNA detection, even in localized disease. Despite these promising results, several limitations must be acknowledged. CtDNA detection rates, although improved, remained moderate, reflecting both biological constraints and current technical limitations. The cohort size was moderate, with limited representation of advanced and metastatic disease. Further studies with larger, independent cohorts, as well as prospective clinical trials, will be required to validate multimodal liquid biopsy approaches and to evaluate their clinical applicability. Future perspectives include the integration of multimodal liquid biopsy data with additional diagnostic modalities, supported by machine learning-based approaches for data integration, to achieve comprehensive, multimodal representations of PCa biology and ultimately support improved risk stratification and personalized treatment decisions.

Zusammenfassung

Das Prostatakarzinom ist eine biologisch und klinisch heterogene Erkrankung. Etablierte diagnostische Verfahren, darunter die Bestimmung des prostataspezifischen Antigens (PSA), die bildgebungsbasierte klinische Stadieneinteilung sowie die histopathologische Beurteilung von Gewebe, sind nur eingeschränkt geeignet, diese Heterogenität vollständig abzubilden und darauf aufbauend eine individualisierte Risikostratifizierung zu ermöglichen. Daraus ergibt sich ein klinischer Bedarf an zusätzlichen molekularen Biomarkern sowie integrativen diagnostischen Ansätzen, die eine personalisierte und patientenorientierte klinische Entscheidungsfindung unterstützen. Liquid Biopsy Analysen ermöglichen die minimalinvasive Untersuchung tumorassoziierter Moleküle in Körperflüssigkeiten und erlauben somit eine umfassende Beurteilung der Tumorerkrankung und ihrer Heterogenität. Insbesondere die Analyse zellfreier DNA (cfDNA) und der darin enthaltenen zirkulierenden Tumor-DNA eröffnet die Möglichkeit einer genomischen und epigenomischen Tumorcharakterisierung. Die Detektion von zirkulierender Tumor-DNA (ctDNA) ist beim Prostatakarzinom jedoch aufgrund der geringen DNA-Freisetzung und der vergleichsweise niedrigen Mutationslast, insbesondere bei lokalisierten Tumoren, technisch anspruchsvoll. Ziel dieser Promotionsarbeit war die Etablierung eines multimodalen Liquid Biopsy Ansatzes, der genomische und epigenomische Analysen zellfreier DNA aus gepaarten Plasma- und Urinproben von Prostatakarzinom-Patienten zum Zeitpunkt der Erstdiagnose integriert. Hierbei wurde angestrebt, die Detektion zirkulierender Tumor-DNA zu verbessern und eine umfassende molekulare Tumorcharakterisierung zu ermöglichen, insbesondere zur Identifikation lokalisierter Tumoren sowie zur Stratifikation aggressiver Krankheitsverläufe. Hierzu wandte ich eine Ganzgenomsequenzierung mit niedriger Sequenziertiefe sowie die Sequenzierung angereicherter methylierter DNA mittels Immunpräzipitation auf zellfreier DNA von Plasma und Urin an, um chromosomale Instabilität, Kopienzahlveränderungen, Fragmentierungsmuster zellfreier DNA sowie aberrante DNA-Methylierung zu untersuchen. Die Analysen umfassten 109 Plasma- und 102 Urinproben von insgesamt 73 Prostatakarzinom-Patienten zum Zeitpunkt der Erstdiagnose sowie 36 krebsfreien Kontrollpersonen. Der überwiegende Teil der Patienten wies eine lokalisierte Erkrankung auf, während etwa ein Viertel der Patienten lymphogene Metastasierung oder Fernmetastasen aufwies. Zusätzlich analysierte ich acht Prostatakarzinom-Gewebeproben mit jeweils korrespondierender Buffy Coat-DNA, um tumorassoziierte Merkmale zu identifizieren und einen Vergleich mit den Ergebnissen der Liquid Biopsy Analyse zu ermöglichen. Vergleichende Analysen von Plasma- und Urinproben dienten der Charakterisierung gemeinsamer und komplementärer Merkmale der zellfreien DNA in beiden Körperflüssigkeiten sowie der Identifikation tumorspezifischer Muster zur Abgrenzung von Krebspatienten gegenüber Kontrollpersonen. Ich wählte repräsentative Merkmale für die Detektion von Tumor-DNA aus und definierte dabei die Detektionsschwellen auf Basis der Kontrollproben. Im Anschluss setzte ich die Ergebnisse in Bezug zu klinischen und pathologischen Parametern. Die Analyse der Fragmentlängen zirkulierender DNA in Plasma und Urin zeigte ausgeprägte und tumorinformative Fragmentierungsprofile, die sowohl spezifische Merkmale der jeweiligen Körperflüssigkeiten als auch Unterschiede zwischen Tumorpatienten und Kontrollen widerspiegelten. Die genomische Analyse identifizierte Kopienzahlveränderungen sowie eine erhöhte chromosomale Instabilität in der zellfreien DNA aus beiden Körperflüssigkeiten und lieferte damit komplementäre Hinweise auf das Vorhandensein von zirkulierender Tumor-DNA. Liquid Biopsy Proben mit nachweisbarer Tumor-DNA wiesen rekurrente genomische Alterationen auf, die mit den Ergebnissen aus der Gewebeanalyse sowie mit bekannten Veränderungen im primären Prostatakarzinom aus der Literatur übereinstimmten. Die epigenomische Analyse identifizierte differentiell methylierte Regionen in zellfreier DNA aus Plasma und Urin, die metastasierte Prostatakarzinom-Patienten von Patienten ohne Fernmetastasen sowie von Kontrollpersonen

unterschieden. Darüber hinaus zeigten Plasma- und Urinproben von Prostatakarzinom-Patienten erhöhte Methylierungsniveaus in gewebebasierten, Prostatakarzinom-spezifischen Methylierungsregionen, die mithilfe eines externen Prostatakarzinom-Gewebedatensatzes validiert wurden, was die Detektion von zirkulierender Tumor-DNA zusätzlich unterstützte. Insgesamt waren genomische und epigenomische Unterschiede insbesondere bei fortgeschrittenem Prostatakarzinom am ausgeprägtesten. Darüber hinaus ermöglichten Merkmale der DNA-Methylierung und der chromosomalen Instabilität eine Unterscheidung zwischen lokalisierten und fortgeschrittenen Erkrankungsstadien. Die integrierte Analyse von Fragmentierung, Kopienzahlveränderungen, chromosomaler Instabilität und DNA-Methylierung der zellfreien DNA führte im Vergleich zur Analyse von einzelnen Parametern oder Körperflüssigkeiten zu einer deutlich erhöhten Detektionsrate zirkulierender Tumor-DNA in Plasma und Urin der Prostatakarzinom-Patienten. Zirkulierende Tumor-DNA konnte bei 45% der Gesamtkohorte, darunter bei 42% der Patienten mit lokalisierter Erkrankung und bei 56% der Patienten mit fortgeschrittener Erkrankung, nachgewiesen werden. Von besonderer klinischer Relevanz ist, dass unter den Patienten mit nachweisbarem Tumorsignal auch ein beträchtlicher Anteil mit niedrigen bis intermediären PSA-Werten (< 10 ng/ml) war, bei denen die Risikostratifizierung und klinische Entscheidungsfindung besonders herausfordernd sind. Diese Ergebnisse unterstreichen die klinische Notwendigkeit und das Potenzial multimodaler Liquid Biopsy Analysen als Ergänzung zu etablierten diagnostischen Verfahren und zur Optimierung der Risikostratifizierung zum Zeitpunkt der Erstdiagnose. Zusammenfassend stellt diese Proof-of-Concept-Studie den ersten beschriebenen multimodalen Liquid Biopsy Ansatz dar, der genomische und epigenomische Analysen zellfreier DNA aus Plasma und Urin bei neu diagnostizierten Prostatakarzinom-Patienten integriert. Die Ergebnisse zeigen, dass die Kombination mehrerer Merkmale in zwei Körperflüssigkeiten eine umfassendere molekulare Tumorcharakterisierung ermöglicht und die Detektion von Tumor-DNA verbessert. Trotz dieser vielversprechenden Ergebnisse sind mehrere Limitationen zu berücksichtigen. Die Detektionsraten der zirkulierenden Tumor-DNA blieben trotz der erzielten Verbesserungen insgesamt moderat, was sowohl biologische Faktoren als auch derzeitige technische Limitationen widerspiegelt. Zudem war die Kohortengröße begrenzt, insbesondere im Hinblick auf fortgeschrittene und metastasierte Erkrankungsstadien. Zukünftige Studien mit größeren, unabhängigen Kohorten sowie prospektive klinische Studien sind erstrebenswert, um multimodale Liquid Biopsy Ansätze zu validieren und deren klinische Anwendbarkeit zu evaluieren. Perspektivisch stellt die Integration von Liquid Biopsy Daten mit weiteren diagnostischen Modalitäten, unterstützt durch den Einsatz künstlicher Intelligenz zur Datenintegration, einen vielversprechenden Ansatz dar, um umfassendere Darstellungen der Biologie des Prostatakarzinoms zu generieren und letzlich eine verbesserte Risikostratifizierung sowie personalisierte Therapieentscheidungen zu ermöglichen.

Table of Contents

List of Tables

List of Figures

List of Appendix Tables

List of Appendix Figures

Abbreviations

5C	unmethylated
5mC	methylated
ADCY1	*adenylate cyclase 1*
ADT	androgen deprivation therapy
AMP	adenosine monophosphate
APC	*APC regulator of Wnt signaling pathway*
AR	androgen receptor
ARID1A	*AT-rich interaction domain 1A*
ARPI	androgen receptor pathway inhibitors
ASAP	atypical small acinar proliferation
ATM	*ATM serine/threonine kinase*
ATP	adenosine triphosphate
AUA	American Urological Association
AUC	area under the curve
BAM	binary alignment map
bmi	body mass index
bp	base pair
BPH	benign prostate hyperplasia
BR	Broad Range
BRCA1/BRCA2	*BRCA1/BRCA2 DNA repair associated*
CAPP-seq	CAncer Personalized Profiling by deep Sequencing
CCND1/CCND2	*cyclin D1/D2*
CDKN1B	*cyclin dependent kinase inhibitor 1B*
cfDNA	cell-free DNA
cfMeDIP-seq	cell-free methylated DNA immunoprecipitation sequencing
CGI	CpG island
CHD1	*chromodomain helicase DNA binding protein 1*
CHEK2	*checkpoint kinase 2*
CHRM1	*cholinergic receptor muscarinic 1*
CI	confidence interval
CIA	chromosomal instability analysis
CNI	copy number instability
CNS	central nervous system
CNV	copy number variation
CPC-GENE	Canadian Prostate Cancer Genome Network
CPVL	*carboxypeptidase vitellogenic like*
CRIP3	*cysteine rich protein 3*
CRPC	castration-resistant prostate cancer
CSPC	hormone-sensitive prostate cancer
csPCa	clinically significant prostate cancer
CT	computed tomography
Ct value	cycle threshold value
CTC	circulating tumor cell
ctDNA	circulating tumor-derived DNA
D statistic	distance statistic
DDR	DNA damage repair
DKFZ	German Cancer Research Center
DLX1	*distal-less homeobox 1*
DMR	differentially methylated region
DNMT	DNA methyltransferase
DOK5	*docking protein 5*
DPBS	Dulbecco's phosphate-buffered saline
dPCR	digital polymerase chain reaction

DRE	digital rectal examination
DST	*dystonin*
EAU	European Association of Urology
EDTA	ethylenediaminetetraacetic acid
EGA	European Genome-Phenom Archive
EPCAM	*epithelial cell adhesion molecule*
ERG	*ETS transcription factor ERG*
ERSPC	European randomized study of screening for prostate cancer
ETS	*E26 transformation-specific*
ETV1/ETV4/ETV5	*ETS variant transcription factor 1/4/5*
FANCD2	*FA complementation group D2*
FDA	Food and Drug Administration
FGFR1	*fibroblast growth factor receptor 1*
FLI1	*Fli-1 proto-oncogene, ETS transcription factor*
FOXA1	*forkhead box A1*
gDNA	genomic DNA
GLM	generalized linear model
GnRH	Gonadotropin-releasing hormone
GSTP1	*glutathione S-transferase pi 1*
HMM	hidden Markov model
HOXB13	*homeobox B13*
HOXC4/HOXC6	*homeobox C4/C6*
HR	homologous recombination
HS	High Sensitivity
IDH1	*isocitrate dehydrogenase (NADP(+)) 1*
IGFBP3/IGFBP7	*insulin like growth factor binding protein 3/7*
INVAR	high-sensitivity integration of variant reads
INVAR-TAPAS	INtegration of VAriant Reads – TAilored PAnel Sequencing
ISUP	International Society of Urological Pathology
kb	kilobase
KLK2	human kallikrein 2
lcWGS	low-coverage whole-genome sequencing
LINE	long interspersed nuclear elements
LNCaP	Lymph Node Carcinoma of the Prostate
logFC	log fold change
logR	log ratio
LTR	long terminal repeats
M	million
MAF	mutant allele frequency
MALBAC-NGS	multiple annealing and looping-based amplification cycles-next-generation sequencing
MAP3K1	*mitogen-activated protein kinase kinase kinase 1*
MAPQ	mapping quality score
MBD(-seq)	methyl-CpG binding domain (sequencing)
mCRPC	metastatic castration-resistant prostate cancer
MeDIP	methylated DNA immunoprecipitation
MIC-1	macrophage inhibitory cytokine-1
miRNA	microRNA
MLH1	*mutL homolog 1*
MMR	mismatch repair
(mp)MRI	(multiparametric) magnet resonance imaging
mRNA	messenger ribonucleic acid
MSMB	microseminoprotein-β
MSH2/MSH6	*mutS homolog 2/mutS homolog 6*
MYC	*MYC proto-oncogene, bHLH transcription factor*
NCOA2	*nuclear receptor coactivator 2*

NCT	National Center for Tumor Diseases
NEPC	neuroendocrine PCa
NET	neutrophil extracellular trap
NGS	Next Generation Sequencing
NKX3-1	*NK3 homeobox 1*
NMIBC	nonmuscle-invasive bladder cancer
NODAL	*nodal growth differentiation factor*
NRPKM	normalized reads per kilobase million mapped reads
NSCLC	non-small cell lung cancer
P163–169 bp	proportion of fragments with 163–169 bp length
PCA	principal component analysis
PCa	prostate cancer
PCA3	*prostate cancer associated 3*
PCR	polymerase chain reaction
PDE4D	*phosphodiesterase 4D*
PET	positron emission tomography
PIK3CA	*phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha*
PIK3R1	*phosphoinositide-3-kinase regulatory subunit 1*
PIN	prostatic intraepithelial neoplasia
PI-RADS	Prostate Imaging-Reporting and Data System
PITX2	*paired like homeodomain 2*
PoN	Panel of Normals
PPV	positive predictive value
PR-AUC	precision-recall area under the curve
PROM1	*prominin 1*
PSA	prostate specific antigen
PSMA	prostate-specific membrane antigen
PSP94	prostate secretory protein 94
PTEN	*phosphatase and tensin homolog*
PTGS2	*prostaglandin-endoperoxide synthase 2*
PTPRN2	*protein tyrosine phosphatase receptor type N2*
QC	quality control
qPCR	quantitative polymerase chain reaction
RARB	*retinoic acid receptor beta*
RASSF1	*Ras association domain family member 1*
RB1RB	*transcriptional corepressor 1*
Reprimo (RPRM)	*reprimo, TP53 dependent G2 arrest mediator homolog*
RGMB	*repulsive guidance molecule BMP co-receptor b*
ROC	receiver operating characteristic
RPM	values reads per million values
SAM	sequence alignment map
SD	standard deviation
SFRP2	*secreted frizzled related protein 2*
SKIDA1	*SKI/DACH domain containing 1*
SLC2A2	*solute carrier family 2 member 2*
SNV	single nucleotide variant
SOX2	*SRY-box transcription factor 2*
SOX2-OT	*SOX2 overlapping transcript*
SPOP(L)	*speckle type BTB/POZ protein (like)*
TAC1	*tachykinin precursor 1*
TERT	*telomerase reverse transcriptase*
TFx	tumor fraction
TIG1 (GPRC5A)	*G protein-coupled receptor class C group 5 member A*
TMEM106A	*transmembrane protein 106A*
TMM	values trimmed mean of M values

TMPRSS2/TMPRSS4	*transmembrane serine protease 2/4*
TNM	UICC Tumour, Node, Metastasis (classification)
TP53	*tumor protein p53*
TUBA4B	*tubulin alpha 4b*
UCSC	University of California Santa Cruz
UDI	unique dual index
UICC	Union for International Cancer Control
UMI	unique molecular identifier
UTR	untranslated region
WES	whole-exome sequencing
WGBS	whole-genome bisulfite sequencing
WGS	whole-genome sequencing
WHO	World Health Organization
ZBTB46	*zinc finger and BTB domain containing 46*
ZFP64	*ZFP64 zinc finger protein*
ZSCAN12	*zinc finger and SCAN domain containing 12*

Note: Gene nomenclature in this PhD thesis follows the official designations provided by the National Center for Biotechnology Information (NCBI) Gene database.

1 Introduction

1.1 Prostate cancer

1.1.1 Epidemiology and tumorigenesis

Prostate cancer (PCa) is the second most common tumor entity in men worldwide with 1,467,854 cases in 2022, accounting for 14.2% of all cancer cases after lung cancer (15.2%) and before colorectal cancer (10.4%)[1]. Thereby, the PCa incidence varies highly between the different countries: PCa is the most frequently diagnosed cancer entity in 118 from 185 countries worldwide, such as the majority of European countries, North and South America, Australia, and many countries in Africa[1]. With regard to cancer-related mortality, PCa ranks 8th in overall cancer-related death (5th rank only in men), with 397,430 PCa-related deaths (4.1%)[1]. PCa ranks as leading cause of cancer-related death in 52 countries, particularly in parts of Central and South America, Africa, and Sweden as only European country[1]. The incidence rates of PCa have shown varying trends over the past centuries, with a notable association observed between these trends and the implementation of PCa screening[2]. In North America and Australia, PCa incidences increased in the late 1980s and early 1990s as consequence of prostate specific antigen (PSA) testing, but declined until 2014 due to a depletion of prevalent cancers and reduced PSA testing with consequently reduced diagnosis of localized PCa disease[1-4]. The decline in PSA testing was linked to updated recommendations from the United States Preventive Services Task Force, initially advising against screening for men over 75 years, and later extending this guidance to all men[5,6]. From 2014 until 2019, a new increase in PCa incidence was reported for the USA, whereas PCa-related mortality stabilized after previous decline, partly because of increased rates in advanced-stage PCa[1,4]. Similar associations between PCa screening and incidences were visible in European countries. Moderate increases in the PCa incidence were observed in countries in which PSA testing was introduced more slowly, such as the United Kingdom or the Netherlands, whereas other countries with structured PCa screening programs experienced rapid increases in PCa incidence, for example Lithuania[2,7]. In Germany and several other Western European countries, incidence increased from the mid-to-late 1990s to the mid-2010s, followed by subsequent declining trend. In some Northern European countries, the decline was less pronounced or an upward trend was still observed[8,9].

The developments of mortality rates were also associated to PCa screening and available therapy options. Declining mortality rates were observed in high-income countries, such as in North America and Northern or Western Europe, where effective treatment and earlier detection through testing were achieved[1,8]. In contrast, mortality rates increased in other regions with less advanced resources, including Central or Eastern Europe, Asia, and Africa. In Germany, 65,820 new PCa cases and 15,403 PCa-related deaths were reported in 2020[9]. In 2023, mortality declined to 15,196 PCa-related deaths, representing 6.6% of all cancer-related deaths after lung, pancreas, and breast cancer[10]. Among cases with known stages at initial PCa diagnosis, the majority were identified at localized stages (Union for International Cancer Control (UICC) stage I-III), while only one fifth of cases harbored UICC stage IV[9] (Table 1), including large, localized tumors (T4 N0 M0) and tumors with lymph node or distant metastases (T1–4 N1 M0 or N0–1 M1)[11].

Table 1: Distribution of UICC stages at initial PCa diagnosis and corresponding 5-year relative survival.
Distribution of UICC stages at initial PCa diagnosis for cases with known stage (61%; 39% of cases with missing data; 4% of cases diagnosed by Death-Certificate-Only). Data based on German Cancer Statistics, 2019-2020[9].

UICC stage	distribution of PCa cases (%)	5-year relative survival (%)
1	40%	99%
2	28%	101%
3	15%	100%
4	18%	54%

Abbreviations: PCa = prostate cancer, UICC = Union for International Cancer Control

There are several known risk factors for PCa tumorigenesis, of which advanced age, family history and genetic predisposition are the most substantial ones[1,12,13]. The life time risk for developing PCa is approximately 14%, affecting one out of seven men[9]. PCa occurs rarely in younger men, and early-onset PCa in men younger than 56 years is only seen in approximately 10% of cases[14]. The PCa risk for men at the age of 45 and 55 is 0.4% and 2.5%, respectively, and increases to 6.2% at the age of 65 and further to 6.7% at the age of 75[9]. PCa belongs to the cancer entities with the highest hereditary influence, and up to 60% of PCa risk have been related to inherited factors[15,16]. PCa has been observed to occur within family clusters, similar to what is known for breast and colon cancers[17]. The relative PCa risk for a man is 3.14, if his brother was diagnosed with PCa, and increases to 4.39, if two or more first-degree relatives were affected. Gene alterations in *BRCA1 DNA repair associated* (*BRCA1)* and *BRCA2 DNA repair associated* (*BRCA2)* are known hereditary changes that increase PCa risk, and *BRCA2* mutations are associated with a worse prognosis in PCa[17,18]. Alterations in the *homeobox B13 (HOXB13)* gene are equally associated to hereditary PCa[19]. The Lynch Syndrome has been linked to hereditary non-polyposis colorectal cancer, but also increases the risk for PCa, as well as for other gynecological and gastrointestinal tumor entities[19]. In this context, germline mutations in DNA mismatch repair genes (*mutL homolog 1 (MLH1), mutS homolog 2 (MSH2), mutS homolog 6 (MSH6),* and *epithelial cell adhesion molecule (EPCAM)*) were observed, leading to microsatellite instability. Overall, hereditary PCa is associated with rare mutations or combinations of genetic variants that can be captured using polygenic risk scores[17,20]. Besides age and genetic predisposition, ethnicity has been proposed as an additional PCa risk factor. PCa occurs more frequently in African American men or in Caribbean men of African ancestry, and these men tend to be younger at initial diagnosis[1]. Other potential PCa risk factors include smoking, high-proof alcohol, excessive body weight/obesity, taller height and nutritional aspects[21-25]. Additionally, inflammatory processes of the prostate gland (prostatitis), urinary retention, vasectomy, and aspects of sexual behavior have been suggested to influence PCa risk[26-30].

1.1.2 PCa screening and diagnostics

The current diagnostic workflow for PCa includes clinical assessment, such as the digital rectal examination (DRE), PSA testing, ultrasound and the performance of a (multiparametric) magnet resonance imaging ((mp)MRI) of the prostate, and finally the prostate tissue biopsy with histopathological assessment, which is the current gold standard for the PCa diagnosis (Figure 1). Initiation of this diagnostic workflow is recommended based on certain indications, such as the presence of PCa risk factors, clinical symptoms, suspicious palpatory finding in the DRE, and increased or dynamic PSA levels within population-based or self-initiated screening[31]. Clinical symptoms are rare in localized PCa, but might occur with local expansion of the tumor[31]. These symptoms include a variety of lower urinary tract symptoms, pain, hematuria and erectile dysfunction, several of which can also occur with benign conditions, such as prostate hyperplasia[31,32]. In Germany, a nation-wide PCa screening program from the health insurance fund offers an annual DRE and a counselling interview for men at the age of 45 or older. PSA testing is not included in the program and can be performed as self-payer service, unless there is a suspicion of PCa that requires further evaluation[33]. DRE has previously been shown to harbor a positive predictive value (PPV) of 5–30% in localized disease, even with low PSA levels < 4 ng/ml[34]. Evaluation of the European randomized study of screening for prostate cancer (ERSPC) revealed that a suspicious DRE, in combination with elevated PSA levels, resulted in a PPV of 48.6% for the detection of PCa at initial screening, instead of 22.4% based on a non-suspicious DRE[35]. The combination of increased PSA levels and suspicious DRE findings was shown to double the likelihood of being diagnosed with significant PCa, compared to abnormal PSA results alone[36]. However, the hazard ratio for abnormal PSA levels were higher compared to suspicious DRE findings, and PSA testing alone achieved higher sensitivities and specificities in PCa screening compared to DRE[37,38]. Recent investigations in the German PROBASE study on PCa screening equally revealed that DRE alone was not effective for early detection of PCa in young men[39], and its usage as single diagnostic tool has not been recommended for an early PCa screening[40]. Following these results, the revised German PCa guidelines have not recommended the application of DRE for PCa screening anymore[41]. The implementation of population-based PCa screening varies distinctly between different countries and is controversially discussed. Large Cochrane reviews with data from five randomized controlled trials revealed that PCa screening significantly increased the overall diagnosis of PCa, localized PCa was more commonly diagnosed with screening, and the proportion of advanced PCa was significantly lower in screening cohorts[42,43]. No significant decrease in PCa-specific mortality was observed in the summary of all five studies, and no benefit on all-cause mortality was detected[42,43]. Only the ERSPC revealed a significant reduction of the PCa-related death in the long-term, and also reported a significantly reduced risk for the development of metastatic disease[44-46]. Besides the reduction of advanced or metastatic disease and cancer-related mortality, early detection of PCa could offer individual benefits for the patients, such as higher possibility of cure, less aggressive treatment options, and improved quality of life[47]. However, the chances associated with PCa screening have to be balanced with potentially harms which are particularly associated with the limited specificity of the PSA testing[47]. The PSA is a prostate-specific serine proteinase produced and secreted by epithelial cells of the prostate[48]. The majority of PSA in serum is bound to alpha-1antichymotrypsin and the remaining PSA circulates unbound as free PSA[48-50]. Although PSA is prostate-specific, it is not PCa-specific and can be also increased by physiological factors (e.g., physical exercise) or due to benign conditions (benign prostate hyperplasia (BPH), prostatitis) [48,51,52]. Increased PSA levels have been associated with a greater likelihood of PCa[53,54], while PCa can also occur at low PSA levels[55]. Limited specificity of PCa screening can lead to false-positive findings, resulting in unnecessary biopsies and associated side-effects as well as psychological burden[42,56]. In addition, multiple studies have shown that screening is associated with overdiagnosis of indolent or low-risk PCa that would otherwise not lead to clinical symptoms or cancer-related death[42,57], particularly in men with limited life expectancy at the time of diagnosis[58]. Consequently, overdiagnosis may result in overtreatment which is associated with side-effects, reduced quality of life, psychological burden, and increased economic burden[57,58]. Overdiagnosis and overtreatment can be reduced through the application of careful selection criteria for PCa screening and appropriate indication for active surveillance or watchful waiting, instead of aggressive therapies, in eligible patients with low-risk PCa[31,57-59]. Furthermore, the risks and benefits of PCa screening should be discussed with men prior to PSA testing to support informed decision making[31].

The European Association of Urology (EAU) recommends a risk-adapted workflow for the decision on a prostate biopsy which includes an initial risk assessment based on several patient-individual selection criteria, such as age, life expectancy, comorbidities, family history, known carriage of *BRCA2* mutations, and ethnicity (African descent), but also available results from PSA testing and DRE[31]. Genetic counselling and germline testing should be considered for men with a known *BRCA* mutation on somatic testing, or in case of family history of high-risk germline mutations or multiple cancers on one family side[31]. This initial risk assessment should be followed by risk stratification to guide tissue biopsy, integrating a risk calculator, the performance of a prostate MRI, and potentially additional serum or urinary tests[31] (Figure 1). Risk calculators may include extended prostate-associated biomarkers, such as the percentage of free PSA or the PSA density, which is calculated as the ratio of total PSA to prostate volume. The assessment of the PSA density has been shown to enhance PCa detection[60-64] and may serve as an independent predictor for clinically significant PCa (csPCa) [65-67], but needs standardization of the prostate volume estimation with either DRE, trans-rectal ultrasound or MRI[31]. The American Urological Association (AUA) guidelines equally recommend the integration of risk calculators into the traditional PCa screening algorithm, including the assessment of PSA levels and the free PSA percentage, as well as age, ethnicity, and family history, combined with shared decision making with the patients[68].

Several blood-based tests and -risk scores that combine biomarkers with clinical patient characteristics have been shown to improve PCa diagnostics compared with standard PSA testing. The Food and Drug Administration (FDA)-approved *Prostate Health Index test* combines free PSA, total PSA, and [-2]pro-PSA isoform, while the *four kallikrein score* integrates free PSA, intact PSA, total PSA and kallikrein-like peptidase 2 with age, DRE findings, and prior biopsy status, in order to enhance risk stratification in men with elevated PSA levels[69-71]. The *IsoPSA test* is based on the detection of cancer-specific structural PSA isoforms[72], and the *Proclarix® test* combines thrombospondin-1, cathepsin D, total PSA, the percentage of free PSA, and age[73]. The comprehensive *Stockholm3 test* integrates several clinical variables (age, first-degree family history of PCa, prior biopsy status), multiple biomarkers (total PSA, free PSA, ratio of free PSA to total PSA, human kallikrein 2 (KLK2), macrophage inhibitory cytokine-1 (MIC-1 (GDF15)), and microseminoprotein-β (MSMB, (prostate secretory protein 94, PSP94)), and a polygenic risk score based on single nucleotide variants for risk stratification[74,75]. Additionally, urine-based tests are commercially available or implemented in investigational settings. The FDA-approved *Progensa urine test* measures the long non-coding RNA, *prostate cancer associated 3* (*PCA3*), which is known to be overexpressed in urine, in association to PCa[76]. The *SelectMDX score* evaluates the expression of the messenger ribonucleic acids (mRNA) *homeobox C6* (*HOXC6)* and *distal-less homeobox 1 (DLX1)* in urine, combined with clinical risk factors, such as age, DRE, total PSA and prostate volume[77,78]. The *Michigan Prostate Score* combines the detection of the *transmembrane serine protease 2* (*TMPRSS2)- ETS transcription factor ERG* (*ERG)* fusion and the *PCA3* expression in urine with serum PSA levels for prediction of (high-grade) PCa on biopsy[79], while the *ExoDx Prostate IntelliScore test* investigates exosomal mRNA in urine[80]. The different blood and urine tests or combined risk scores aim to enhance the detection of csPCa and to improve risk stratification prior to tissue biopsy to avoid unnecessary examinations.

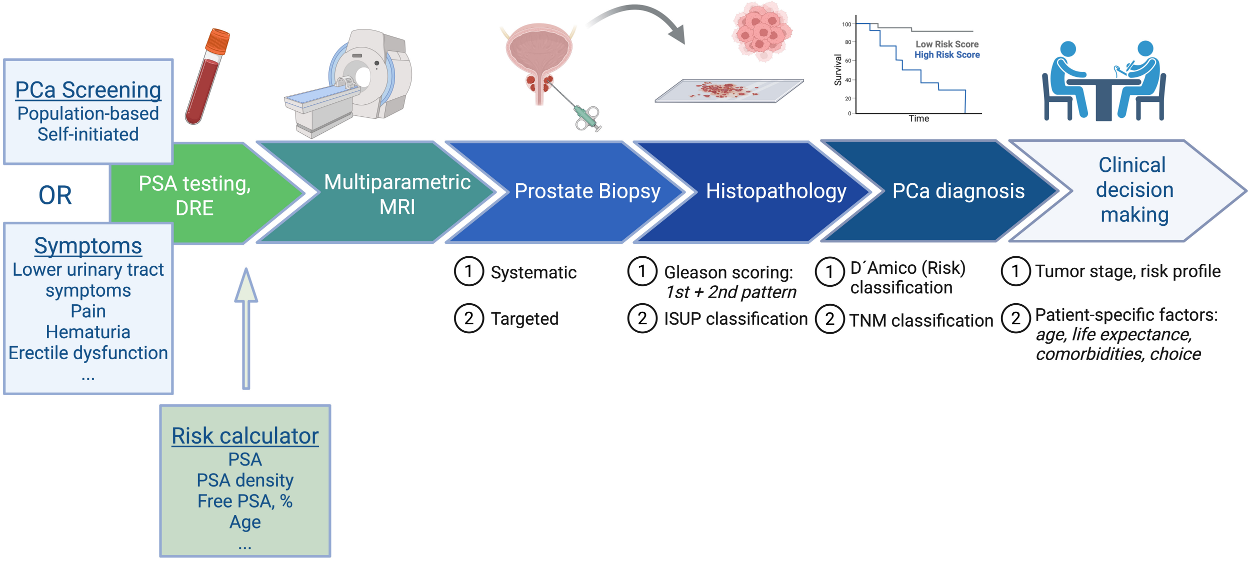


Figure 1: Diagnostic workflow and initial risk assessment for PCa.
Schematic overview of the current standard diagnostic workflow for PCa, including clinical assessment, prostate imaging by (mp)MRI, and histopathological evaluation of prostate tissue obtained by biopsy. Initiation of the diagnostic pathway is recommended based on risk factors, clinical symptoms, suspicious findings on DRE, and/or elevated or dynamic PSA levels. International guidelines further recommend the assessment of additional parameters for initial risk stratification. Abbreviations: DRE = digital rectal examination, (mp)MRI = (multiparametric) magnetic resonance imaging, PCa = prostate cancer, PSA = prostate-specific antigen. Figure created with BioRender.com.

The performance of an mpMRI, which combines different sequences such as T1-weighted, T2-weighted, diffusion-weighted-, and dynamic contrast-enhanced imaging, represents an essential examination in the diagnostic workflow for PCa[81,82]. MRI is commonly performed after the initial risk assessment process based on PSA testing, DRE, and clinical risk factors (Figure 1), but has also been evaluated as initial PCa screening method[83]. MpMRI was shown to harbor a high accuracy and sensitivity in the identification of csPCa, confirmed with histopathological assessment of the prostate tissue biopsy or prostatectomy tissue, especially for higher Gleason Score and larger tumor size[84-86]. Recent approaches have also investigated the extraction of multiple radiological parameters from mpMRI (radiomic features), as well as the implementation of new technologies such as artificial intelligence to integrate these features into predictive models for improved radiological decision-making and PCa detection[87,88]. The assessment of prostate MRI is recommended using the Prostate Imaging-Reporting and Data System (PI-RADS), which enables a standardized interpretation of the imaging data and subsequent estimation of the PCa risk[89,90]. The PI-RADS classification applies a scaling from 1 to 5 for the estimation of the likelihood of csPCa (Table 2). The radiological assessment guides the prostate tissue biopsy and is used to identify suspicious index lesions which can subsequently be targeted in the tissue biopsy for histopathological assessment (Figure 1).

Table 2: Likelihood of csPCa estimated with the PI-RADS classification.
Reported PCa detection rates result from a meta-analysis (including 17 studies) which evaluated the predictive ability of MRI assessment using PI-RADS version 2.1[90] for PCa detection[91].

PI-RADS score	likelihood of malignancy/presence of csPCa	PCa detection rate (95% CI)
1	very low; csPCa is highly unlikely	lesion level: 2% (0–8%) patient level: 6% (0–20%)
2	low; csPCa is unlikely	lesion level: 4% (1-9%) patient level: 9% (5-13%)
3	intermediate; presence of csPCa is equivocal	lesion level: 20% (13-27%) patient level: 16% (7-27%)
4	high; csPCa is likely	lesion level: 52% (43-61%) patient level: 59% (39-78%)
5	very high; csPCa is highly likely	lesion level: 89% (76-97%) patient level: 85% (73-94%)

Abbreviations: CI = confidence interval, csPCa = clinically significant PCa, PI-RADS = Prostate Imaging-Reporting and Data System

The performance of the prostate tissue biopsy is recommended as an ultrasonic-guided, systematic biopsy of the whole prostate gland assessing 10–12 cores, combined with an additional targeted biopsies of radiological defined or palpatory suspicious lesions[31,40,68]. The tissue biopsy can be performed using either a transrectal or transperineal approach. The EAU guidelines recommend the performance of a transperineal biopsy[31], due to a decreased risk of infectious complications[92]. Further reported adverse events following prostate tissue biopsy include haematuria, haematospermia/-ejaculate, infection, fever, and pain, with minor events occurring more frequently, while major events have been rarely observed[42,93,94]. The final PCa diagnosis is based on the histopathological detection of tumor cells in the systematic and/or targeted tissue biopsies[31,40] (Figure 1). The combination of systematic and targeted biopsies showed the best performance for accurate detection of csPCa, since both techniques add substantial value to each other[86,95,96].

1.1.3 PCa risk classification and staging

The EAU guidelines[31] recommend clinical decision-making for an appropriate, risk-adapted therapy management and estimation of the prognosis of newly diagnosed PCa based on the staging with the UICC Tumour, Node, Metastasis (TNM) classification (8th edition[11]) and the clinical risk assessment with the D'Amico classification system[97]. The expansion of the primary tumor (clinical T-status) in the TNM classification, which is also one criterion in the D'Amico classification system (Table 3[97]), is determined using the DRE, as also recommended by the American Joint Committee on Cancer staging[98]. The pathological TNM staging is performed based on the histopathological assessment of the prostatectomy tissue specimen[31]. Computed tomography (CT) and MRI might be complemented for the clinical assessment of lymph node metastases, but with limited sensitivity[99]. Prostate-specific membrane antigen (PSMA) positron emission tomography (PET)/CT uses fluorid-18- or gallium-68-labelling of PSMA which is known to be overexpressed on the surface of PCa cells[100]. PSMA-PET/CT revealed similar or higher sensitivity for the detection of nodal metastases compared to MRI and/or CT scan, and has been shown to improve performance of nomograms for the preoperative prediction of pelvic lymph node metastases[101-104]. Additionally, PSMA-PET/CT showed improved performance for primary staging of PCa with an intermediate- to high-risk for metastatic diseases, revealing higher accuracy and sensitivity for the detection of pelvic lymph nodes or distant metastases compared with conventional imaging modalities, such as bone scintigraphy and CT scan[105,106]. However, bone scintigraphy remains an essential conventional imaging technique with high sensitivity and specificity for the detection of bone metastases in PCa, and continues to be recommended in clinical guidelines[31,40,107], particularly for symptomatic patients[31,40] and patients with increased PSA ≥ 20 ng/ml or Gleason Score ≥ 8[108].

Following prostate tissue biopsy, patients with localized or locally-advanced PCa are stratified with regard to their expected (low-, intermediate-, high-) risk for biochemical recurrence after curative intended radical prostatectomy or radiotherapy, using the D'Amico classification system[97]. This system integrates three criteria: the PSA level, the histopathological grade derived from the tissue biopsy, and the clinical T-stage based on the TNM classification (Table 3). The pathological assessment of the prostate tissue biopsy applies the Gleason grading system, considering the histological architecture of the tumor, which is subsequently used for the grouping based on the International Society of Urological Pathology (ISUP) grade system (Table 4)[109].

Table 3: D'Amico Risk classification system.
Overview of the risk classification system developed by D'Amico and colleagues[97], which considers initial PSA levels, Gleason grading based on prostate biopsy, and clinical T stage to stratify patients with localized PCa according to the estimated risk of biochemical recurrence following radical prostatectomy or radiotherapy.

risk category	criteria
Low-risk	PSA ≤ 10 ng/ml and Gleason Score 6 (Gleason Grade Group/ISUP I) and cT1c or cT2
Intermediate-risk	PSA > 10 ng/ml and ≤ 20 ng/ml or Gleason Score 7 (Gleason Grade Group/ISUP II and III) or CT2b
High-risk	PSA > 20 ng/ml or Gleason Score ≥ 8 (Gleason Grade group/ISUP IV and V) or cT2c

Abbreviations: ISUP = International Society of Urological Pathology, PSA = prostate-specific antigen

The discrimination between csPCa or insignificant PCa is essential for a risk-adapted therapy decision to avoid both over-treatment of indolent (insignificant) PCa and under-treatment of potentially aggressive (significant) PCa[31]. This distinguishment can be made based on tumor stage and grade, as well as patient-specific factors, but a uniform consensus is lacking[31]. While low-risk PCa is commonly considered as clinically insignificant PCa, intermediate-risk PCa has been shown to harbor heterogeneous risk profiles, resulting in an unmet need for enhanced stratification[110-112]. The D'Amico classification established in the late 1990s[97] is still considered as an essential guide for risk stratification, but adapted or extended risk classifications have been proposed, such as the EAU risk groups[31], the Cambridge Prognostic Groups[113] or the National Comprehensive Cancer Network guidelines[114]. The EAU risk group classification extends the criteria of the D'Amico classification by introducing an additional sub group which categorizes all locally-advanced tumors with detectable lymph node metastases in the CT or bone scan, regardless of the PSA level or the Gleason Score[31].

1.1.4 Histopathology of PCa

The majority of prostate carcinomas are classified as acinar adenocarcinoma which forms glandular structures and originates from the epithelial prostate cells[115]. The Fifth Edition of the World Health Organization (WHO) Classification of Tumours of the Urinary and Male Genital Systems categorizes overall 11 PCa subtypes, which can be assigned to two major groups: epithelial or mesenchymal tumors of the prostate[116]. The category of epithelial tumors comprises six subgroups of glandular neoplasms, and three subgroups of (baso)squamous neoplasms. The group of glandular tumors includes the predominantly occurring prostatic acinar adenocarcinoma, benign lesions (prostatic cystadenoma) and precursor tumors (high-grade prostatic intraepithelial neoplasia (PIN)), as well as the intraductal carcinoma of the prostate, the prostatic ductal adenocarcinoma, and the treatment-related neuroendocrine prostatic carcinoma[116]. The category of mesenchymal tumors comprises stromal tumors of the prostate, including stromal tumors of uncertain malignant potential and stromal sarcoma[116].

Primary PCa frequently occurs as multifocal disease with at least two or more topographically separate tumor lesions in the prostate gland[117]. These lesions can harbor different subclones and mutually exclusive molecular alterations, suggesting separate tumor evolution[117]. The prostate gland is partitioned into three zones based on their histological features: the central zone (20–25%), the transition zone (5–10%), and the peripheral zone (~70%)[118,119]. PCa incidence varies highly between the three zones, and tumor lesions harbor different molecular and cellular characteristics, depending on the zone they are located in, which are associated with varying aggressiveness and prognosis[118]. The highest incidence of PCa (~70%) has been found in the largest area of the prostate, the peripheral zone, and these tumors tend to show an infiltrative growth pattern, higher Gleason scores, a more aggressive molecular risk profile and a worse clinical outcome[118]. Up to 25% of prostate tumors are located in the transition zone and these tumors often harbor well-differentiated morphological patterns, similar to BPH nodules which are also predominantly located in the transition zone[118]. These tumors in the transition zone are associated with an overall better prognosis compared with the ones located in the peripheral zone[118]. PCa lesions occur rarely in the central zone (5–10%), but display morphological features similar to tumors located in the peripheral zone and tend to be highly aggressive[118].

The histopathological assessment of the prostate tissue biopsy is considered as the current gold standard for the PCa diagnosis and the determination of the PCa risk profile. The tumor is diagnosed as acinar adenocarcinoma, if all three of the following criteria are met: architectural disorder of the glands, nuclear atypia, and exclusion of a benign lesion[40]. The lesion is diagnosed as atypical small acinar proliferation (ASAP) or atypical glands suspicious for prostate cancer, if only two of three criteria are met[40]. Precursor lesions of the prostatic carcinoma (PIN) show prominent nucleoli and other architectural changes, mimicking the cytology of cancer cells[120,121]. Extensive high-grade PIN and ASAP have been associated with an increased risk to develop into invasive PCa[122,123], and should be controlled with repeated tissue biopsy[40]. In cases of uncertain malignancy based on morphological criteria, additional immunohistochemical analyses are recommended[40]. These include the assessment of basal cell markers to demonstrate the absence of basal cells in PCa (e.g., transcription factor p63 or basal cell-specific cytokeratin 5/6 and 34bE12) or the analysis of indicative markers for PCa, such as Alpha-methylacyl-CoA racemase, which is frequently overexpressed in PCa cells[124].

International grading of PCa is commonly performed using the Gleason grading system (Figure 2), as defined by the ISUP Consensus Conference, in which grades are assigned based on the appearance of glandular formations[109,125,126]. The Gleason grading system was established in 1966 and originally involved five grades (Gleason 1–5)[126,127], but in practice has been often reduced to three grades (Gleason 3–5) [125]. For the assessment of prostate tissue biopsies, both the appearance of the most common and the worst glandular pattern are graded, and the final Gleason score is calculated based on the combination of the two grades[109,125] (Figure 2). If only one pattern is present, then the respective grade is doubled to determine the Gleason score. Accordingly, Gleason scores range from 6 to 10, with a score of 10 representing a highly undifferentiated structure. All cores obtained from the systematic biopsy and all targeted lesions should be assessed and graded separately to define an additional overall Gleason Score[31,40,109]. The resulting Gleason Score guides assignment of ISUP grade groups, ranging from 1 to 5 (Table 4), which have been shown to more accurately reflect the cancer biology than the Gleason system alone and can be used for prognostic estimation[125].

Table 4: Gleason score and corresponding ISUP grade groups.
Assignment of Gleason scores to the corresponding ISUP grade groups according to the International Society of Urological Pathology 2014 grading system[125]. Gleason scores range from 6 to 10, with higher scores indicating decreasing tumor differentiation. Gleason scores are determined based on pathological evaluation of prostate tissue biopsies.

Gleason score	ISUP grade
≤ 6	1
7 (3+4)	2
7 (4+3)	3
8 (4+4 or 3+5 or 5+3)	4
9-10 (4+5 or 5+4 or 5+5)	5

When radical prostatectomy is performed, subsequent histopathological assessment of the available prostatectomy specimen complements the existing staging and grading, but may differ from the preoperative evaluation due to the multifocality and heterogeneity of PCa. For this reason, the entire prostatectomy tissue sample should be examined histopathologically to allow a comprehensive determination of the highest grade and identification of residual tumor cells, corresponding to the resection margin status (R-status) within the TNM classification[40]. Pathological staging is performed using the TNM classification[11], assigning the T-status based on the extent of the primary tumor and the N-status based on the presence or absence of lymph node metastases when lymphadenectomy has been performed[31,40]. Gleason grading is also applied to prostatectomy specimens, but the Gleason score is calculated based on the most frequent and the second most prevalent glandular pattern (Figure 2)[109,125]. In addition, tertiary (minor) Gleason patterns 4 and 5 can be reported, if the affected region extends > 5% of the tumor volume[109].

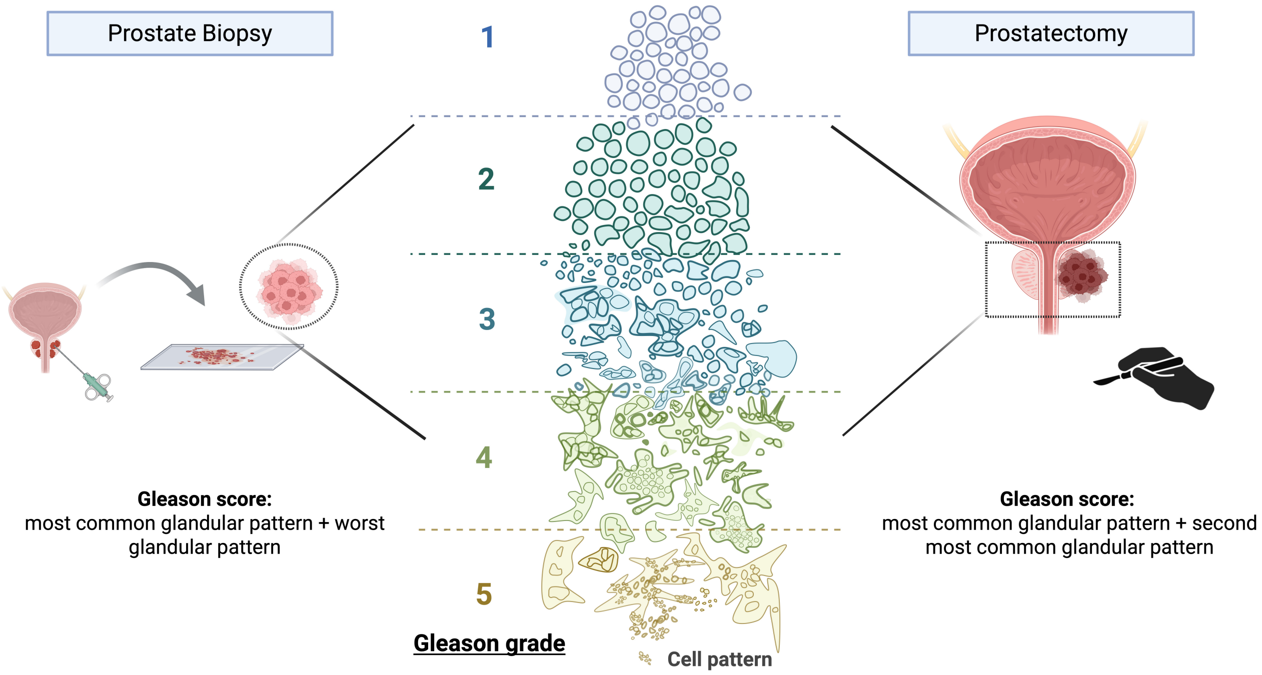


Figure 2: Histopathological assessment and grading of PCa.
Histopathological grading of PCa is performed using the Gleason grading system, which assigns grades based on the architectural appearance of glandular formations. The Gleason system includes five grades (Gleason patterns 1–5)[126,127]. For prostate biopsy specimens, the most prevalent and the highest-grade glandular patterns are identified and summed to obtain the final Gleason score. For prostatectomy specimens, the Gleason score is calculated based on the most frequent and the second most frequent glandular patterns. Gleason scores range from 6 to 10, with higher scores indicating increasing tumor dedifferentiation. Figure created with BioRender.com.

1.1.5 Treatment of PCa

Available therapy options for primary PCa are multifactorial and involve risk-adapted decision for active surveillance or low-dose brachytherapy in low-risk PCa, radical prostatectomy or definitive (percutaneous) radiotherapy for significant, localized (low-), intermediate- or high-risk PCa with curative intention, and systemic therapy in the advanced, metastatic setting (Figure 3). Watchful waiting can be applied in the palliative setting to patients with all PCa stages and life expectancy of less than 10 years or in case of other patient-specific circumstances[31]. The decision-making for a therapy regime at initial PCa diagnosis is based on clinical staging, PSA levels, histopathological assessment of the tissue biopsy, and patient-individual factors[31,40,118] (Figure 1). These include age and life expectancy of the patient, his health status and potential comorbidities, as well as the patient's attitude toward therapy options for shared decision-making[31,40].

The implementation of active surveillance with curative intention aims to reduce overtreatment of indolent, low-risk PCa. Active surveillance is recommended for patients with life expectancy of more than 10 years and low-risk PCa (D'Amico criteria or ISUP grade 1; Figure 3), but could also be considered for PCa with ISUP grade II and favorable risk profile based on histopathological assessment of the prostate tissue biopsy[31,40,128]. The surveillance is performed using DRE, PSA testing, MRI and re-biopsies based on a defined schedule, and should reduce treatment-related side effects without shortening survival of the patients with indolent PCa[31]. A large prospective study with 1818 PCa patients, categorized as ISUP 1, under active surveillance with an intensive structured program revealed a cumulative incidence of 0.1% for PCa-specific mortality or progression to metastatic disease over 10–15 years[129]. Similarly, the Prostate Cancer Intervention versus Observation Trial and the ProtecT trial compared curative-intended treatment (radical prostatectomy or external-beam radiotherapy) versus (active) monitoring of patients with localized PCa, and did not identify significantly decreased overall or PCa-related mortality within a 12- or 10-year time period, respectively, related to treatment compared with observation[59,130]. However, active monitoring was associated with higher incidence of disease progression and development of metastases compared to both prostatectomy and radiotherapy in the ProtecT trial[130]. The Prostate Cancer Guideline Panel of the EAU established the DETECTIVE study and a systematic review aiming to achieve standardization and to develop consensus recommendations with regard to inclusion, monitoring and reclassification criteria, as well as structured programs for active surveillance in PCa[128,131].

The curative-indented therapy options for eligible patients with localized PCa and locally-advanced PCa include radical prostatectomy and radiotherapy[31,40] (Figure 3). The selection of one therapy option must consider the PCa risk profile, the patient's health status and comorbidities, the expected therapy-related side-effects, as well as patient and physician preferences. Radical prostatectomy may be selected for patients with localized PCa of all stages, particularly if there is a high likelihood of achieving a complete tumor removal through surgery[40]. Several surgical approaches are available for the complete resection of the prostate gland, each associated with specific advantage, disadvantage and adverse events. These approaches include retropubic or perineal surgical access, as well as laparoscopic or robot-assisted laparoscopic procedures using intra- or extraperitoneal access[40]. In addition, a nerve-sparing approach targeting parasympathetic nerve branches of the pelvic plexus can be applied for eligible patients to maintain the erectile function[132,133]. The radical prostatectomy can be combined with a pelvic lymph node dissection for patients with an increased risk for lymph node metastases or with suspicious clinical or radiological findings[31,40]. An adjuvant radiotherapy might be considered for patients with locally-advanced PCa and increased risk profiles, especially for PCa with high Gleason scores or in cases with residual tumor cells (R1 resection) after surgery[40]. Major postoperative complications following radical prostatectomy include urinary incontinence and erectile dysfunction, with reported frequencies of up to approximately 20–30% and 70%, respectively, depending on patient cohort and surgical technique[134,135]. In addition, extensive pelvic lymph node dissection was associated with increased occurrence of lymphoceles[136,137]. Further intra- and peri-operative complications, such as urinary retention, wound complication, pain, anastomotic leak, and infection[31,138,139] have also been reported and should be comprehensively discussed with the patients prior to the surgical treatment[139].

The percutaneous radiotherapy can be selected as alternative therapy strategy for patients with localized PCa of different risk profiles[40] (Figure 3). The large randomized controlled ProtecT trial reported 10-years outcomes for patients who were diagnosed with localized PCa, undergoing either active surveillance, prostatectomy or external-beam radiotherapy[130]. While PCa-specific mortality was low in all cases, without significant differences among the treatments, both prostatectomy and radiotherapy revealed lower incidences of disease progression and development of metastatic disease than active surveillance[130]. Additional studies on patients with locally-advanced PCa revealed comparable outcomes for the percutaneous radiotherapy and prostatectomy, with regard to the biochemically 5-year progression free survival and the PCa-specific 5- and 10-year survival, but slightly increased 10 year-biochemically progression free survival and overall survival for prostatectomy[40]. Percutaneous, dose-escalated radiotherapy is recommended to be applied using intensity-modulated techniques, combined with image-guided technologies, employing standard fractionation with a minimum dose of at least 64–80 Gy[31,40]. In case of intermediate-risk PCa, high-risk PCa or locally-advanced PCa, it is recommended to combine radiotherapy with androgen deprivation therapy (ADT) for at least 24 or 36 months which can already be started 4–6 months prior to the radiotherapy[40]. The brachytherapy can be considered as monotherapy option for low-risk, localized PCa, or combined with percutaneous radiotherapy and/or ADT for localized PCa with different risk profiles[40] (Figure 3).

For lymph-node positive PCa, both radical prostatectomy and percutaneous radiotherapy can be applied as primary therapy option[40] (Figure 3). The radical prostatectomy can be additionally combined with lymphadenectomy or adjuvant radiotherapy of the pelvine lymph nodes[40]. Furthermore, the radical prostatectomy can and the radiotherapy should be combined with adjuvant ADT for at least 24 or 36 months[40].

The current standard therapy options focus on the treatment of the whole prostate gland. However, it is known that prognosis, as well as morphological and molecular features can differ highly, depending on the prostate zone in which tumors are located, and these aspects are currently not comprehensively considered in the clinical guidelines[118]. In this context, a zone-specific risk stratification based on clinical and molecular biomarkers has been proposed, which could guide clinical decision-making[118]. New approaches investigate treatment strategies that focus on local tumor regions, such as partial prostatectomy, high-intensity focused ultrasonography, radiofrequency ablation and focal high-dose rate brachytherapy[40,118].

The assessment of the serum PSA levels can be used to monitor therapy response after radical prostatectomy or radiotherapy, as well as for the detection of biochemical recurrence. Following radical prostatectomy the PSA level is expected to decline to zero, and a subsequent increase of the PSA level > 0.2 ng/ml (> 0.4 ng/ml) is considered as specific sign for PCa recurrence[31,40]. Following radiotherapy, an increase of PSA levels of > 2 ng/mL above the post-interventional PSA nadir, regardless of the nadir's level, is considered as biochemical recurrence[31,40]. The recurrence can occur locally or systemically with lymph node or distant metastases[31]. Biochemical recurrence occurs in approximately 27% of patients undergoing radical prostatectomy and 53% following radiotherapy[31]. However, not every PSA recurrence is associated with progressive disease and may instead remain limited to a PSA-only recurrence[31]. It has been shown that biochemical recurrence does negatively impact clinical outcomes, indicating that a subgroup of patients with poor, clinical and pathological risk profile have a higher risk of developing distant metastases, as well as PCa-specific and overall mortality[140]. For patients with biochemical recurrence following radical prostatectomy, percutaneous salvage radiotherapy may be performed alone, or in combination with either ADT or anti-androgen medication with bicalutamide[40] (Figure 3). Conversely, salvage radical prostatectomy may be performed in patients with local tumor recurrence following radiotherapy[31,40] (Figure 3).

When a curative-intended local treatment is not desired or not applicable, due to patient-related factors or an advanced tumor stage at initial PCa diagnosis, an appropriate palliative setting has to be pursued[31,40] (Figure 3). This includes either watchful waiting or systemic ADT, combined with a symptom-oriented therapy[31,40]. The watchful waiting strategy aims to avoid side-effects associated with androgen deprivation, but it may be converted to symptom-oriented systemic or local treatment in the presence of clinical symptoms[31]. Primary treatment option for metastatic, androgen-sensitive PCa include different ADT combination therapies, depending on the PCa risk profile and patient-related factors (Figure 3). Metastatic, hormone-sensitive PCa can be classified as “synchronous” for primary, metastatic PCa or “metachronous” for recurrent, metastatic PCa[40,141]. Disease course and prognosis are impacted by the tumor volume, tumor burden and the risk profile, as classified by the presence of visceral metastases, the number of bone metastases and/or a high Gleason score[40,141-143]. Synchronous, metastatic PCa tends to be more aggressive and is associated with higher tumor burden and poorer prognosis than metachronous, metastatic PCa, which is linked to longer therapy response and improved overall survival[144,145]. For synchronous, metastatic PCa with low tumor volume, local therapy of the primary tumor, followed by adjuvant ADT combination therapy, may be applied[141]. Patients with synchronous, metastatic PCa and high tumor volume or with metachronous, metastatic PCa predominantly undergo systemic treatment with ADT combination therapies. Metastasis-directed, local therapy may be considered for oligometastases[141,146,147] and symptomatic metastases, in addition to pain therapy[40]. ADT can be performed with Gonadotropin-releasing hormone (GnRH) agonists or antagonists, and in rare cases with surgical castration[141]. ADT combination therapies include antiandrogen drugs, such as abiraterone which inhibits the androgen synthesis catalyzing enzyme CYP17A1, or androgen receptor pathway inhibitors (ARPI) apalutamide, darolutamide and enzalutamide. In addition, triple therapies which combine ADT with docetaxel and abiraterone/prednisolone or darolutamide, respectively, can be applied[141]. The systemic treatment aims to stabilize or delay tumor progression and to reduce tumor-associated, local complications and metastasis-related symptoms[148,149]. It has been shown that 70–80% of metastatic PCa patients respond to ADT, achieving a stabilization of the tumor disease for up to 24 months, until resistance occurrence and further progression[144]. Two trials have demonstrated that the combination of ADT and docetaxel prolongs overall survival for 10 months in metastatic PCa[150,151]. Similarly, improved overall survival and/or reduced mortality risk were observed in the combination therapy of ADT with other antiandrogen drugs[143,152-154]. Triple therapy of ADT and docetaxel combined with either abiraterone/prednisolone or darolutamide achieved hazard ratios of 0.75 and 0.68, respectively, for PCa-related mortality compared with ADT monotherapy in metastatic, hormone-sensitive PCa[155,156].

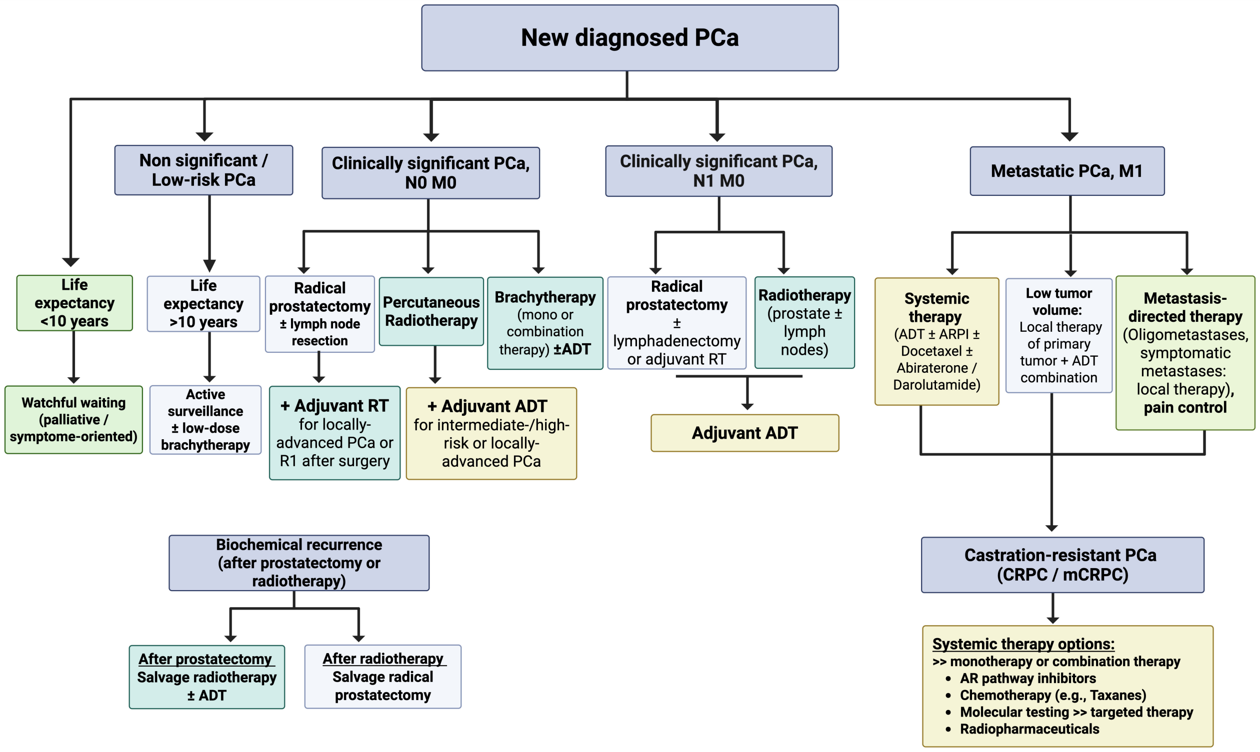


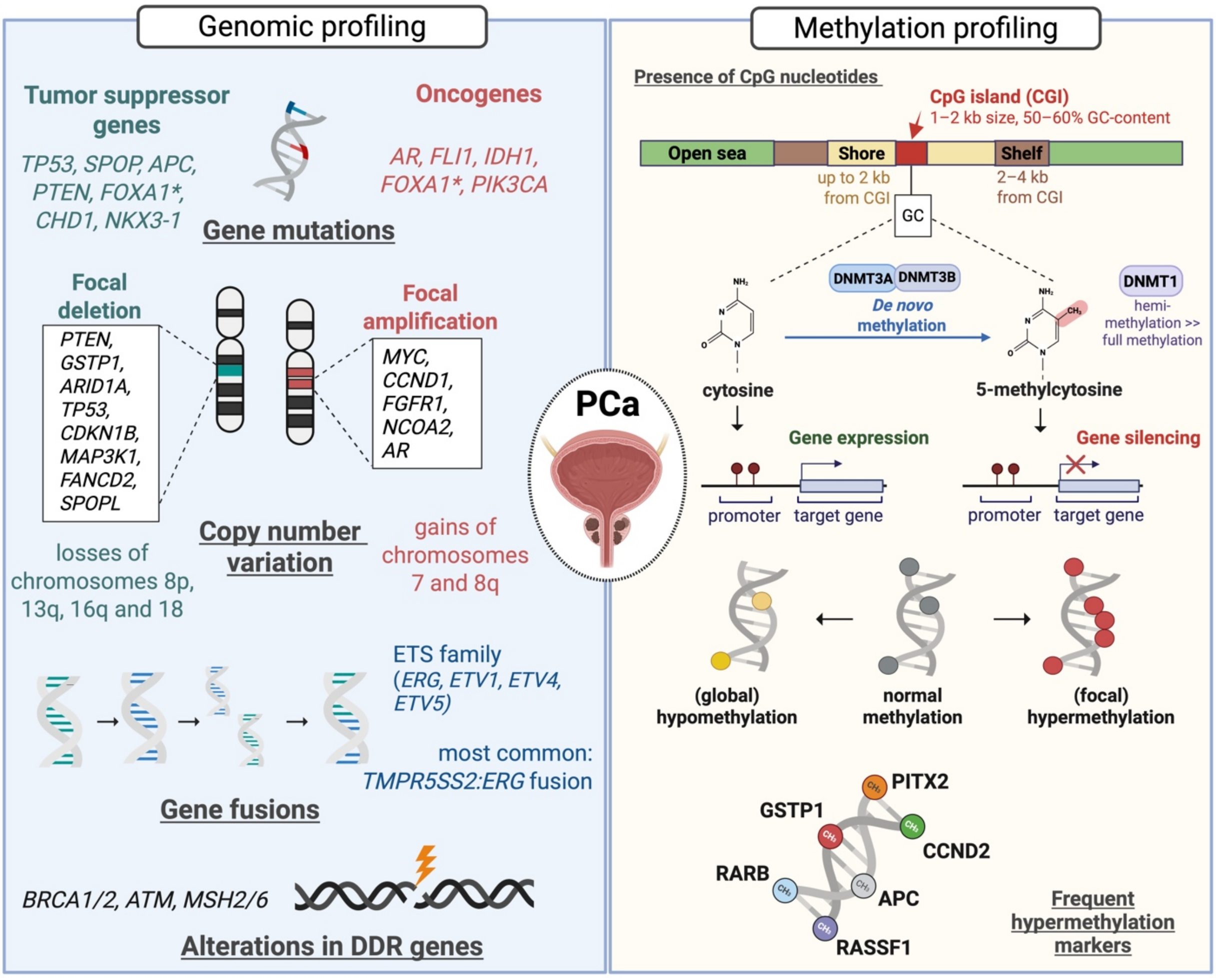
Figure 3: Treatment options for localized and advanced PCa.
Schematic overview of treatment strategies for localized and advanced PCa. Therapeutic decision-making in PCa is multifactorial and risk-adapted, involving active surveillance or low-dose brachytherapy in low-risk disease, radical prostatectomy or definitive (percutaneous) radiotherapy in localized intermediate- or high-risk disease, and systemic therapy or watchful waiting in the advanced and metastatic setting. Abbreviations: ADT = androgen deprivation therapy, AR = androgen receptor, (m)CRPC = (metastatic) castration-resistant prostate cancer, M0/M1 = absence/presence of distant metastases, N0/N1 = absence/presence of lymph node metastases, RT = radiotherapy. Figure created with BioRender.com.

More than 90% of patients treated with ADT develop androgen-independent PCa during therapy course[144]. They are subsequently classified as (metastatic) castration-resistant PCa ((m)CRPC), when the testosterone level is below 50 ng/dl, accompanied by biochemical progression with elevated PSA levels and/or clinical evidence for progression on imaging[31,40]. In this context, molecular testing is recommended with somatic testing for homologous recombination (HR) and mismatch repair (MMR) defects, as well as germline testing for *BRCA1* and *BRCA2*, *ATM serine/threonine kinase* (*ATM)* and MMR genes in high-risk- and mCRPC[31,40]. For the treatment of mCRPC, approved therapies in Europe are docetaxel, abiraterone/prednisolone, enzalutamide, cabazitaxel, olaparib, niraparib + abiraterone/prednisolone, talazoparib + enzalutamide, radium-223 and lutetium-177 (^{177}Lu) vipivotide tetraxetan, whereas for non-mCRPC, darolutamide and apalutamide are approved[31]. A wide range of mono- and combination treatment regimens are available, and can be selected based on the molecular status, previously applied therapies, and patient-individual factors.

Alongside tumor progression to CRPC, differentiation from adenocarcinoma to neuroendocrine PCa (NEPC) has been observed[157,158], occurring in up to 20% of CRPC[117,159]. The development of *de novo* NEPC, without any prior treatment, is rare (< 1%)[160]. NEPC resembles the morphology of small-cell carcinoma or high-grade neuroendocrine carcinoma, and harbors additional characteristics, such as low amount or absence of androgen receptors (AR) and expression of neuroendocrine markers[117,157,158]. It has a similar mutational landscape as castration-resistant adenocarcinoma, with frequent occurrence of genomic alterations in *RB transcriptional corepressor 1 (RB1)* and *tumor protein p53 (TP53)* [158,159]. Characteristic epigenetic changes have also been identified[117,158], involving pathways associated with neuronal, developmental and stem cell programs, as well as cell-cell adhesion and epithelial-mesenchymal transition[158].

1.1.6 Molecular profiling of PCa

PCa is a heterogeneous disease with regard to its molecular landscape and the associated risk profiles[117,161]. Comprehensive molecular profiling using next generation sequencing (NGS) methods and gene expression analyses revealed insights into the most frequent genomic alterations, epigenetic changes and associated biological pathways (Figure 4), supporting a better understanding of PCa carcinogenesis and progression, and potentially serving as targets for diagnostic and therapeutical applications[161,162].



*tumor suppressor gene or oncogene

Figure 4: Molecular landscape of PCa.
Schematic overview of the molecular heterogeneity of PCa, highlighting genomic and epigenomic alterations associated with PCa carcinogenesis and disease progression. Genomic alterations include mutations in tumor suppressor genes and oncogenes, alterations in DNA damage repair genes, copy number variations, and gene fusions. Epigenomic changes include aberrant DNA methylation patterns, such as focal hypermethylation, particularly affecting gene promoter regions, and global hypomethylation. Abbreviations: *APC = APC regulator of Wnt signaling pathway*; *AR = androgen receptor; ARID1A = AT-rich interaction domain 1A; ATM = ATM serine/threonine kinase; BRCA1/2 = BRCA1/2 DNA repair associated*; *CCND1/2 = cyclin D1/D2*; *CDKN1B = cyclin dependent kinase inhibitor 1B*; CGI = CpG island; *CHD1 = chromodomain helicase DNA binding protein 1*; DDR = DNA damage repair; DNMT = DNA methyltransferase; *ERG = ETS transcription factor ERG*; *ETS = E26 transformation-specific*; *ETV1/4/5 = ETS variant transcription factor 1/4/5*; *FANCD2 = FA complementation group D2*; *FGFR1 = fibroblast growth factor receptor 1*; *FLI1 = Fli-1 proto-oncogene, ETS transcription factor*; *FOXA1 = forkhead box A1*; *GSTP1 = glutathione S-transferase pi 1*; *IDH1 = isocitrate dehydrogenase (NADP(+)) 1*; *MAP3K1 = mitogen-activated protein kinase kinase kinase 1*; *MSH2/6 = mutS homolog 2/6*; *MYC = MYC proto-oncogene, bHLH transcription factor*; *NCOA2 = nuclear receptor coactivator 2*; *NKX3-1 = NK3 homeobox 1*; *PIK3CA = phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha*; *PITX2 = paired like homeodomain 2*; *PTEN = phosphatase and tensin homolog*; *RARB = retinoic acid receptor beta*; *RASFF1 = Ras association domain family member 1*; *SPOP(L) = speckle type BTB/POZ protein (like)*; *TMPRSS2 = transmembrane serine protease 2*; *TP53 = tumor protein p53*. Figure created with BioRender.com.

1.1.6.1 Genomic landscape of PCa

Genomic alterations involve mutations, insertions, deletions and fusions of genes, as well as chromosomal rearrangements, copy number variations (CNVs) and focal chromosomal gains or losses. Gene mutations can act as drivers for tumor development, but cancer cells frequently need more than one driver mutation for tumor initiation and progression[163,164]. However, the majority of mutations are passengers which do not substantially impact carcinogenesis[164]. Oncogenes are involved in normal cellular functions, such as cell growth, proliferation or apoptosis, but get frequently altered or highly expressed in tumor cells, leading to an overactivation of the subsequent molecular mechanisms and, consequently, to tumor development or progression[165]. In contrast, tumor suppressor genes control cell growth and are involved in the regulation of different cellular processes, such as proliferation or differentiation[166,167]. They are also frequently altered or lower expressed in tumor cells, resulting in a reduction or loss of function with subsequent abnormal cell growth[165-167]. Chromosomal amplifications and deletions equally influence the function of genes which are located on the respective aberrant genomic region, consequently leading to up- or downregulation of associated pathways. Another important gene class related to tumor development comprise DNA repair genes, which are involved in the repair of DNA single-strand damage or double-strand breaks[168]. Alterations in repair genes result in defects of the DNA repair function and irreparable DNA damage in the cells, consequently increasing their susceptibility for carcinogenesis[168,169].

The important role of genetic alterations for PCa development has previously been described in the context of PCa etiology. Alterations in DNA damage repair (DDR) genes are known to be key drivers for PCa tumorigenesis and aggressiveness[20,170]. They are associated with hereditary PCa and were also detected as somatic mutations in primary PCa, involving several frequent DDR genes, such as *BRCA1/2*, *ATM*, and *MSH2/6*[161,171] (Figure 4). Alterations in the DDR genes occur more frequently in advanced PCa, especially CRPC, than in localized PCa, suggesting their important role as drivers for PCa progression[172]. Additional prevalent genomic alterations in PCa involve gene fusions from the *E26 transformation-specific* (*ETS*) family genes (*ERG*, *ETS variant transcription factor 1* (*ETV1*), *ETS variant transcription factor 4 (ETV4)*, and *ETS variant transcription factor 5 (ETV5)*), representing one of the largest families of transcription factors, combined with different fusion partners, among which the gene *TMPRSS2* (21q22.3) is the most common one[161,173] (Figure 4). The most frequent fusion mechanism involves the deletion of segments from both genes, while translocations or insertions represent a rare mechanism. Consequently, the fusion of 5'-UTR of *TMPRSS2* with either *ERG* (21q22.2) or *ETV1* (7p21.2) results in the overexpression of these genes[174]. The *TMPRSS2-ERG* fusion has been observed in up to 50–60% of prostate tumors[20,117]. The presence of the *TMPRSS2-ERG* fusion has also been detected in precancerous, high-grade-PIN lesions[175], but not in benign hyperplasia or normal prostate tissue, indicating PCa-specificity[20] and highlighting its role for tumor development and progression. Another frequent genomic alteration in PCa is the deletion of the tumor suppressor gene *phosphatase and tensin homolog* (*PTEN)* on chromosome 10 (Figure 4), which has also been detected in high-grade PIN[176]. *PTEN* loss is associated with aggressive PCa and tumor progression to metastatic disease[172,177,178]. It often co-occurs with the *TMPRSS2-ERG* fusion and this combination has been suggested to promote carcinogenesis and has been linked to poorer prognosis[20,177,179].

Somatic point mutations occur less frequently in PCa compared to other tumor entities[161]. PCa is known to harbor low mutation rates, even in metastatic disease, with a mean mutation frequency of 0.9 per megabase in primary PCa[180] and few reported cases of increased mutation rates[178,181]. Thereby, *speckle type BTB/POZ protein* (*SPOP)*, *TP53*, *forkhead box A1 (FOXA1)* and *PTEN* have be identified as the most frequently mutated genes, with reported frequencies of up to 15% in primary

PCa, or even higher frequencies in metastatic disease[20,178,182] (Figure 4). Further known mutations in primary and advanced PCa involve the genes *Fli-1 proto-oncogene, ETS transcription factor (FLI1)*, *isocitrate dehydrogenase (NADP(+)) 1 (IDH1)*, *chromodomain helicase DNA binding protein 1 (CHD1)*, *NK3 homeobox 1NKX3-1, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA)* and *APC regulator of Wnt signaling pathway (APC)* [160,182,183]. Several of these somatic mutations were observed to occur in the absence of *ETS* gene fusions, and tended to be mutually exclusive[20,160,182]. Alterations in the tumor suppressor genes *TP53* and *RB1*, leading to the loss of function of these genes, were associated with PCa progression and aggressiveness[172,178,184] (Figure 4). *TP53* alterations have been shown to be enriched in mCRPC compared with primary PCa[178], similar to *AR* gene alterations, both mutations and amplifications, which occur frequently in advanced PCa and affect AR signaling[161,178]. Although primary PCa is predominantly androgen dependent, variable AR activity has been reported, potentially influenced by *SPOP* or *FOXA1* mutations and low levels of AR splice variants have also been observed[161].

Chromosomal arm-level CNVs are the most frequent molecular alterations in primary PCa[161] (Figure 4). However, studies have shown that their occurrence and extent vary considerably, ranging from cases with largely unaltered genomes to intermediate levels and a high CNV burden[161,162]. Common CNVs involve gains of chromosomes 7 and 8q, as well as heterozygous losses of chromosomes 8p, 13q, 16q, and 18[161]. With regard to focal changes, relatively few chromosomal gains and losses have been described in PCa, similar to its low mutation rates[20]. Focal amplifications are less common than deletions and affect the oncogenes *MYC proto-oncogene, bHLH transcription factor (MYC), cyclin D1 (CCND1),* and *fibroblast growth factor receptor 1 (FGFR1),* and other genes such as *nuclear receptor coactivator 2 (NCOA2)* and *AR*[161,183]*.* Focal deletions involve tumor suppressor genes such as *PTEN*, *glutathione S-transferase pi 1 (GSTP1), AT-rich interaction domain 1A (ARID1A)*, *TP53*, *cyclin dependent kinase inhibitor 1B (CDKN1B)*, *mitogen-activated protein kinase kinase kinase 1 (MAP3K1)*, *FA complementation group D2 (FANCD2)* and *speckle type BTB/POZ protein like (SPOPL)* [160,161,183]. Genomic instability and chromosomal aberrations have been associated with PCa progression and aggressiveness[160,162,183].

1.1.6.2 Epigenomic landscape of PCa

Epigenomic changes have emerged as an additional key mechanism contributing to tumorigenesis. Consequently, they have gained interest in cancer research for improving the understanding of biological processes and for the discovery of novel diagnostic biomarkers and therapeutic targets[185]. Epigenetic mechanisms involve modification of the DNA, such as DNA methylation, posttranslational modification of histones (acetylation, phosphorylation, methylation and ubiquitylation), and chromatin remodeling[186].

DNA methylation is characterized by the enzymatic transfer from the methyl donor S-adenosylmethionine to the carbon-5 position of the nucleotide base cytosine, resulting in 5-methylcytosine[185] (Figure 4). This cytosine methylation occurs predominantly at CpG nucleotides, and around 70% of approximately 50 million CpG dinucleotides are methylated in the human genome[185,186]. Non-CpG methylation, such as CpA and CpT, has also be identified in embryonic stem cells, but not in somatic cells[187]. The enzymatic transfer of the methyl-group is catalyzed by three DNA methyltransferases (DNMTs) from the same enzyme family: DNMT1, DNMT3A and DNMT3B[185,186]. DNMT1 is ubiquitously expressed and the main enzyme for the conversion of hemi-methylated sites to fully methylated DNA after replication, whereas DNMT3A and DNMT3B are involved in *de novo* methylation of new methylation sites and patterns[185,186]. While the majority of CpG dinucleotides are predominantly unequally distributed, the human genome contains approximately

45,000 dense clusters of CpG dinucleotides with 1–2 kilobase (kb) size and > 50–60% GC-content, which are called CpG islands and account for approximately 2% of the genome[185,188,189]. The CpG islands are predominantly located at the 5' end of genes, and are surrounded by CpG shores with up to 2 kb distance from the CpG island, followed by shelfs with an additional distance of 2–4 kb, followed by open sea regions[188-190] (Figure 4). CpG islands are unevenly distributed between the chromosomes and some chromosomes contain many of them (e.g., chromosomes 19 and 22), while others harbor only a few (e.g., chromosome 18) [188]. Cytosines located within CpG islands are predominantly unmethylated in normal cells, with few exceptions associated to gene regulation, while the majority of methylated cytosines are located outside these regions, with a predominant occurrence in repetitive or satellite sequences[185,189]. Two effects have been observed physiologically during aging and, to a greater extent, pathologically in tumor cells: increased focal methylation in CpG islands and a genome-wide decrease of DNA methylation[185]. Aberrant DNA methylation is known to occur earlier in carcinogenesis[185,191,192]. The gain of methylated CpG sites in tumor cells relative to the physiological state is termed “hypermethylation”, whereas “hypomethylation” refers to a loss of DNA methylation[189]. Global DNA hypomethylation is a common tumor characteristic and predominantly occurs in repetitive elements of centromeric regions or satellite sequences[163,189]. In addition, hypomethylation has been identified in association to protooncogenes, consequently leading to gene activation[163,189,193]. DNA hypomethylation may occur early in tumorigenesis, but has also been observed during tumor progression[194]. Focal hypermethylation of promoter regions, predominantly associated with tumor suppressor genes, frequently occurs in tumor cells, leading to gene silencing[163,185,189]. Furthermore, DNA hypermethylation has been associated to silencing of transcription factors[163]. Aberrant methylation has also been identified, to varying extent, in gene bodies, 3' untranslated regions (UTR), and close to transcription sites of coding genes (-200 to -1500 bp, 5′ UTR, and exons 1) in tumor cells[190,195,196]. Since DNA methylation is both dynamically and tightly regulated[189], revealing specific methylation patterns associated with tissue differentiation, its assessment can provide information about the tissue of origin[197-199]. In this context, specific methylation markers have been shown to occur more homogenously and prevalently in tumors of the same entity compared with recurrent genomic markers[185].

Both regional and gene-specific hypermethylation is a common characteristic in primary PCa, accumulating during tumor progression[161,191] (Figure 4). The most frequent methylation marker in PCa is hypermethylation associated with the *GSTP1* gene, which occurs in > 80–90% of primary PCa cases[161,185,191]. Furthermore, promoter hypermethylation of additional genes, including *Ras association domain family member 1 (RASSF1)*, *APC, cyclin D2 (CCND2)*, *retinoic acid receptor beta (RARB)*, and *paired like homeodomain 2 (PITX2)*, has been reported in PCa[183] (Figure 4). Massie and colleagues summarized findings from 17 studies investigating aberrant gene methylation in PCa and identified 861 genes that were reported in two or more studies, 168 genes in three or more studies, and 45 genes in four or more studies[191]. Several of these methylation markers have been tested as diagnostic biomarkers for PCa detection or for risk stratification of aggressive disease[183]. Global hypomethylation may co-occur with focal hypermethylation at early stages of PCa (Figure 4), but has also been observed as late event associated with tumor progression and aggressiveness[191,200-203]. Furthermore, DNA methylation patterns may vary depending on the genomic landscape of tumors, revealing distinct molecular subsets of primary PCa characterized by combined genetic and epigenetic signatures[20,161].

1.2 Liquid Biopsy

Liquid biopsy analyses comprise the assessment of molecular analytes in body fluids, primarily in blood (i.e., plasma and serum), or in other sources, such as urine[204,205], saliva[206,207], pleural fluid[208,209], exhaled breath condensate[210], pancreatic juice[211], bile[211,212] and cerebrospinal fluid[213,214]. These assessments have been applied in the context of both physiological and pathological processes[215], with a particular focus on cancer research and diagnostics. It has been shown that tumor-derived markers can be detected in body fluids, and can subsequently be analyzed with regard to different scientific and clinical questions[216,217]. These tumor-associated molecules involve circulating cell-free DNA (cfDNA), cell-free RNA (microRNA (miRNA[218]) and mRNA[219]), mitochondrial DNA[220,221], extracellular vesicles[222,223], proteins, tumor-educated platelets[224,225], and circulating tumor cells (CTCs[226,227]). Histopathological and molecular assessment of tissue biopsies are considered the current gold standard for tumor diagnostics and genotyping. However, tissue samples may be difficult or even impossible to obtain, and tissue biopsy is an invasive procedure associated with an increased risk of adverse events for patients[228]. Needle biopsies of individual regions may fail to capture tumor heterogeneity or may entirely miss tumor cells[229-232]. In contrast, liquid biopsy analyses represent minimally invasive approaches and can complement tissue biopsies in molecular tumor profiling[217,230]. Liquid biopsies harbor the potential to reveal molecular information from different tumor sites, such as primary tumors and metastases, capturing the entire tumor heterogeneity[233-236]. The analyses may provide molecular information in the absence of available tumor tissue, and can be performed as serial assessments over disease or therapy courses[217,237-243].

The presence of cfDNA in the bloodstream was first reported in 1948 by Mandel and Metais[244], but cfDNA analyses have gained increasing interest over the past 10–20 years among researchers worldwide[217,240,245-250]. Comprehensive cfDNA analyses in blood and other body fluids revealed varying or elevated cfDNA concentrations, in both physiological and pathological conditions[206,217,246,251-253]. Molecular cfDNA profiling, and particularly the assessment of circulating tumor-derived DNA (ctDNA) released from cancer cells, has revealed great potential as a diagnostic tool for precision medicine. In this context, various clinical needs have been addressed, including tumor characterization and risk stratification at initial diagnosis, assessment of minimal residual disease, therapy and disease monitoring, as well as management of therapy resistance[217,254]. In addition, liquid biopsy analyses may contribute to an improved understanding of biological mechanisms, including treatment response and (acquired) resistance to cancer therapies[217,254-257]. Serial molecular cfDNA profiling and the identification of changes in ctDNA fractions over disease or therapy course have been shown to provide earlier identification of tumor recurrence or progression, with varying reported lead times, up to several months, compared with clinical detection using conventional imaging techniques[237,258-260]. The proportion of detectable ctDNA in the blood has been shown to vary widely between different tumor entities and tumor stages[250], posing challenges for tumor signal detection in cancers with low ctDNA content. It has been hypothesized that the ctDNA passage is hindered by physiological barriers (i.e., blood-brain barrier, capsules), in several tumor entities, including renal cancer, prostate cancer and brain tumors[250,261]. The ctDNA proportion has been shown to increase during tumor progression, whereas ctDNA concentrations in early-stage diseases remain low, even in those tumor entities which have been associated with high ctDNA content. The ctDNA fraction may be influenced by several parameters, including tumor burden, tumor growth, location and vascularization, cell turnover rate, histology, and tumor aggressiveness, among others[217,262]. CfDNA is rapidly degraded by nucleases and cleared from the circulation by the liver and kidneys, with reported half-lives spanning several minutes to a few hours[261]. Consequently, cfDNA analysis enables the observation of dynamic changes, but also poses specific challenges.

PCa is known to shed low ctDNA amounts into the bloodstream[250]. This is particularly evident in early-stage PCa[263], while patients with advanced PCa may demonstrate increased ctDNA levels[264-266]. It has been suggested that the low plasma ctDNA content observed in early-stage disease may be associated with an intact blood-testis, blood-prostate and blood-epididymis barrier which limits the passage of molecules into the blood stream[264]. The permeability of these barriers may be pathological increased during tumor progression, leading to improved detectability of tumor-associated molecules[264]. Total cfDNA concentrations have been shown to be significantly increased in PCa patients compared with both healthy controls or BPH patients[264,265,267-269].

Urine represents an additional promising source for ctDNA detection, particularly for urological tumor entities such as PCa, due to their anatomical proximity to the urinary tract. Urinary ctDNA has been detected in PCa and other urothelial cancers[204,264,270], but has also been observed in non-urological tumor entities such as non-small cell lung cancer (NSCLC) or glioma[205,271].

1.2.1 Origin and composition of cfDNA in plasma and urine

Comprehensive investigations have explored the origin, the underlying release mechanisms and the biological significance of cfDNA[245,249]. Several mechanisms are known by which cfDNA and ctDNA are shed into the bloodstream[272] (Figure 5). Plasma cfDNA fragments harbor a characteristic length profile centered around 167 base pair (bp) and multiples thereof[246,272]. These small cfDNA fragments with < 200 base pairs (bp) length originate from apoptotic cell death, representing DNA that is wrapped around the nucleosomal core unit (~147 bp) plus a 20 bp linker DNA molecule. Larger fragments (>700 bp) are associated with cells undergoing other forms of cell death (e.g., necrosis)[245,249,273]. Further release mechanisms have equally been discussed, including oncosis, pyroptosis, phagocytosis, neutrophil extracellular trap release (NETosis), excision repair, and active secretion, particularly from tumor cells[272,273]. CtDNA has been shown to harbor shorter fragment lengths compared with non-tumorous cfDNA, suggesting differences in the underlying fragmentation mechanisms[198,274-276].

CfDNA presence has also been investigated in additional body fluids, including urine, but the origin of urinary cfDNA remains less well explored and understood than that of plasma cfDNA. Glomerular filtration represents one clearance pathway of cfDNA, but the glomerular barrier is selective with respect to size and charge of molecules[277]. Consequently, it restricts the passage of longer cfDNA fragments[277] and allows only fragments shorter than 100–250 bp to pass[272,278,279] (Figure 5). During their passage through the urinary tract, cfDNA fragments may be further degraded by enzymatical and mechanical effects[280] (Figure 5). The nuclease-activity in urine has been reported to be 100-fold higher than in blood[281]. Furthermore, it has been suggested that deviating enzymatic DNA degradation in plasma and urine is associated with different nucleotide patterns at starts and ends of plasma and urinary cfDNA fragments [282], and may also influence cfDNA fragment length. The enzyme DNase1-like3 predominantly performs chromatin degradation in plasma, whereas the highly active enzyme DNAse1 in urine preferentially degrades of naked DNA from which DNA-bound proteins have been removed[282]. Longer DNA fragments with heterogeneous sizes > 1 kb may also be present in urine, but predominantly origin from urinary tract cells[277]. In addition to this cfDNA from healthy uroepithelial cells, ctDNA from urological tumor cells may reach the urine through direct shedding. Depending on tumor location as well as its invasiveness and aggressiveness, both transrenal and directly released ctDNA may be detectable in urine samples of patients with urological tumors[283] (Figure 5).

Release mechanisms and nucleosomal packaging also impact the cfDNA composition. Loyfer and colleagues established a comprehensive DNA methylation atlas using cell type-specific methylation patterns and examined the composition of plasma cfDNA in healthy individuals[197,284]. They identified blood cells as the main contributors to cfDNA (Granulocytes: 32%, Megakaryocytes: 27%, Monocytes/Macrophages: 19%, Erythroid progenitors: 6%, Lymphocytes: 4%), alongside vascular endothelial cells (7%), hepatocellular cells (2%), and other sources (2%). Depending on the underlying pathological processes in specific organs or body compartments, these proportions may be altered. In cancer patients, ctDNA released from tumor cells represents an essential contributor to overall cfDNA levels, albeit to a varying extent[250]. The cellular contribution to urinary cfDNA has been shown to be more heterogeneous and to exhibit higher interpatient variability[270]. Methylation-based deconvolution analysis of urinary cfDNA from both bladder cancer patients and controls identified blood cells as essential contributors to cfDNA (mean: 29%[270]). However, increased proportions of cfDNA derived from kidney and urothelial cells were also observed in urine[270]. Tumor patients and non-tumorous controls have been shown to harbor similar distributions of cellular contributors to urinary cfDNA, albeit with interindividual variability[270]. Tumor-derived cfDNA contribution was significantly increased in cancer patients, with a trend toward higher levels in aggressive disease[270]. These findings were partially confirmed by a study in metastatic PCa, demonstrating that the majority of methylation signals in plasma and urine derived originated from non-tumor sources, with kidney and bladder cells contributing to the non-tumorous urinary DNA[285]. Prostate content was also detectable and showed a higher contribution in urine (sediment) than in plasma[285].

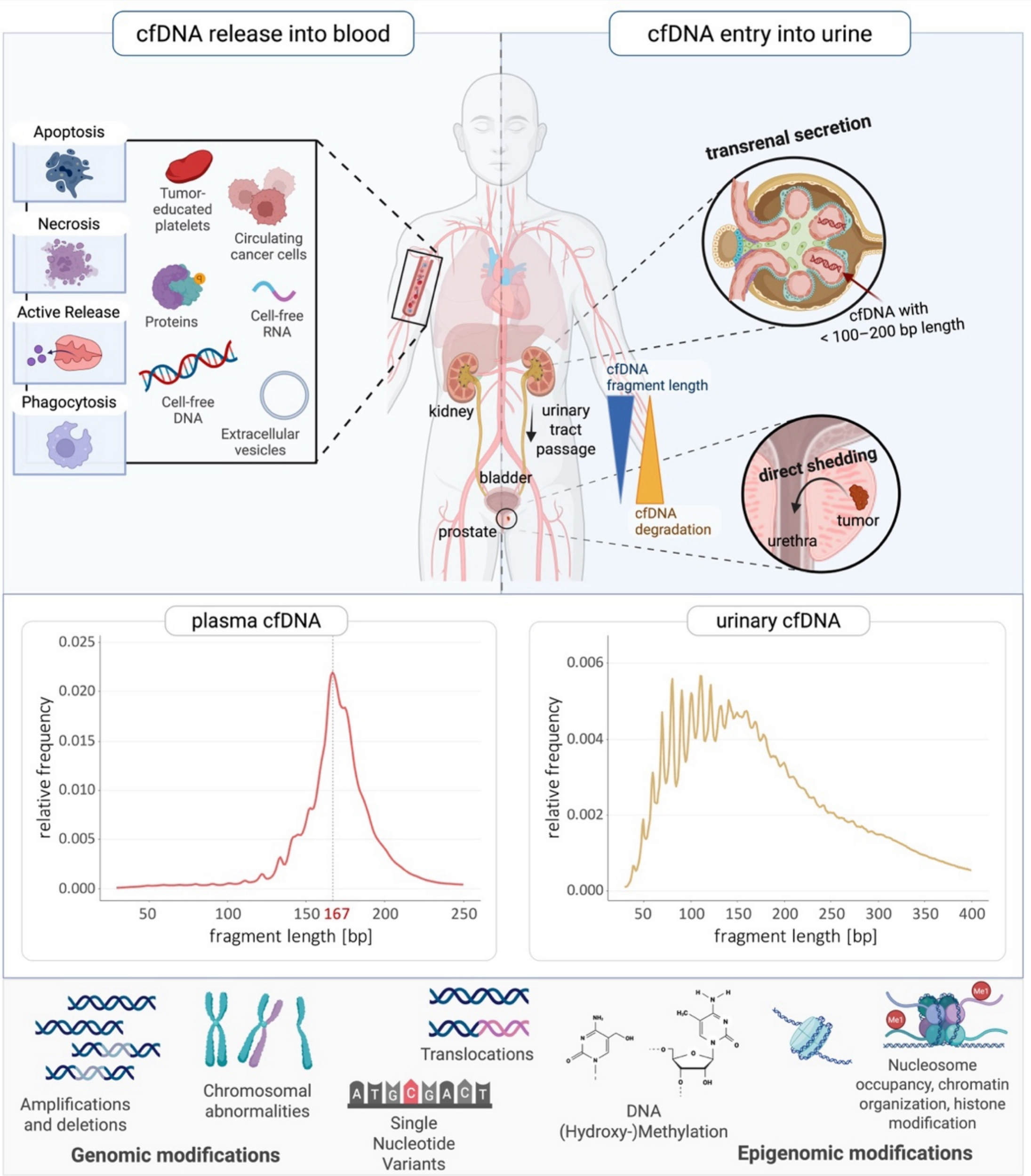


Figure 5: Biological origins and characteristics of plasma and urinary cfDNA.
Schematic overview of the underlying mechanisms for cfDNA release into the blood circulation and the entry of cfDNA into urine, including characteristic plasma and urinary cfDNA fragmentation profiles. Molecular profiling of cfDNA enables the detection of genomic and epigenomic alterations indicative for the presence of ctDNA. Abbreviations: bp = base pair, cfDNA = cell-free DNA, ctDNA = circulating tumor-derived DNA. Figure created with BioRender.com.

1.2.2 Tumor-associated genomic alterations in cfDNA

The genomic assessment of cfDNA is a well-established component of liquid biopsy analyses. The detection of genomic alterations in the cfDNA, such as single nucleotide variants (SNVs), gene arrangements and CNVs indicates the presence of ctDNA[247] (Figure 5). Accordingly, genomic cfDNA analyses have been used to estimate the proportion of ctDNA in total cfDNA, as well as to evaluate tumor burden[247,272,286,287]. Furthermore, the identification of entity-specific, actionable or otherwise indicative genomic alterations may support molecular characterization and risk stratification of tumors[217,288]. These analyses can be utilized for diagnostic applications, and dynamic changes can be monitored to track therapy response and disease progression[217]. The greatest potential is currently seen for advanced-stage disease or tumor entities characterized by increased ctDNA release[238,239,250,289]. However, the implementation for early-stage disease and tumor entities with low ctDNA abundance or a limited number of recurrently altered genes, as observed in (localized) PCa[263], remains challenging. Technological advances with improved sensitivity and specificity for ctDNA detection have emerged to address the challenge of low ctDNA abundance against a dominant non-tumorous background[247]. Different platforms have achieved detection limits for mutation allele frequencies below 10%, with some technologies demonstrating ctDNA detection sensitivities as low as 1–10 parts per 10^5–10^6 [241,247,290-296]. These advances have established suppression of background errors and/or enrichment of tumor-derived signals as key elements for improved sensitivity[241,290,292]. However, the achievable technical sensitivity is ultimately limited by biological factors, such as the frequency of mutant cfDNA fragments present in the liquid biopsy samples[297]. Targeted approaches enable the sensitive detection of single or multiple known tumor-specific alterations, even at very low variant allele frequencies and with low (cf)DNA input[217,243,298]. These approaches include digital polymerase chain reaction (dPCR) and target enrichment methods, such as hybridization capture and amplicon sequencing[217]. In addition to universally applicable, commercially available or custom-designed panels, personalized and patient-individual assays have demonstrated high sensitivities for ctDNA detection, especially in the context of minimal residual disease or tumor progression over the therapy course[291,297,299,300]. While targeted approaches allow for deeper sequencing of specific regions of interest, they potentially fail to detect novel or unexpected genetic alterations[217,298]. In contrast, whole-genome sequencing (WGS) and whole-exome sequencing (WES) are untargeted approaches, providing a comprehensive insight into the entire genome or exome, respectively, without prior tumor information[217]. This enables the identification of a wide range of genetic alterations and the discovery of novel changes[217] which can be applied for initial tumor characterization as well as for the evaluation of genomic alterations arising during tumor progression or therapy resistance. Untargeted approaches are typically associated with lower sensitivity for ctDNA detection and reduced sequencing depths at specific genomic regions or genes, while requiring higher cfDNA input and higher sequencing costs[298]. Low-coverage WGS (lcWGS), performed at 1–2x coverage and also referred to as “shallow WGS”, represents a cost-efficient alternative for genome-wide molecular profiling, requiring only low (cf)DNA input amounts (down to ~1 ng)[238,286]. This approach has been applied to assess CNVs and chromosomal instability in cfDNA, thereby enabling ctDNA detection particularly in advanced tumor stages and metastatic disease[238,286,301]. Depending on the tumor entity, CNVs may also be detectable in ctDNA by lcWGS at earlier disease stages[204]. In addition, lcWGS data has been used for the comprehensive analysis of cfDNA fragmentation[205,274], and the potential of lcWGS for identifying tumor-specific somatic mutations or mutational signatures in plasma cfDNA has been explored, with machine-learning algorithms applied to enhance ctDNA detection[302-304].

LcWGS has also been applied to plasma cfDNA from patients with hormone-sensitive PCa and CRPC, and multiple CNVs were identified in both disease groups, including known alterations from PCa tissue analyses[305]. In addition, whole-genome or targeted cfDNA sequencing have identified recurrent genomic alterations (i.e., SNVs, gains and losses) in driver genes, such as *AR*, *BRCA1/2*, *TP53*, *SPOP*, *ATM*, *APC*, *checkpoint kinase 2 (CHEK2)*, *PTEN*, *MYC, RB1, phosphoinositide-3-kinase regulatory subunit 1 (PIK3R1), CDKN1B,* as well as the *TPRSS2-ERG* gene fusion[183,306,307]. However, the identification of tumor-specific alterations in cfDNA has been shown to remain difficult, even in metastatic PCa, largely due to low ctDNA fractions[250,307]. Characterization of recurrent genomic alterations in localized PCa is even more challenging[263]. Point mutations and few deletions are characteristic for localized PCa, whereas amplifications predominantly emerge during disease progression[308]. Other key drivers, such as *AR* aberrations, are mainly associated with castration resistance during hormone treatment[305,309]. To address these challenges, various approaches have been explored for ctDNA detection in localized PCa, including the use of lcWGS for CNV analysis and sensitive targeted panel sequencing for somatic mutation detection[263,310].

1.2.3 Tumor-associated epigenomic modifications in cfDNA

Following the successful implementation of epigenomic analyses in tumor tissue and increasing recognition of the essential role of epigenetics in tumorigenesis, profiling of the cell-free epigenome in cancer patients has emerged as alternative or complementary approach to genomic cfDNA analyses. The exploration of tumor-informative epigenomic cfDNA markers has gained particular interest in the pursuit of developing advanced approaches for more sensitive and specific ctDNA detection, especially for early tumor stages or for cancer entities with a low genomic burden. Aberrant DNA methylation is known to occur early during carcinogenesis and, in some cases, precedes the emergence of genomic events[185,191,192,288]. This characteristic has made DNA methylation an appealing target for liquid biopsy analyses, with the aim of developing enhanced technologies for early detection and risk stratification.

Epigenomic cfDNA profiling involves the assessment of cfDNA methylation[197,311,312] and hydroxymethylation[313-315], as well as various aspects of cfDNA fragmentation (i.e., genome-wide fragmentation and fragment lengths[205,213,274,316], fragment ends and recurrent sequence motifs[317-319], nucleosome occupancy and chromatin organization[288,320,321]; Figure 5). It has been investigated that cfDNA fragmentation follows a non-random process, influenced by release mechanisms and nucleosomal packaging of cfDNA[245,318,322]. Pathological conditions may impact these mechanisms, leading to aberrant cfDNA fragmentation patterns compared with the physiological state[288]. Consequently, assessment of cfDNA fragmentation may extend insights into underlying biological processes and may support the identification of differences between ctDNA and healthy cfDNA for tumor characterization[254,288]. Tumor-derived DNA fragments have been shown to exhibit shorter length profiles than non-tumorous cfDNA, and this characteristic has been applied for the enrichment and detection of tumor signals in liquid biopsy samples[205,274,275,316,323,324]. Genome-wide fragmentation features have been incorporated into machine-learning models, enabling sensitive tumor detection across multiple cancer types in plasma[204,274,316], as well as in other body fluids[213]. Accordingly, cfDNA fragmentation analyses have also enabled the discrimination between cancer patients and healthy individuals without prior knowledge of genomic alterations[274,288,316]. The combination of cfDNA fragmentation with genomic cfDNA features, such as CNVs or mutations, has further improved sensitivity for tumor classification[204,274,316].

Beyond fragment length, cfDNA fragmentation analyses have also focused on DNA fragment end characteristics, including their genomic location relative to the nucleosomal phasing[318], their sequence composition at the 5′-end[325-327], as well as single-stranded 5′ protruding (“jagged”) ends[254,319,328]. CfDNA fragmentation patterns have been successfully investigated using lcWGS data, providing a straightforward, cost-effective approach[204,213,274]. Similarly, chromatin accessibility inferred from nucleosomal occupancy as assessed by lcWGS enabled tumor detection and classification from cfDNA[329]. Nucleosomal positioning and cfDNA fragmentation have also been shown to be informative for the determination of the cell or tissue of origin[316,320,321,330]. Likewise, methylome profiling has been applied for cfDNA deconvolution to determine the contributors to cfDNA in liquid biopsy samples from both cancer-free individuals and cancer patients[197-199].

However, the application of available methylation analysis techniques on cfDNA has proven to be challenging due to its high fragmentation and frequently low concentration[331-333]. Several methods have been established for sensitive cfDNA methylation analysis, including targeted and whole-genome approaches, analogous to genomic cfDNA analyses[254,331,332,334]. Genome-wide or targeted bisulfite sequencing represent well-established methods for methylome profiling, enabling the assessment of DNA methylation with single-base resolution[332]. However, bisulfite conversion induces additional degradation and reduced recovery from already limited amounts of fragmented cfDNA, and may result in an uneven representation of thymine and cytosine bases[331,332]. Genome-wide methylome profiling with whole-genome bisulfite sequencing (WGBS) requires high DNA input and high sequencing depth, leading to elevated sequencing costs[332]. To overcome these limitations, enzymatic conversion-based alternatives to bisulfite conversion have been developed, such as TET-assisted pyridine borane sequencing[335] and enzymatic methyl-sequencing[336,337], and have been successfully applied to cfDNA methylation analysis[332]. Genome-wide methylation profiling has also been performed using affinity-based capture methods, which enrich methylated DNA fragments through the use of either methylation-specific monoclonal antibodies (methylation DNA immunoprecipitation, MeDIP[333,338]) or methyl-CpG binding domain- (MBD-) harboring proteins[339-341] that selectively bind to DNA fragments with all methylated cytosines or only methylated CpGs, respectively. Enrichment-based methylome profiling has shown high potential for tumor detection and characterization across different tumor entities and disease stages, and also for the classification of cancers of unknown primary[342-345]. Shen and colleagues published a cell-free MeDIP sequencing (cfMeDIP-seq[333]) protocol as advancement of the existing MeDIP method for low DNA input amounts <10 ng[312,332,333], enabling a cost-effective, genome-wide assessment of methylated regions with a resolution of 100–200 bp. The successful application of cfMeDIP-seq for cfDNA methylation profiling has been demonstrated across multiple tumor entities, including early tumor stages[237,312,344-346].

CfDNA methylation analysis has also been investigated in the context of PCa, and established methylation markers identified in PCa tissue (e.g., hypermethylation in *GSTP1* or *APC*) have been detected in plasma or urine samples from PCa patients[183,347-351]. Most cfDNA methylation studies in PCa have focused on targeted approaches, frequently using PCR-based techniques to assess methylation panels comprising few markers or genes[347-351]. More recently, a multi-cancer early detection test has been developed that targets several thousand methylation markers, previously identified with WGBS, and integrates them into a machine-learning model for tumor signal detection in plasma cfDNA across multiple cancer types[352,353]. High overall sensitivities for cancer signal detection were achieved in advanced tumor stages (stage III-IV: 83.9%, stage IV: 90.1%) at high specificity (99.5%), whereas sensitivity was decreased for earlier stages (stage I-II: 27.5%) and remained low in all PCa cases (11.2%, 47/420[352]). These and other findings have highlighted the successful, but still limited detectability of methylation signals in liquid biopsy samples from PCa patients, with improved performance in advanced tumor stages, but persistent challenges in localized PCa[183]. Two studies have applied enrichment-based methylation profiling in plasma cfDNA from PCa patients with cfMeDIP-seq[354] and MBD-based capture and sequencing (methyl-CpG binding domain sequencing, MBD-seq) [355], respectively. These two studies determined differentially methylated regions (DMRs) between metastatic PCa compared with either localized PCa patients or cancer-free controls, respectively, and incorporated the resulting tumor-informative methylation markers into machine-learning classifiers for the detection of metastatic and localized PCa, as well as for the discrimination of the tumor stages[354,355]. These findings have provided insights into the underlying biology of epigenetic changes in localized and advanced PCa, highlighting the potential of cfDNA methylome profiling for earlier disease stages and diagnostic applications[354,355].

1.3 Aim of this PhD Project

The molecular and clinical heterogeneity of PCa is not sufficiently reflected by current diagnostic options. Hence, novel molecular biomarkers and diagnostic algorithms are needed for personalized and patient-oriented clinical decision-making and PCa risk stratification. Liquid biopsy analyses enable a comprehensive view on the tumor disease and its heterogeneity, serving as a promising diagnostic tool for minimally invasive tumor characterization. Previous studies investigating genomic and epigenomic cfDNA profiles in PCa revealed challenges, including low ctDNA shedding and limited burden of genomic alterations, particularly in early-stage PCa. Epigenomic changes are known to occur earlier in tumorigenesis and their assessment might harbor potential for early tumor characterization and for improved ctDNA detection in cases with low (genomic) tumor signals.

My PhD project aimed to establish a multimodal liquid biopsy approach to improve ctDNA detection in primary PCa, thereby supporting the identification of localized PCa and risk stratification of aggressive disease at initial diagnosis. To address the current challenges of limited ctDNA shedding and low numbers of genomic alterations in PCa, I integrated genomic and epigenomic cfDNA analyses in matched plasma and urine samples. This multimodal strategy combined complementary data modalities and two liquid biopsy sources to enable improved ctDNA detection and comprehensive tumor characterization in heterogeneous PCa. For this purpose, I applied cfMeDIP-seq and lcWGS on plasma and urinary cfDNA of PCa patients and cancer-free controls, in order to evaluate genome-wide methylation patterns, chromosomal instability, CNVs, as well as cfDNA fragmentation features.

The first objective of this study was to identify genomic and epigenomic cfDNA characteristics in plasma and urine and to assess their complementarity between the two liquid biopsy sources. The second objective was to evaluated whether plasma and urinary cfDNA features revealed differences between tumor and non-tumorous samples, consequently serving as informative biomarkers for ctDNA detection. In this context, I evaluated whether specific genomic and epigenomic cfDNA features were characteristic of primary PCa by comparing findings in liquid biopsy samples with genomic and epigenomic data obtained from PCa tissue and matched buffy coat samples. In addition, these results were validated, using publicly available PCa tissue datasets. The third objective of this study was to determine whether the complementary evaluation of genomic and epigenomic cfDNA features in matched plasma and urine improved ctDNA detection compared to the analyses based on single parameters or single liquid biopsy sources. Finally, the study aimed to associate the genomic and epigenomic cfDNA analyses with clinical characteristics of the PCa patients, such as tumor stage or initial PSA levels, in order to assess the potential of multimodal liquid biopsy analyses as a complementary tool in PCa diagnostics.

2 Materials and Methods

2.1 Materials

Table 5: Equipment

Name	Manufacturer
Capillary electrophoresis system, Fragment Analyzer 5200	Agilent Technologies, Santa Clara, USA
Centrifuge 5415 R	Eppendorf, Hamburg, Germany
Centrifuge 5810 R	Eppendorf, Hamburg, Germany
Centrifuge Heraeus Megafuge 16 R	ThermoFisher Scientific, Waltham MA, USA
Centrifuge Sigma 4-16	Sigma Laborzentrifugen GmbH, Osterode am Harz, Germany
DynaMag™-2, Magnetic rack for 2 ml tubes	ThermoFisher Scientific, Waltham MA, USA
Fluorometer, Qubit 2.0	ThermoFisher Scientific, Waltham MA, USA
Focused-ultrasonicator M220	Covaris, Woburn MA, USA
Heating bock, Thermomixer comfort	Eppendorf, Hamburg, Germany
LightCycler® 480	Roche, Mannheim, Germany
Magnetic separator, 0.2 mL PCR Strip	Permagen Labware, Peabody, USA
Magnetic rack, 5 mL tubes	Ariumlab, Tallinn, Estonia
Magnetic rack, 15 mL and 25 mL conical tubes	Ariumlab, Tallinn, Estonia
Microcentrifuge, Heraeus Fresco 21	ThermoFisher Scientific, Waltham MA, USA
Micropipettes, Pipetman (10, 20, 100, 200, 1000 µL)	Gilson, Middleton, USA
Minicentrifuge,Spectrafuge 3-1810	neoLab Migge, Heidelberg, Germany
Multichannel pipet, Rainin pipet lite L-10 (1-10 µL)	Mettler-Toledo, Columbus, USA
Multichannel pipet, Rainin pipet lite L-200 (20-200 µL)	Mettler-Toledo, Columbus, USA
Nanodrop 1000 Spectrophotometer	ThermoFisher Scientific, Waltham MA, USA
Picus NxT (120µl), Electronic Pipette, Single Channel	Sartorius AG, Göttingen, Germany
Pipetboy acu (2)	Integra Biosciences, Biebertal, Germany
Thermocycler, T100™	Bio-Rad Laboratories, Hercules, USA
Thermomixer comfort	Eppendorf, Hamburg, Germany
TissueLyser II	Qiagen, Hilden, Germany
Tube rotator, Rotobot	Benchmark Scientific, Edison NJ, USA
Vortexer, Vortex-Genie 2	Scientific Industries, Bohemia, USA

Table 6: Consumables

Name	Manufacturer
Adhesive Clear PCR Seal	Biozym Scientific, Hessisch Oldendorf, Germany
Conical centrifuge tubes 15 mL	TPP Techno Plastic Products, Trasadingen, Switzerland
Conical centrifuge tubes, CELLSTAR® (15 mL, 50 mL)	Greiner Bio-one, Frickenhausen, Germany
Conical Centrifuge Tubes, Corning™ Falcon™, 15mL	ThermoFisher Scientific, Waltham MA, USA

Conical tubes (25mL)	Eppendorf, Hamburg, Germany
Filter tips, ClipTip™ filtered 384-Format Pipette Tips	ThermoFisher Scientific, Waltham MA, USA
Filter tips (10 µL, 20 µL, 100 µL, 200 µL, 1000 µL)	Neptune, San Diego, USA
Filter tips, SafeSeal Professional Tips (10 µL, 20 µL, 100 µL, 200 µL, 1250 µL)	Biozym Scientific, Hessisch Oldendorf, Germany
Filter tips, SafeSeal SurPhob Tips (20 µL)	Biozym Scientific, Hessisch Oldendorf, Germany
Filter tips for Rainin LTS pipets, RT-L10F (20µl)	Mettler-Toledo, Columbus OH, USA
Filter tips for Rainin LTS pipets, RT-L250FLR (200µl)	Mettler-Toledo, Columbus PH, USA
Filter tips, ZAP Slik Typ Rainin LTS (20µl)	Labcon, Petaluma CA, USA
Gloves, StarGuard Comfort	Starlab, Hamburg, Germany
LightCycler® 480 Multiwell Plate 384, white	Roche, Mannheim, Germany
LightCycler® 480 Sealing Foil	Roche, Mannheim, Germany
Microcentrifuge tubes (0.5 mL, 1.5 mL, 2.0 mL, 5mL)	Eppendorf, Hamburg, Germany
microTUBE AFA Fiber Pre-Slit Snap-Cap 6x16mm (130µl)	Covaris, Woburn MA, USA
PCR plate 96, semi-skirted	Eppendorf, Hamburg, Germany
PCR strips (0.2 mL)	Eppendorf, Hamburg, Germany
PCR strips (0.2 mL), StarPCR Tubes	Starlab, Hamburg, Germany
Qubit Assay Tubes	ThermoFisher Scientific, Waltham MA, USA
Reagent Reservoirs, Corning® Costar® 4870, 50ml	ThermoFisher Scientific, Waltham MA, USA
Reagent Reservoirs, 10ml, Polystyrene	Integra Biosciences, Biebertal, Germany
Sample beakers, LABSOLUTE®, 200ml, PP, with screw cap, sterile	Th. Geyer, Renningen, Germany
Serological pipettes, CELLSTAR® (5 mL, 10 mL, 25 mL, 50mL)	Greiner Bio-one, Frickenhausen, Germany
Pasteur Pipettes, 3ml, PE, sterile	Th. Geyer, Renningen, Germany
S-Monovette 9 mL K3 EDTA tubes	Sarstedt, Nürnberg, Germany

Table 7: Molecular biology kits

Kit	Manufacturer
AllPrep DNA/RNA/Protein Mini Kit	Qiagen, Hilden, Germany
Ambion® Buffer Kit	ThermoFisher Scientific, Waltham MA, USA
HS NGS fragment Kit (1-6000bp)	Agilent Technologies, Santa Clara, USA
iPure Kit v2 x100	Diagenode, Seraing, Belgium
KAPA HiFi HS Library Amplification Kit	Roche, Mannheim, Germany
KAPA HyperPrep Kit	Roche, Mannheim, Germany
MagMeDIP qPCR Kit x48	Diagenode, Seraing, Belgium
NEBNext ® Multiplex Oligos for Illumina ® (Unique Dual Index UMI Adaptors DNA Set 1)	New England BioLabs, Ipswich MA, USA
NGS fragment Kit (1-6000bp)	Agilent Technologies, Santa Clara, USA
QIAamp ® DNA Mini and Blood Mini Kit	Qiagen, Hilden, Germany
QIAamp ® MinElute ccfDNA Kit	Qiagen, Hilden, Germany
Qubit dsDNA HS Assay Kit	ThermoFisher Scientific, Waltham MA, USA
Qubit dsDNA BR Assay Kit	ThermoFisher Scientific, Waltham MA, USA

Table 8: Reagents

Reagent	Manufacturer
Ambion nuclease-free water	ThermoFisher Scientific, Waltham MA, USA
AMPure XP magnetic beads	Beckman Coulter, Brea, USA
Dynabeads™ M270	ThermoFisher Scientific, Waltham MA, USA
EDTA (0.5M, pH 8)	ThermoFisher Scientific, Waltham MA, USA
Ethanol absolute	Sigma-Aldrich, St. Louis, USA
Gibco™ Dulbecco's Phosphate-Buffered Saline (without calcium and magnesium)	ThermoFisher Scientific, Waltham MA, USA
Isopropyl alcohol	Sigma-Aldrich, St. Louis, USA
Lambda-DNA	ThermoFisher Scientific, Waltham MA, USA
PrimaQuant CYBR BLUE 2x qPCR SYPR Green master mix	Steinbrenner, Wiesenbach, Germany
Stainless steel beads for TissueLyser systems	Qiagen, Hilden, Germany

Table 9: Software

Software	Company
ChatGPT-4o	OpenAI, San Francisco CA, USA
Elicit	Elicit[356]
LightCycler® 480 Software	Roche, Mannheim, Germany
Microsoft Excel for Mac, Version 16.78.3	Microsoft, Redmond, USA
Perplexity AI	Perplexity AI, San Francisco CA, USA
ProSize Data Analysis Software 4.0.1.4	Agilent Technologies, Santa Clara, USA
RStudio 2023.12.0+369.pro3 "Ocean Storm" Release (2023-12-14) for CentOS 7	R Studio, Boston, USA

Table 10: Bioinformatic tools

Tool	Reference
Bedtools v2.30.0	Quinlan, A. and Hall, I. [357]
Bowtie2 v2.4.4	Langmead, B. and Salzberg, S. [358]
Cutadapt v3.4	Martin, M. [359]
FastQC v0.11.9	Andrews, S. [360]
HMMcopy Utils v0.1.1	Lai, D. *et al.* [361,362]
MultiQC v1.12	Ewels, P. *et al.* [363]
Nextflow v22.07.1-edge	Di Tommaso, P. *et al.* [364]
Picard v2.27.2	Broad Institute [365]
SAMtools v1.15.1	Li, H. *et al.* [366]
Singularity v.1.2.5-1.el7	Kurtzer, G. M. *et al.* [367]
R v.4.0.0	R Core Team [368]
UMI-tools v1.1.2	Smith, T. *et al.* [369]

Table 11: R packages

Tool	Reference
boot v.1.3.28	Davison, A. C. and Hinkley, D. V. [370]
ChIPpeakAnno v.3.24.2	Zhu, L. J. *et al.* [371]
data.table v.1.14.4	Srinivasan, A. and Dowle, M. [372]
dplyr v.1.1.1	Wickham, H. *et al.* [373]
GenomicRanges v.1.42.0	Lawrence, M. *et al.* [374]
ggcorrplot v.0.1.4.1	Kassambara, A. [375]
ggfortify v.0.4.17	Tang, Y. *et al.* [376]
ggplot2 v3.4.2	Wickham, H. [377]
ggpubr v.0.6.0	Kassambara, A. [378]
HMMcopy v1.32.0	Lai, D. *et al.* [361]
liver v.1.15	Mohammadi,R. and Burke, K. [379]
lmtest v.0.9.40	Zeileis, A. and Hothorn, T. [380]
matrixStats v.0.61.0	Bengtsson, H. [381]
MEDIPS v.1.46.0	Lienhard, M. *et al.* [382]
MESA v0.2.2	Chemi, F. *et al.*, Conway, A.-M. *et al.* [342,343]
MLmetrics v.1.1.3	Yan, Y. [383]
pheatmap v.1.0.12	Kolde, R. [384]
plotly v.4.10.2	Sievert, C. [385]
plyranges v1.10.0	Lee, S. *et al.* [386]
pROC v.1.18.2	Robin, X. *et al.* [387]
PRROC v.1.3.1	Grau, J. *et al.* [388]
QSEA v.1.16.0	Lienhard, M. *et al.* [389]
RColorBrewer v.1.1-3	Neuwirth, E. [390]
readr v.2.1.2	Wickham, H. *et al.* [391]
readxl v.1.3.1	Wickham, H. and Bryan, J. [392]
regioneR v.1.22.0	Gel, B. *et al.* [393]
rstatix v0.7.2	Kassambara, A. [394]
stats v.4.0.0	R Core Team [368]
tidyverse v.1.3.1	Wickham, H. *et al.* [395]
zoo v.1.8-12	Zeileis, A. and Grothendieck, G. [396]

2.2 Methods

2.2.1 Study population

A total of 72 PCa patients and 36 cancer-free controls were recruited at the Department of Urology at Heidelberg University Hospital between June 2021 and November 2022. The recruitment was performed by a study nurse of this department. All individuals provided informed consent and the study was approved by the ethic committee of the Medical Faculty Heidelberg (S-130/2021). PCa patients were recruited at time of initial tumor diagnosis and underwent prostate biopsy and/or radical prostatectomy at the University Hospital Heidelberg. Matched blood and urine samples were collected for overall 102 of 109 men, whereas only plasma was available for the remaining seven men. For 47 PCa patients, liquid biopsy samples were collected prior to prostate biopsy which confirmed the PCa diagnosis. One patient received two prostate biopsies within 12 months, and liquid biopsy samples were collected at both timepoints. The first biopsy revealed the diagnosis of a prostate tumor with Gleason 6 pattern, whereas the second biopsy resulted in the upgrade to csPCa (Gleason 7b). For 25 PCa patients, liquid biopsy samples were collected few weeks after the initial PCa diagnosis based on an internal or external prostate biopsy, shortly prior to radical prostatectomy performed at Heidelberg University Hospital. In addition to liquid biopsies, PCa tissue samples were available for eight patients who underwent radical prostatectomy at Heidelberg University Hospital. The control cohort included 16 men who also received a prostate biopsy at Heidelberg University Hospital without detection of any signs for malignancy, and liquid biopsy samples were collected prior to the biopsy. In addition, plasma and urine samples were obtained from 20 men with a PSA level <2.5 ng/ml, who subsequently underwent different examinations or surgeries due to benign urological diseases at the Department of Urology at Heidelberg University Hospital. Clinical data was prospectively collected for all PCa patients and cancer-free controls. Further information regarding family history (first-degree relatives with PCa), lifestyle factors (diet, alcohol consumption, smoking status, and physical examination), and other potential PCa risk factors (infertility, sterilization occurrence of urinary tract infection or urinary retention within the last 6 months) were assessed via questionnaires. MpMRI of the prostate was either performed at the radiological department of the German Cancer Research Center (DKFZ) Heidelberg or at external institutions. MpMRI was assessed following the PI-RADS recommendations[89]. Prostate tissue biopsies were performed as transperineal fusion-targeted biopsy of PIRADS 3–5 lesions identified in the prostate MRI, with elastic software registration using a UroNav system (Philips Invivo, Gainsville, FL, USA), in addition to systematic saturation biopsy adjusted to prostate volume, as previously described by Görtz *et al.* and Radtke *et al.* [84,397]. Histopathological assessments of the prostate tissue biopsies were performed at the Institute of Pathology of Heidelberg University Hospital, according to International Society of Urological Pathology standards[109].

2.2.2 Liquid biopsy and PCa tissue sample collection

Liquid biopsy samples were collected for all PCa patients and cancer-free controls. Sample collection and subsequent preprocessing were performed by a study nurse at the Department of Urology at Heidelberg University Hospital. In addition, PCa tissue samples were available for eight patients who underwent radical prostatectomy. Blood samples were collected via cubital puncture with 20- to 23-gauge butterfly needle between 8 and 11 am in fasting patients. Two 9 ml blood tubes (Sarstedt S-Monovette K3 EDTA) were collected for biobanking and molecular profiling, in addition to regular blood taking for routine laboratory assessments. Additionally, men were asked to collect 30–50 ml midstream urine in 200 ml sterile sample beakers in the same time period during their hospital visit. One part of the urine volume was used for routine assessments (urine status, urine culture). The other part was separated into one or two 15 ml falcons, containing each 200 µl of 0.5 M ethylenediaminetetraacetic acid (EDTA; pH 8.0) for stabilization, and were used for biobanking and molecular profiling. After collection, blood and urine samples were immediately placed on 4°C until further processing within maximum 4–6 hours. Blood samples were centrifuged (Centrifuge 5810 R, Eppendorf) at 1,600 x g for 10 min at 4°C. Subsequently, the two plasma aliquots (3.2–4.8 ml each) were distributed among multiple 2 ml microcentrifuge tubes (Eppendorf), and again centrifuged (Centrifuge 5415 R, Eppendorf) at 14,000 x g for 10 min at 4°C to remove cell debris. The buffy coats obtained from the two blood samples were also transferred to 2 ml microcentrifuge tubes (Eppendorf), and kept for further biobanking. Urine samples were centrifuged (Centrifuge 5810 R, Eppendorf) at 750 x g for 10 min at 4°C. Subsequently, the two urine supernatant aliquots (10 ml each) were transferred into two 15 ml falcons, and again centrifuged (Centrifuge 5810 R, Eppendorf) at 2,600 x g for 10 min at 4°C to remove cell debris. Remaining urine cell pellets from two urine samples were each resuspended in 5 ml Dulbecco's phosphate-buffered saline (DPBS), and the cell suspensions were combined in one 15 ml falcon. This sample was centrifuged (Centrifuge 5810 R, Eppendorf) at 2,600 x g for 10 min at 4°C, the DPBS supernatant was removed, and the urine cell pellet was dried. All centrifugation steps for blood and urine samples, respectively, were performed without brake. Plasma, buffy coat, urine supernatant and urine cell pellet samples were stored at -80°C until further processing. Subsequently, I received the plasma and urine samples and continued with the downstream processing for the molecular analyses.

The PCa tissue samples were extracted intraoperatively during radical prostatectomy. Single embedded slides containing each ~80 mg of tissue were cryopreserved with liquid nitrogen and were stored in 2 ml PCR-clean, safe-lock-tubes (Eppendorf) at -80°C in the Tissue Bank of the National Center for Tumor Diseases (NCT) Heidelberg until further processing. Tissue samples were stored and provided upon request for molecular analyses of genomic DNA (gDNA), in accordance with the regulations of the NCT Tissue Bank and the approval of the ethics committee of Heidelberg University. Histopathological assessment of the PCa tissue slides was performed at the Institute of Pathology of Heidelberg University Hospital. Tissue quality was reported on a scale of 1–5, where 1 represented ideal tissue quality (tissue could be used without restriction) and 5 corresponded to inappropriate tissue quality. In addition, information on the overall tumor cell content, as well as the contribution of viable tumor cells, non-viable tumor cells, and stroma cells, was provided for the bulk tissue samples.

2.2.3 Isolation of cfDNA from plasma and urine samples

I performed cfDNA isolation from 1–5.5 ml of plasma and 9.5–19 ml of urine using the QIAamp MinElute ccfDNA Kit. Frozen plasma and urine samples were thawed on ice, and centrifuged for 2 min at 1,500 x g at room temperature (Heraeus Fresco 21 Centrifuge (plasma) or Heraeus Megafuge 16 R Centrifuge (urine), Thermo Fisher Scientific) to remove remaining cell debris in the supernatant. All subsequent centrifugation steps were performed with the Heraeus Fresco 21 Centrifuge (Thermo Fisher Scientific) at room temperature. Initial preparation of the kit's reagents and the cfDNA isolation were performed according to the manufacturer's instructions with few adaptions, as explained in the following. Briefly, proteinase K (55 µl x *sample volume [ml]*), magnetic bead suspension (30 µl x *sample volume [ml]*) and bead binding buffer (150 µl x *sample volume [ml]*) were added to the samples, in order to bind cfDNA to the magnetic beads. The volumes of proteinase K, magnetic bead suspension and bead binding buffer were adjusted to the initial sample volume (in ml), with 9.5 ml as maximum volume for urine samples. Depending on the final volume of the mixture, 5 ml or 25 ml tubes were used. After 10 min incubation at room temperature with end-over-end rotation, samples were placed on a magnetic rack for precipitation of bead-bound cfDNA. The bead pellets were eluted in 200 µl bead elution buffer (urine: 350 µl bead elution buffer), and incubated for 5 min at room temperature on a tube shaker with 300 rpm to release the cfDNA from the beads. Subsequently, empty beads were captured with a magnetic rack, and the supernatant containing the cfDNA was mixed with 300 µl Buffer ACB (urine: 400 µl ACB buffer). The mixture was placed on a QIAamp UCP MinElute column and centrifuged for 1 min at 6000 x g to bind the cfDNA on the silica membrane and discard the flow-through. One washing step was performed with 500 µl Buffer ACW2 and centrifugation for 1 min at 6000 x g, followed by a second centrifugation at 20,000 x g at full speed for 3 min to remove any remaining fluid and dry the silica membrane of the spin column. The cfDNA was eluted in 30–50 µl of nuclease-free water by centrifugation at 20,000 x g at full speed for 1 min, and stored at -20°C until further processing.

2.2.4 Isolation of gDNA from fresh-frozen PCa tissue samples

Genomic DNA was extracted from eight fresh-frozen PCa tissue samples using the AllPrep DNA/RNA/Protein Mini Kit (Qiagen). The kit was applied for simultaneous purification of DNA, RNA, and proteins from cells and tissues. Initial preparation of the kit's reagents and performance of the workflow for DNA, RNA and protein purification were conducted according to the manufacturer's instructions. Fresh-frozen tissue samples (~80 mg) were placed on dry ice and dissected into three parts, each of which was transferred into a 2 ml microcentrifuge tube containing each one stainless steel bead. Then, 350 µl Buffer RLT was added to each tube at room temperature, and tubes were placed in the TissueLyser II System (Qiagen) for 30 s at 20 Hz frequency to achieve disruption and homogenization. The TissueLyser was applied according to the manufacturer's protocol. Subsequently, the lysate was centrifuged for 3 min at full speed, and the supernatant was transferred to an AllPrep DNA spin column placed in a 2 ml collection tube (both provided in the kit), followed by another centrifugation step for 30 s at 8000 x g. The AllPrep DNA spin column was subsequently used for DNA purification. For DNA extraction, two washing steps of the AllPrep DNA spin column were conducted. First, 500 µl Buffer AW1 was added and the column was centrifuged for 15 s at 8000 x g. Second, 500 µl Buffer AW2 was added and the column was again centrifuged for 2 min at full speed. A third centrifugation step for 1 min at full speed was performed to remove remaining ethanol. The DNA was eluted in 100 µl (2 x 50 µl) Buffer EB (preheated to 70°C) and centrifuged for 1 min at 8000 x g. All centrifugation steps were performed with the Heraeus Fresco 21 Centrifuge (Thermo Fisher Scientific). Extracted gDNA samples were subjected to DNA quantification with the NanoDrop 1000 Spectrophotometer, as described in 2.2.7.2. The isolation of gDNA from fresh-frozen PCa tissue samples and subsequent DNA quantification were performed by a colleague.

2.2.5 Isolation of gDNA from buffy coat samples

Genomic DNA was extracted from eight buffy coat samples using the QIAamp ® DNA Mini and Blood Mini Kit (Qiagen). Initial preparation of the kit's reagents and performance of the protocol for gDNA isolation were conducted according to the manufacturer's instructions. All centrifugation steps were performed with the Heraeus Fresco 21 Centrifuge (Thermo Fisher Scientific) at room temperature. Isolation of gDNA was performed using 200 µl of buffy coat per sample. First, 20 µl proteinase K and 200 µl Buffer AL were added to the sample, and mixed thoroughly. The mixture was incubated on a heating block at 56°C for 10 min. Subsequently, 200 µl ethanol (96–100%) were added, and the total volume was transferred to a QIAamp Mini spin column, followed by a centrifugation step at 6000 x g for 1 min. Two washing steps were performed, first with 500 µl Buffer AW1 and centrifugation at 6000 x g for 1 min, followed by washing with 500 µl Buffer AW2 and centrifugation at 20,000 x g at full speed for 3 min. One additional centrifugation step at 20,000 x g at full speed for 1 min was performed to remove any remaining liquid in the column. The gDNA was eluted in 200 µl nuclease-free water and centrifuged at 6000 x g for 1 min. Isolated gDNA samples were subjected to DNA quantification with NanoDrop 1000 Spectrophotometer, as described in 2.2.7.2. The isolation of gDNA from buffy coat samples and subsequent DNA quantification were performed by a colleague.

2.2.6 Shearing of gDNA from tissue and buffy coat samples

To achieve homogeneously fragmented DNA as input for NGS library preparation, gDNA derived from PCa tissue samples and buffy coat samples was sheared using the Covaris M220 Focused ultrasonicator. First, gDNA samples were quantified with Qubit 2.0 Fluorometer, as described in 2.2.7.1, and concentrations were adjusted to 7.7 ng/µl to obtain an input of 1 µg gDNA in 130 µl elution volume. Sample volume was transferred to a Covaris microTUBE (130 µl) and subsequently processed with the Covaris ultrasonicator. The instrument setup and shearing procedure were performed according to the manufacturer's protocol. The following parameter settings for shearing were applied: Peak incidence power: 50 W; Duty factor: 20%; Cycles per burst: 200; Time: 220 sec. The described setup resulted in a mean gDNA fragment length of ~150–220 bp. The sheared gDNA samples were transferred to a 1.5 ml microcentrifuge tube and were stored on ice (before and after the shearing procedure). Sheared gDNA samples were subjected to DNA quantification with Qubit 2.0 Fluorometer (Qubit dsDNA HS Assay Kit) and to assessment of fragment length distribution using the Fragment Analyzer 5200 (HS NGS fragment Kit (1–6000 bp)), as described in 2.2.7.1 and 2.2.8, respectively. The gDNA shearing and the subsequent quality control (QC) were performed by a colleague.

2.2.7 DNA quantification

2.2.7.1 DNA quantification with Qubit Fluorometer

Quantification of DNA concentrations in cfDNA samples, sheared gDNA samples and NGS libraries was performed using the Qubit 2.0 Fluorometer and the Qubit dsDNA High-Sensitivity (HS; concentrations: 0.005–120 ng/µl) or Broad Range (BR; concentrations: 0.2–2000 ng/µl) Assay Kit reagents. Sample preparation and instrument loading were performed according to the manufacturer's protocol. Briefly, 1 µl of each sample was mixed with 199 µl master mix, consisting of Qubit dsDNA HS buffer (or Qubit dsDNA BR buffer) and fluorescent dye. The mixture was thoroughly vortexed and incubated for 2 min at room temperature and under light protection. The fluorescence intensities for each sample were measured in comparison to an external standard provided by the Qubit dsDNA HS Assay Kit (or Qubit dsDNA BR Assay Kit) to infer the DNA concentrations of the measured samples.

2.2.7.2 DNA quantification with NanoDrop

Quantification of extracted gDNA concentrations was performed using the NanoDrop 1000 Spectrophotometer. Briefly, 1 µl of gDNA sample were loaded on the instrument, and DNA quantification was performed according to the manufacturer's protocol. The quantification of gDNA samples was performed by a colleague.

2.2.8 Assessment of DNA fragment lengths by capillary electrophoresis

Assessment of the fragment size distribution and DNA integrity of cfDNA, sheared gDNA, and NGS libraries was performed with the multiplexed capillary electrophoresis instrument Fragment Analyzer 5200 (Agilent). The HS NGS fragment Kit (1–6000 bp) reagents were used for the measurement of extracted cfDNA or sheared gDNA, while the NGS fragment Kit (1–6000 bp) reagents were used for the assessment of NGS libraries. Sample preparation and instrument loading were performed according to the manufacturer's protocol. Samples with expected high DNA concentrations were diluted prior to measurement (e.g., isolated gDNA samples were diluted 1:10). Briefly, 2 µl of sample or DNA ladder was mixed with 22 µl of Diluent Marker on a 96-well plate. The sample-buffer mixture was vortexed for 3 min at 2000 rpm on a plate shaker, and the 96-well plate was placed into the Fragment Analyzer instrument. After finalization of the measurement, fragment length distributions were assessed with the ProSize data analysis software v4.0.1.4, using the NGS analysis mode. A smear analysis was performed to assess the proportions of the following fragment length ranges: 20–150 bp, 20–175 bp, 130–160 bp, 130–180 bp, 160–180 bp, 20–520 bp, 520–5200 bp. Additionally, total DNA concentration, total molarity and average fragment length of the samples were obtained.

2.2.9 Preparation of NGS libraries and (cf)DNA methylation immunoprecipitation

CfMeDIP-seq was performed to generate genome-wide DNA methylation profiles of cfDNA from plasma (n = 109) and urine (n = 102), as well as of sheared gDNA from PCa tissue and matched buffy coat samples (n = 8 each). This method is based on the immunoprecipitation of methylated DNA fragments using a methylation-specific monoclonal antibody. The cfMeDIP-seq protocol was performed according to the workflow published by Shen and colleagues[333], who adapted existing MeDIP-seq workflows (Taiwo *et al.* [338]) to very low DNA input amounts, facilitating its application to cfDNA. The (cf)MeDIP-seq workflow consisted of a standard WGS library preparation workflow (i.e., end-repair/A-tailing, adapter ligation, and library amplification) with an embedded DNA methylation enrichment procedure (Figure 6). In addition to the methylation-enriched (cf)MeDIP-seq libraries hereafter referred to as IP samples), the workflow also generated non-enriched libraries (hereafter referred to as IC samples) for lcWGS (Figure 6). The IC samples served as experimental and bioinformatic quality controls to assess methylation enrichment efficiency. The individual steps of the (cf)MeDIP-seq workflow, the generation of lcWGS libraries, and the final methylation enrichment QC are described in the following sections. I performed the entire cfMeDIP-seq workflow, including the generation of cfMeDIP-seq and lcWGS libraries for all plasma and urine samples and all associated QC procedures.

The MeDIP-seq workflow using gDNA derived from eight PCa tissue and matched buffy coat samples was performed according to the same protocol as for the liquid biopsy samples, with few adaptions (described below). The experimental work for the PCa tissue and matched buffy coat samples was carried out by a colleague, who subsequently provided the raw sequencing data for my bioinformatic analyses.

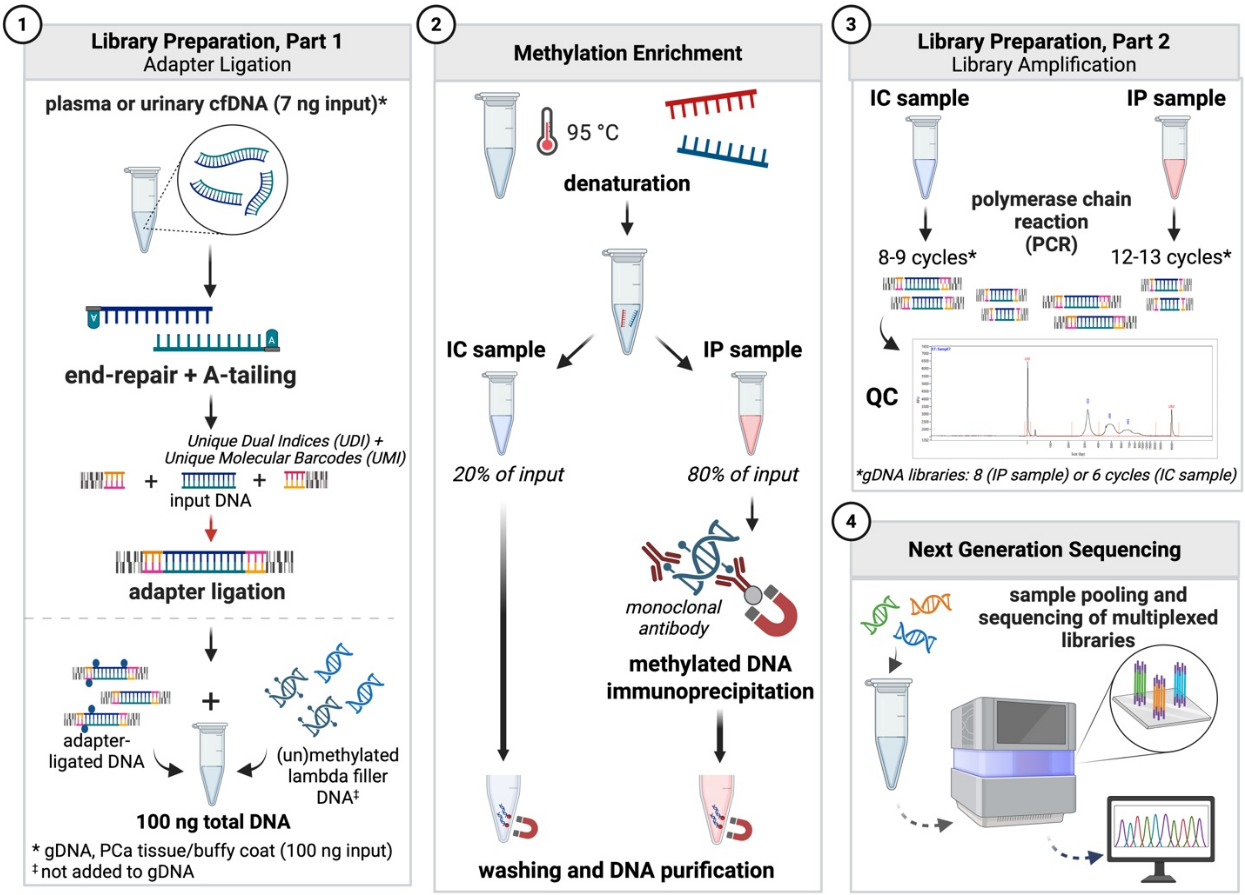


Figure 6: Library preparation workflow for lcWGS and (cf)MeDIP-seq.
Schematic overview of the library preparation workflow for lcWGS and (cf)MeDIP-seq using plasma and urinary cfDNA, as well as gDNA from PCa tissue and buffy coat samples as input. Library preparation includes adapter ligation and library amplification for lcWGS libraries, and an additional intermediate methylation enrichment step for (cf)MeDIP-seq libraries. LcWGS and (cf)MeDIP-seq libraries were generated from the same initial samples by splitting the input material (7 ng input cfDNA or 100 ng input gDNA) during library preparation. Abbreviations: gDNA = genomic DNA, IC sample = non-enriched (lcWGS) library, IP sample = methylation-enriched ((cf)MeDIP-seq) library, lcWGS = low-coverage whole-genome sequencing, (cf)MeDIP-seq = (cell-free) methylated DNA immunoprecipitation sequencing, PCR = polymerase chain reaction, UDI = unique dual index, UMI = unique molecular identifier, QC = quality control. Figure created with BioRender.com.

2.2.9.1 Generation of lambda filler DNA

While previous MeDIP-seq protocols required large DNA quantities as input (≥ 50 ng)[338], Shen and colleagues adapted this protocol for successful application on very low DNA input quantities of 1–10 ng[333]. One essential feature was the introduction of lambda (λ) 'filler' DNA which was added to low input samples to achieve an overall DNA quantity of 100 ng, enhancing the specificity of the DNA methylation enrichment. The filler DNA comprised a mixture of unmethylated (20SCpG) and *in vitro*-methylated (1CpG, 5CpG, 10CpG, 15CpG and 20LCpG) λ DNA fragments (i.e., polymerase chain reaction (PCR) amplicons) of different sizes, reflecting the fragment length distribution of cfDNA (Table 12).

For the generation of filler DNA amplicons, the 300 ng/µl λ DNA stock was diluted to a concentration of 0.1 ng/µl in nuclease-free water, and 1 µl of the diluted λ DNA was added to one PCR master mix per filler DNA amplicon, each containing 10 µl 5x SuperFi Buffer, 1 µl dNTPs (10 mM), 0.5 µl Platinum™ SuperFi™ DNA Polymerase (2 U/µl), 36 µl nuclease-free water, and 3 µl of the appropriate primer pairs (10 µM each, listed in Table 12). Amplification was performed in the thermocycler with the following program: 98 °C for 30 sec (initial denaturation), 30 cycles of 98 °C for 10 sec (denaturation), 57 °C for 10 sec (annealing), 72 °C for 15 sec (extension), followed by 5 min at 72 °C for final extension. Resulting PCR products were purified with a bead clean-up, using AMPure XP magnetic beads with a DNA-to-bead ratio of 1:1. Therefore, 100 µl AMPure XP magnetic beads were added to each pool of two PCR reactions, and the mixture was incubated for 10 min at room temperature. Amplicon-bound beads were captured with a magnet rack, and two washing cycles with each 200 µl ethanol (80%) were followed. The amplicons were finally eluted in 50 µl nuclease-free water. Amplicon sizes were verified using the Fragment Analyzer 5200 (NGS fragment Kit (1–6000 bp)), and quantified using the Qubit 2.0 Fluorometer (Qubit dsDNA BR Assay Kit), as described in 2.2.7.1 and 2.2.8.

In the next step, the amplicons 1CpG, 5CpG, 10CpG, 15CpG, and 20LCpG, but not 20SCpG, were subjected to *in vitro* methylation using the CpG methyltransferase (M.SssI). An amplicon methylation master mix was prepared, containing 2 µl M.SssI buffer (10 x), 0.4 µl S-adenosylmethionine (SAM, 50 x) and 1 µl M.SssI enzyme per 1 µg amplicon. The amplicon volume was adjusted based on the concentration to a final volume of 20 µl. The mixture was incubated for 15 min at 37 °C, followed by 20 min at 65 °C. Methylated amplicons were again purified with a bead clean-up in a pool of maximum five reactions, using AMPure XP magnetic beads with a DNA-to-bead ratio of 1:1, as previously described. The methylated amplicons and the unmethylated 20SCpG amplicon were quantified using the Qubit 2.0 Fluorometer and the Qubit dsDNA HS Assay Kit reagents. Samples with high concentrations were diluted in nuclease-free water prior to quantification. Efficiency of the methylation reaction was verified by digestion of both methylated and unmethylated amplicons with the restriction enzyme HpyCh4IV, whose cleavage activity is known to be blocked by CpG methylation. Therefore, 2 µl CutSmart buffer, and 1 µl HpyCH4IV enzyme were added to 100 ng amplicon. The amplicon volume was adjusted based on the quantified concentration to a final volume of 20 µl. The mixture was incubated for one hour at 37 °C, followed by 20 min at 65 °C. Methylated and unmethylated amplicon sizes were verified using the Fragment Analyzer 5200 and the NGS fragment Kit (1–6000 bp), expecting a successful digestion only in unmethylated fragments. The final pool of filler DNA consisted of a mixture of all five methylated amplicons at equal concentrations combined with five parts of the unmethylated 20SCpG amplicon, resulting in a 1:1 ratio between all pooled methylated amplicons and the unmethylated amplicon (Table 12). Size distribution of the filler DNA pool was assessed using the Fragment Analyzer 5200 and the NGS fragment Kit (1–6000 bp). The DNA concentration was quantified using the Qubit 2.0 Fluorometer (Qubit dsDNA HS Assay Kit), and adjusted to a concentration of 5 ng/µl.

Table 12: Lambda DNA fragments and primer pairs for filler DNA generation.
Overview of λ DNA fragments used for filler DNA generation. The filler DNA consisted of a mixture of unmethylated (20SCpG) and *in vitro*-methylated (1CpG, 5CpG, 10CpG, 15CpG, and 20LCpG) λ DNA fragments generated as PCR amplicons of varying sizes. The table lists the forward and reverse primer sequences, amplicon sizes (bp), CpG content, and the relative proportion of each fragment in the final filler DNA pool. Primers were synthesized and purchased at Eurofins Genomics (Ebersberg, Germany).

Fragment name	Forward primer (5'-3')	Reverse primer (5'-3')	Amplicon size (bp)	CpG sites; GC content (%)	Pooling ratio
1CpG	GAGGTGATAAAATTAACTGC	GGCTCTACCATATCTCCTA	196	1–30.5	1
5CpG	CATGTCCAGAGCTCATTC	GTTTAAAATCACTAGGCGA	269	5–34.1	1
10CpG	CTGACCATTTCCATCATTC	GTAACTAAACAGGAGCCG	359	10–38.6	1
15CpG	ATGTATCCATTGAGCATTGCC	CACGAATCAGCGGTAAAGGT	461	15–43.9	1
20LCpG	GAGATATGGTAGAGCCGCAGA	TTTCAGCAGCTACAGTCAGAATTT	495	20–43.3	1
20SCpG	CGATGGGTTAATTCGCTCGTTGTGG	GCACAACGGAAAGAGCACTG	274	20–49.5	5

Abbreviations: bp = base pair

2.2.9.2 Adapter ligation

The standard WGS library preparation workflow, comprising DNA end-repair and A-tailing as well as adaptor ligation (Figure 6), was performed using the KAPA Hyper Prep Kit (Roche) according to the manufacturer's protocol. To improve the identification of PCR duplicates within the bioinformatic deduplication process of the sequencing data, NEBNext® Multiplex Oligos for Illumina® (New England Biolabs), containing both unique dual indices (UDIs) and unique molecular identifiers (UMIs), were used. The workflow was performed with 7 ng of cfDNA (plasma and urine) or 100 ng of sheared gDNA (PCa tissue and buffy coat) input. Samples with lower available cfDNA quantities were included down to 2.49 ng, still resulting in adequate NGS libraries. First, the input volume was adjusted with nuclease-free water to a final volume of 50 µl. Then, 10 µl of master mix, containing 7 µl End-repair & A-tailing Buffer and 3 µl End-repair & A-tailing Enzyme Mix were added. The mixture was incubated in the thermocycler at 20 °C for 30 min, followed by an incubation for 30 min at 65 °C. Subsequently, the adapter ligation master mix, containing 30 µl Ligation Buffer, 10 µl DNA Ligase and 2.5 µl (cfDNA) or 9 µl (gDNA) nuclease-free water, was added to the prepared samples. Then, 7.5 µl of pre-diluted 750 nM NEBNext UDI-UMI Adaptors (original stock: 20 µM) were added to cfDNA samples, assigning a unique barcode for each sample. For sheared gDNA samples, 1 µl of undiluted 20 µM NEBNext UDI-UMI Adaptors was added. Adapter ligation was performed overnight at 16 °C. Adapter-ligated DNA was purified with AMPure XP magnetic beads, using a 0.8x bead-to-sample ratio (110 µl DNA sample + 88 µl AMPure XP magnetic beads). Purified DNA was eluted in 50 µl nuclease-free water. After DNA purification, a double-sided size selection was performed to obtain adapter-ligated DNA fragments with lengths > 50–100 bp and < 800 bp. Therefore, two magnetic bead purifications were performed with 0.5x bead-to-sample ratio (50 µl eluted DNA sample + 25 µl AMPure XP magnetic beads) and 0.8x bead-to-sample ratio (75 µl DNA sample + 60 µl AMPure XP magnetic beads), respectively. The final cfDNA samples were eluted in 32.4 µl nuclease-free water. For cfDNA samples with an initial input of 7 ng, 18.6 µl of filler DNA (5 ng/µl) was added to reach a total DNA content of 100 ng in a final volume of 51 µl. For samples with lower initial cfDNA input, both elution volume and filler DNA amount were adjusted accordingly. The gDNA samples, which already had an initial input of 100 ng, did not require additional filler DNA, and were directly eluted in 51 µl nuclease-free water. Of note, the filler DNA was not sequenced, as it lacked the required sequencing adapters.

2.2.9.3 Enrichment of methylated DNA fragments

The DNA methylation immunoprecipitation workflow contained the following procedures within a 2-days protocol: sample denaturation, addition of methylated (5mC-) and unmethylated (5C-) spike-ins for subsequent QC, separation of IP and IC samples, methylation enrichment using a monoclonal antibody, as well as sample washing and DNA purification (Figure 6). The MagMeDIP qPCR kit (Diagenode) was used for the DNA methylation enrichment workflow, and the final DNA purification was performed with the iPure Kit (Diagenode). Initial preparation of the kits' reagents and performance of the entire workflow were performed according to the manufacturer's protocols for the two kits.

The protocol was initiated by the addition of 5mC- and 5C-spike-ins, which were quantified during the final methylation enrichment QC using quantitative PCR (qPCR) to assess methylation enrichment efficiency. The 5C- and 5mC-spike-ins as well as the respective qPCR primers were provided in the MagMeDIP qPCR kit. Briefly, 39 µl of a master mix, containing 24 µl of 1:5-diluted MagBuffer A, 6 µl MagBuffer B, 8 µl nuclease-free water, and 1 µl of 5C-spike-in (0.1 ng/µl) and 5mC-spike-in (0.1 ng/µl) were added to the purified, adapter-ligated DNA samples for a total volume of 90 µl. The mixture was incubated for DNA denaturation in the thermocycler at 95 °C (heated lid set to 105 °C) for 10 min, with a ramp up from room temperature to 95 °C. Subsequently, samples were snap cooled in an ice water slush mixture for at least 10 min to maintain single-stranded DNA. Then, sample volume was measured and topped up to 87 µl with ice-cold MagBuffer A (1x), to account for volume loss due to evaporation. In the next step, 20% of the volume (14.5 µl) was separated and stored at 4 °C until further processing. This fraction, containing ~2 ng of cfDNA or ~20 ng of sheared gDNA, represented the non-enriched IC samples, which served as reference for methylation QC and as input for lcWGS. The remaining 80% (72.5 µl) were subjected to methylation enrichment using the monoclonal 5mC-antibody. The following temperature-sensitive steps were performed in the cold room at 4 °C to maintain denatured, single-stranded DNA and ensure successful methylation enrichment. For methylation enrichment using the 5mC-antibody, 0.6 µl of 1:5-diluted MagBuffer A, 2 µl MagBuffer C, 2.4 µl of 1:15-diluted 5mC-antibody in ice-cold nuclease-free water, as well as pre-washed MagBeads diluted in ice-cold MagBuffer A (1x) were added to the IP samples. The mixture was incubated overnight on a rotator (10 rpm) at 4 °C. Subsequently, DNA-bound MagBeads were precipitated on a pre-cooled magnetic stand, and three washing cycles were performed with each 100 µl of MagWash Buffer-1 and incubation on a rotator (10 rpm) at 4 °C for 4 min, followed by one washing cycle with 100 µl MagWash Buffer-2 with only short incubation. MagBeads were finally eluted in 50 µl iPure elution buffer and incubated on a rotator (10 rpm) for 15 min at room temperature, followed by the separation of the DNA-containing supernatant from the MagBeads using a magnetic stand. The supernatant was transferred to a fresh 0.2 ml PCR strip, and the procedure was repeated with another 50 µl of iPure elution buffer to increase the DNA recovery. IC samples were also placed at room temperature, and 85.5 µl of iPure elution buffer was added for a final volume of 100 µl. Purification of 5mC-enriched and non-enriched DNA in IP and IC samples, respectively, was subsequently performed in parallel. All incubation steps were performed on a rotator (10 rpm) at room temperature. Briefly, 2 µl iPure carrier DNA, 100 µl isopropanol (100%) and 10 µl of iPure beads were added to each sample, and the mixture was incubated for 10 min. Two washing cycles were performed using 100 µl of iPure wash buffer-1 and 100 µl of iPure wash buffer-2, each time with a 10 min incubation. Finally, DNA-bound iPure beads were incubated in 25 µl iPure buffer C for 30 min to dissolve DNA from the beads. The samples were placed on a magnetic stand, and the supernatant containing 5mC-enriched DNA (IP sample) or non-enriched DNA (IC sample), respectively, was transferred to a fresh 0.2 ml PCR strip for library amplification. In addition, 2 µl of each eluted sample was transferred to a fresh 0.2 ml PCR strip and diluted with 28 µl nuclease-free water for qPCR quantification of the spike-ins.

2.2.9.4 Library amplification

The cfMeDIP-seq protocol was finalized with the library amplification (Figure 6), using the KAPA Hyper Prep Kit (Roche) reagents. The procedure was performed according to the manufacturer's protocol. Briefly, 25 µl of KAPA HiFi HotStart Ready Mix (2x) and 5 µl of KAPA Library Amplification Primer mix (10x) were added to the final IP or IC samples, respectively, for a final volume of 53 µl. Library amplification was performed in the thermocycler with the following program: 98 °C for 45 sec (initial denaturation), a determined number of cycles at 98 °C for 15 sec (denaturation), 60 °C for 30 sec (annealing), 72 °C for 30 sec (extension), followed by 60 sec at 72 °C for final extension. For cfDNA samples, 12–13 PCR cycles were applied for IP samples, and 8–9 PCR cycles for IC samples. Due to higher initial gDNA input, gDNA samples were amplified using eight PCR cycles for IP samples and six PCR cycles for IC samples. Amplified libraries were purified using AMPure XP magnetic beads at a DNA-to-bead ratio of 0.8 (53 µl amplified sample + 42.4 µl AMPure XP magnetic beads), as previously described. The final libraries were eluted in 12 µl nuclease-free water. Fragment size distributions were assessed using the Fragment Analyzer 5200 (NGS fragment Kit (1–6000 bp)), and libraries were quantified using the Qubit 2.0 Fluorometer (Qubit dsDNA HS Assay Kit).

2.2.9.5 Laboratory QC of DNA methylation enrichment

DNA methylation enrichment efficacy was assessed by quantifying the recovered 5mC- and 5C-spike-ins in both IP and IC samples via qPCR. Briefly, two master mixes were prepared, containing each 5.5 µl of PrimaQuant qPCR Buffer (2x), 0.44 µl of the respective 5mC- or 5C-spike-in primer mix (5 µM each), and 3.06 µl nuclease-free water. The 5mC- and 5C-master mixes were added to a 2 µl sample volume, respectively, and 11 µl of the resulting mixture was measured in triplicate in a 384-well plate on the LightCycler480 instrument. The qPCR was performed at 95 °C for 3 min (initial denaturation), followed by 50 cycles at 95 °C for 10 sec (denaturation) and 60 °C for 30 sec (annealing). Dilution series with (1:1, 1:10, 1:100, 1:200, 1:1000) of both spike-ins were regularly quantified by qPCR to confirm appropriate PCR efficacy. The resulting cycle threshold (Ct) values of recovered 5mC- and 5C-spike-ins in IP samples were compared, aiming for a difference of at least 5–6 cycles, in accordance with Shen *et al.* [333]. In addition, the recovery of the 5mC- and 5C-spike-ins in IP samples and the specificity of the methylation enrichment were calculated, comparing the qPCR results to the ones from matched IC samples. The 5mC-enriched IP samples were expected to harbor high recoveries of the 5mC-spike-ins and low recoveries of the 5C-spike-ins, whereas the non-enriched IC samples were expected to harbor similar recoveries for both the 5mC- and 5C-spike-ins.

First, the Ct value of the IC sample (Ct_{IC}) was adjusted according to the dilution factor (DF) derived from the separation of the input DNA volume into 80% IP and 20% IC sample (i.e., DF = 5; eg. 1):

eg. 1

$$adjusted\ Ct_{\mathrm{IC}} = Ct_{\mathrm{IC}} - \frac{log10(DF)}{log10(2 * AE)}$$

with amplification efficiency (AE) set to 1, assuming 100% AE.

Next, the recovery of the 5mC- and 5C-spike-ins in the IP sample was calculated, as shown in eq. 2:

eg. 2

$$spike\text{-}in\ recovery\ [\%] = (2AE^{adjusted\ Ct_{\mathrm{IC}} - Ct_{\mathrm{IP}}}) \times 100$$

The specificity of the methylation enrichment was calculated based on the ratio of 5C- to 5mC-spike-in recoveries in the IP sample (eq. 3):

eq. 3

$$specificity = 1 - \frac{5C\text{-}spike\text{-}in\ recovery\ [\%]}{5mC\text{-}spike\text{-}in\ recovery\ [\%]}$$

2.2.10 NGS of (cf)MeDIP-seq and lcWGS libraries

NGS of (cf)MeDIP-seq and lcWGS libraries was performed at the High Throughput Sequencing Unit of the DKFZ Genomic and Proteomics Core Facility within five sequencing runs (12/2021, 05/2022, 09/2022, 02/2023, 08/2023). (Cf)MeDIP-seq or lcWGS libraries were pooled equimolarly to a final concentration of 10 µM for multiplexed sequencing. Each multiplex consisted of 56–57 (cf)MeDIP-seq or lcWGS libraries derived from IP and IC samples, respectively. The majority of libraries were subjected to 100 bp paired-end sequencing on the Illumina NovaSeq 6000 S4 instrument. Seven paired (cf)MeDIP-seq and lcWGS libraries were sequenced in two separate multiplexes on the Illumina NovaSeq 6000 SP instrument, also with 100 bp paired-end sequencing mode. Flow-cell loading, cluster formation and NGS were performed by the DKFZ Genomic and Proteomics Core Facility according to the manufacturer's recommendations.

2.2.11 NGS data processing

The raw sequencing data from (cf)MeDIP-seq and lcWGS was provided by the High Throughput Sequencing Unit of the DKFZ Genomic and Proteomics Core Facility as demultiplexed FASTQ files, generated with the tool *bcl-convert 3.9.3*. Three FASTQ files were generated for each sample, containing the sequence data from read 1 (R1), read 2 (R2) and the UMIs (I1). The three FASTQ files were used as input for a pre-processing and alignment workflow, involving removal of remaining adapter sequences (adapter trimming), read alignment to the human reference genome (mapping), identification and removal of duplicate reads (deduplication), and quality filtering (Figure 7).

I implemented the entire pipeline in the workflow management system *Nextflow*, ensuring scalability, portability and reproducibility of the workflow[364]. The software dependencies within the pipeline were managed using the Singularity container runtime[367]. The bioinformatic processes were run on the high-performance computing cluster of the DKFZ Heidelberg, working with the IBM Spectrum LSF workload management software. The *Nextflow* pipeline consisted of a main workflow, including several subworkflows that contained single processes, such as the adapter trimming or the read alignment. I acquired several subworkflows and modules from the nf-core community[398], and implemented them either without any changes or with few individual adaptions. Furthermore, I created additional subworkflows and modules, and integrated them into the pipeline. The dynamic workflow was established for the processing of (cf)MeDIP-seq and lcWGS data from cfDNA or gDNA samples.

In the first step, the UMI information was extracted from the respective FASTQ file, and the individual UMI barcode was added to the read name, using the function `umitools extract` of *UMI-tools v1.1.2*[369]. Subsequently, remaining Illumina sequencing adapters were identified and cleaned with *cutadapt v3.4*[359]. Adapter-trimmed reads in FASTQ format were then aligned to the human reference genome (Human GRCh37/hg19, UCSC genome browser) using the aligner *Bowtie2 v2.4.4*[358] with default parameters and an insert size range restricted to 30–700 bp. Sequence alignment map (SAM) files were converted to binary alignment map (BAM) files, coordinate-sorted, and indexed using the functions `sort` and `index` of *Samtools v1.15.1*[366]. Quality filtering of the aligned sequencing reads was performed based on the properties stored in SAM flags with the function `view -F 12 -f 3` of *Samtools v1.15.1*, keeping only properly paired reads which were mapped in proper pair (`-f 3`), and excluding unmapped reads or reads with unmapped mate (`-F 12`). Additionally, only reads with a mapping quality score (MAPQ) > 10 were kept (`view -q 10`). Filtered BAM files were again indexed using *Samtools v1.15.1*. Subsequently, duplicate reads were identified and collapsed to one representative read based on the mapping coordinate and the UMI barcode, using the function `umitools dedup` of *UMI-tools v1.1.2* with default parameter settings, but discarding all unmapped and unpaired reads, as well as chimeric read pairs. Groups of duplicate reads with similar UMI barcodes were generated by first grouping reads with identical mapping positions and then applying the directional, network-based deduplication method (default). The process resulted in final BAM files with deduplicated and sorted reads, which were again indexed with *Samtools v1.15.1*. QC was performed using *fastqc v0.11.9*[360] for the raw sequencing data (FASTQ files), the final aligned, filtered and deduplicated reads, as well as for all intermediate steps. Additionally, statistical functions (`stats`, `flagstat`, `idxstat`) of *Samtools v1.15.1* were applied to aligned and final processed reads, respectively. QC results were aggregated using *multiQC v1.12*[363].

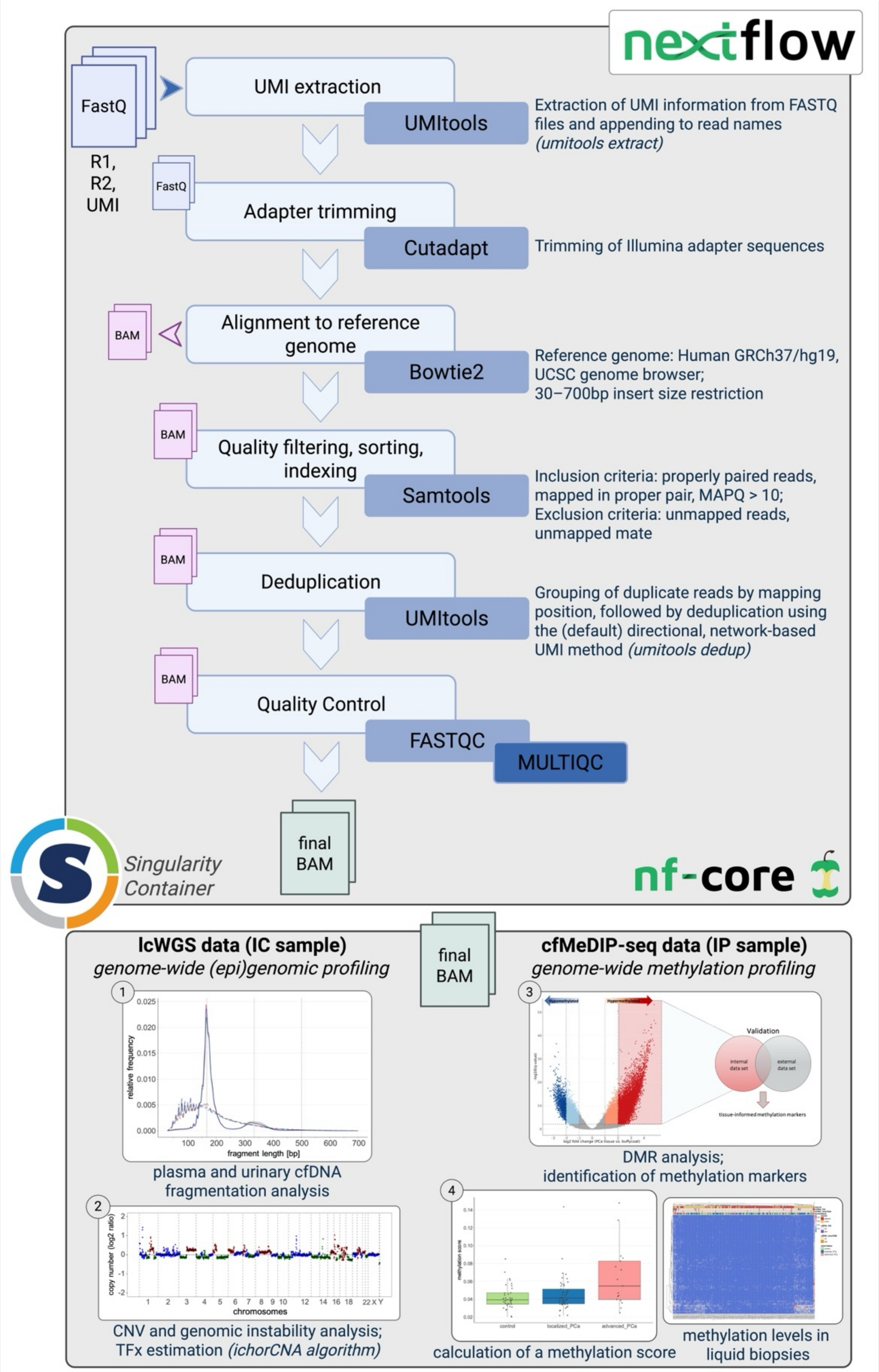


Figure 7: Bioinformatic workflow for preprocessing and downstream analyses of lcWGS and (cf)MeDIP-seq data. Schematic overview of the preprocessing of raw sequencing data from lcWGS and (cf)MeDIP-seq, including extraction of UMI barcodes, adapter trimming, alignment to the human reference genome, deduplication based on UMI information, quality filtering, and QC. The bioinformatic pipeline was implemented using the workflow management system Nextflow[364], and software dependencies were managed using Singularity containers[367]. Subsequent downstream analyses include genomic profiling of CNVs and chromosomal instability, as well as epigenomic profiling of cfDNA fragmentation patterns and aberrant DNA methylation. Abbreviations: BAM = binary alignment map, CNV = copy number variation, DMR = differentially methylated region, MAPQ = mapping quality score, TFx = tumor fraction, UCSC = University of California Santa Cruz. Figure created with BioRender.com.

2.2.12 Bioinformatic QC of DNA methylation enrichment

In addition to the experimental QC after the library preparation, methylation enrichment efficacy was evaluated based on bioinformatic analyses of the processed (cf)MeDIP-seq data. The QC of the sequencing data included the assessment of the sufficient sequence coverage for genome-wide DNA methylation analysis (saturation), CpG coverage, and the methylation enrichment score. The analyses were performed using the R package *MEDIPS v.1.46.0*[382], applying the functions `MEDIPS.saturation()`, `MEDIPS.CpGenrich()` and `MEDIPS.seqCoverage()` to assess the three parameters, respectively.

2.2.12.1 Saturation

The saturation analysis assessed whether the mapped reads in the (cf)MEDIP-seq data of each sample contained a sufficient number of short reads to present a genome-wide saturated coverage profile of the reference genome, which could also be reproduced by an independent dataset with a similar number of short reads[382]. Within the saturation analysis, the available regions for each sample were randomly divided into two equally sized sets. For each set, the read coverage at genome-wide 300-bp windows was iteratively (n = 10) calculated for an increasing number of reads. For each iteration, the resulting coverage of the two sets was compared using Pearson correlation. A higher correlation between the two sets was expected for an increasing number of considered reads. The result of the highest correlation between the two sets was reported as “saturation” value for the given sample. In addition, a second value, termed “estimated saturation”, was calculated using the same approach, but the available regions were first artificially doubled before being randomly divided into two equally sized sets and subsequently analyzed. This step enabled the assessment of the complete dataset rather than only half of the set, as in the first scenario.

2.2.12.2 Methylation enrichment score

Methylation enrichment analysis investigated the appropriate enrichment of CpG-harboring DNA fragments during library preparation relative to all genome-wide DNA fragments. The CpG coverage in genomic regions, which were covered by the mapped sequencing reads, were compared to the entire reference genome. The relative frequency of CpGs, defined as the number of existing CpGs compared to the total number of bases, was calculated for both the covered regions and the entire reference genome[382]. The methylation enrichment score was defined as the ratio between these two parameters. This enrichment score was assessed for all (cf)MeDIP-seq and lcWGS data derived from plasma and urinary cfDNA, as well as from sheared gDNA from PCa tissue and buffy coat samples.

2.2.12.3 CpG coverage

The CpG coverage analysis evaluated the number of CpGs and their presence in mapped reads of given genomic regions. Detectable CpGs within the reads were compared with the known CpG density in the corresponding regions in the reference genome. Several parameters were reported for each sample, including the total number of CpGs, the number of reads without present CpG pattern, as well as the proportion of CpGs that were covered by reads at 0x, 1x, 2x, 3x, 4x, 5x and > 5x coverage, respectively. Three key parameters were assessed for each sample to characterize CpG coverage. These were the following: the proportion of CpGs relative to all present CpGs in the reference genome that were covered with at least one read or more, the proportion of CpGs that were covered by more than five reads, and the proportion of reads without CpG pattern[382]. The latter represented the unspecific bound DNA fragments during the immunoprecipitation step in the library preparation.

2.2.13 Plasma and urinary cfDNA fragmentation analysis

I investigated the fragment length distributions of plasma and urinary cfDNA by assessing insert size lengths in lcWGS and cfMeDIP-seq data. The insert size described the length of both reads (R1 and R2) in paired-end sequencing data plus the inner distance between the two reads, in case that they were not attached or overlapping (Figure 8A). For each sample and sequencing type, the insert sizes were extracted to a text file from the final aligned, filtered and deduplicated BAM files, using *Samtools v1.15.1* and the following code:

```
samtools view -f66 file.bam | cut -f 9 > insert_sizes.txt
```

The insert sizes corresponded to the plasma and urinary cfDNA fragment lengths, and will be referred to interchangeably in the following sections. For each sample, the frequency of each insert size was determined, and its relative frequency was calculated as the proportion of the total size range (30–700 bp). The cfDNA fragmentation analysis was inspired by previous studies from Mouliere and colleagues[205,213,274], and slightly adapted or supplemented with additional features. Several statistical parameters were assessed based on the relative frequency distributions of cfDNA fragment lengths. These were the following: median and mean fragment length, maximum relative frequency, modal fragment length (fragment length with the highest frequency), cumulative frequency at the modal fragment length, fragment length at a cumulative frequency of 0.5, as well as the relative and cumulative frequencies at the theoretical mono-nucleosomal (167 bp fragment length) and di-nucleosomal (334 bp fragment length) peaks. Additionally, proportions of various cfDNA fragment length ranges compared to the entire fragment length range (30–700 bp length) were calculated based on the relative frequency distributions (Figure 8A). The proportions of the following fragment length ranges were assessed: P30–60 bp, P30–100 bp, P30–150 bp, P30–180 bp, P90–150 bp, P160–180 bp, P163–169 bp, P180–220 bp, P150–300 bp, P250–320 bp, P250–420 bp, P324–344 bp, and P420–700 bp. Subsequently, pairwise ratios between various fragment length ranges were calculated: P30–150 bp/P160–180 bp, P90–150 bp/P163–169 bp, P30–150 bp/P150–300 bp, P30–150 bp/P163–169 bp, P30–100 bp/P160–180 bp, P30–100 bp/P163–169 bp and P160–180 bp/P250–420 bp. Both the individual proportions and the ratios between proportions of different plasma or urinary cfDNA fragment length ranges, respectively, were statistically compared between tumor and control samples.

In addition, the oscillation pattern with 10-bp periodicity was assessed following previous publications by Mouliere and colleagues[205,274]. This oscillation pattern was characterized by alternating local maxima and minima with an approximate spacing of 10 bp between adjacent maxima or minima (Figure 8B). These distances between two local maxima or two local minima were referred to as interpeak and intervalley distances, respectively. Local maxima and minima were defined as the fragment length with the highest or lowest relative frequency, respectively, within a sliding interval of 5 bp, applying the function `rollapply()` of the R package *zoo v.1.8-12*[396]. The height of the maxima and the depth of the minima were obtained from the relative frequencies. An oscillation score was calculated based on the sum of the heights of all local maxima minus the sum of the depths of all local minima detectable within the two fragment length ranges 30–150 bp and 150–300 bp, respectively (Figure 8B).

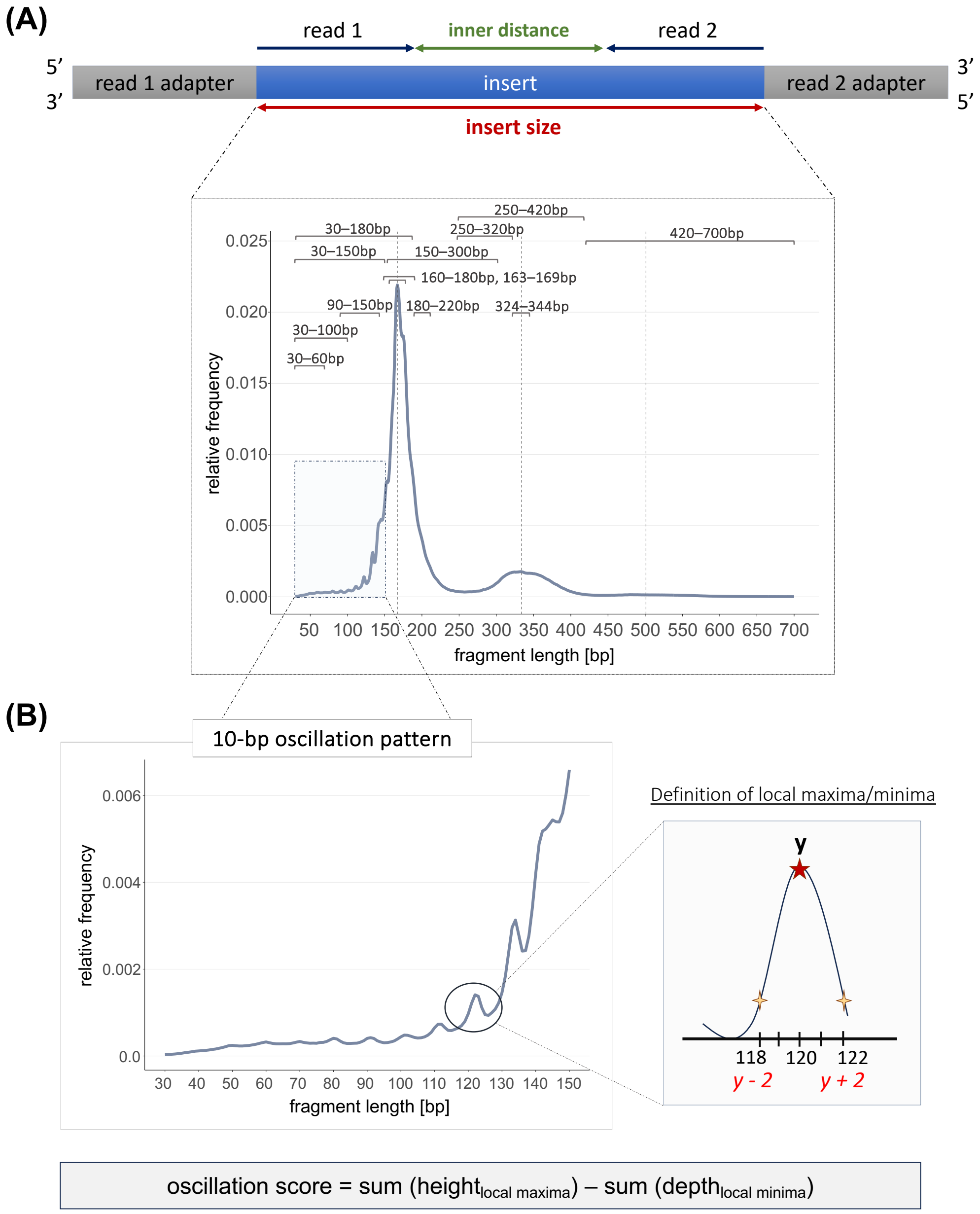


Figure 8: Fragmentation analysis of plasma and urinary cfDNA based on insert size distributions derived from lcWGS and cfMeDIP-seq data.
(A) Insert sizes were extracted from sequencing reads of lcWGS and cfMeDIP-seq data to assess plasma and urinary cfDNA fragmentation. Fragmentation profiles were generated as relative and cumulative frequency distributions of insert sizes (fragment lengths), and the proportions of predefined fragment length ranges were quantified. **(B)** Plasma and urinary cfDNA fragmentation profiles showed a characteristic oscillation pattern with approximately 10-bp periodicity. Local maxima and minima were defined as data points representing the highest or lowest values, respectively, within a ±2 bp window. A synoptic oscillation score was calculated as the difference between the sum of the heights of all local maxima and the sum of the depths of all local minima.

2.2.14 Analysis of CNVs and chromosomal instability

2.2.14.1 Genome-wide analysis of CNVs

Liquid biopsy samples

The assessment of genomic instability in plasma and urinary cfDNA was used to identify CNVs and to estimate the ctDNA abundance (tumor fraction, TFx) based on the CNV burden in the liquid biopsy samples. Genome-wide copy number profiles were generated from both lcWGS and cfMeDIP-seq data, using the processed BAM files as input for the CNV analysis. I established the workflow in *Nextflow*, including several pre-processing steps and the actual CNV analysis using the R package *ichorCNA v.0.4.0*[286]. In the first step, sequencing reads mapping to ENCODE blacklisted genomic regions, which are known to increase the noise in genomic analyses of high-throughput sequencing data and may lead to inaccurate interpretation of the results[399], were excluded from further processing, using the function `intersect -v` of *Bedtools v2.30.0*[357]. The resulting BAM files were indexed with *Samtools v1.15.1,* QC was performed with *fastqc v0.11.9*, and the remaining number of sequencing reads was determined using the function `view` of *Samtools v1.15.1*. Subsequently, sequencing reads were randomly downsampled to 20 million (M) paired reads to achieve a uniform, comparable coverage across all samples, applying the function `view -s DOWNSAMPLING FACTOR` of *Samtools v1.15.1*. The final BAM files were indexed with *Samtools v1.15.1*. In the next step, the function `readcounter` of *HMM Copy Utils v0.1.1*[361,362] was used to segment the genome into non-overlapping 1000-kb bins, count the number of sequencing reads per bin, and store the resulting information in WIG file format for each sample. The WIG files served as input for the CNV analysis using *ichorCNA v.0.4.0*. For both tumor and control samples, the segmented reads in 1,000-kb bins were corrected for GC content and mappability bias using the R package *HMMcopy v1.32.0*[361]. A Panel of Normals (PoN) was created based on the normalized reads from control samples, which served as copy-number-neutral reference for the identification of positive (gains, amplifications) or negative (losses, deletions) deviations from the copy-number-neutral state in tumor samples. For the analysis of tumor samples, all controls were included in the PoN, while individual PoNs were generated for each control sample including all controls except for the analyzed sample. A hidden Markov model (HMM), implemented as Bayesian statistical framework, was used for segmentation as well as for subsequent prediction of large-scale CNVs and estimation of the tumor fraction in the samples[286]. The *ichorCNA* algorithm was executed with adjusted parameter settings for expected low tumor-derived DNA content in the cfDNA samples. All chromosomes were analyzed, but only the autosomal chromosomes were used for parameter estimation. As all tumor patients and controls were men, the gender was set to “male”. The following shows a selection of the most relevant parameters:

```
--ploidy  "c(2)"  --normal  "c(0.9,  0.95,  0.99,  0.995,  0.999,  1)"  --maxCN  3  \
--includeHOMD  False  --chrs  "c(1:22,\"X\",\"Y\")"  --chrNormalize  "c(1:22)"  \
--chrTrain   "c(1:22)"   --estimateNormal   True   --estimatePloidy   True   \
--estimateScPrevalence  False  --fracReadsInChrYForMale  0.001  --scStates  "c()"  \
--txnE 0.9999999 --txnStrength 10000000 --gender "male" --genderNormal "male" \
--genomeBuild "hg19" --genomeStyle "UCSC" --minSegmentBins 50
```

The *ichorCNA* algorithm generated CNV profiles and estimated TFx values for both tumor and control samples, which were subsequently subjected to statistical comparison between cohorts.

The described CNV profiling approach was initially performed based on genome-wide sequencing reads with 30–700 bp insert size. In addition, *in silico* size selection of sequencing reads was carried out, intending to enrich for tumor-derived DNA[274], and the described CNV workflow was repeated using the size-selected sequencing reads as input. This analysis was applied to the lcWGS data derived from plasma and urinary cfDNA samples. *In silico* size selection led to a high reduction of the number of sequencing reads, requiring further downsampling to 2 M paired reads to maintain comparability across the samples. For plasma samples, I performed *in silico* size selection for the insert size range 90–150 bp, which has previously been shown to enhance ctDNA detection based on genome-wide CNV profiling[204,274]. For urinary cfDNA, I evaluated 15 different size ranges to identify the most appropriate range for a potential improvement of the CNV analysis. The following ranges were investigated: 90–150 bp, 110–170 bp, 20–100 bp, 20–110 bp, 20–120 bp, 20–130 bp, 20–140 bp, 20–150 bp, 20–160 bp, 20–170 bp, 140–170 bp, 140–180 bp, 140–190 bp, 140–200 bp and 140–210 bp. For each fragment length range, it was tested whether the estimated TFx values in the tumor and control cohorts were significantly different from those obtained by the CNV analysis without size selection. Only fragment length ranges, for which tumor samples showed significant differences and the control cohort revealed a non-significant result, were considered. For each remaining fragment length range, the absolute difference and the ratio between TFx values with and without size selection were calculated to assess, whether size selection resulted in an increase or decrease in the TFx values. Absolute difference and ratio were calculated, as following:

$$absolute\ difference = TFx_{\text{size selection}} - TFx_{\text{without size selection}}$$

and

$$ratio = \frac{TFx_{\text{size selection}} + 0.001}{TFx_{\text{without size selection}} + 0.001}$$

The median values of all tumor and control samples were compared, and the ratio between the median absolute differences of tumor and control samples was calculated:

$$ratio = \frac{median(absolute\ difference_{\text{size selection}})_{\text{tumor}}}{median(absolute\ difference_{\text{size selection}})_{\text{control}}}$$

This comparison aimed to assess for which fragment length ranges the size selection resulted in a higher increase in the median TFx values in tumor samples compared with control samples, potentially indicating ctDNA enrichment.

Fresh-frozen PCa tissue samples

I also performed genome-wide CNV profiling performed based on the lcWGS data derived from eight PCa tissue and matched buffy coat samples. For each tissue sample, the data from the corresponding buffy coat sample severed as copy-number-neutral reference. CNVs were also assessed in buffy coat samples, and individual PoNs, including all buffy coat samples except for the analyzed sample, were generated as copy-number-neutral references for each sample. The same CNV analysis workflow as for the liquid biopsy samples was applied, with a few adaptions for gDNA samples. The *ichorCNA* algorithm was executed with default parameter settings and the following most relevant parameters:

```
--ploidy "c(2)" --normal "c(0.2, 0.3, 0.4,0.5,0.6,0.7,0.8,0.9)" --maxCN 5 \
--includeHOMD False --chrs "c(1:22,\"X\",\"Y\")" --chrNormalize "c(1:22)" \
–-chrTrain "c(1:22)" --estimateNormal True --estimatePloidy True \
--estimateNormal True --estimateScPrevalence True --fracReadsInChrYForMale 0.001 \
--scStates "c(1,3)" --txnE 0.9999 --txnStrength 10000 --sex "male" \
--genomeBuild "hg19" --genomeStyle "UCSC" --minSegmentBins 50
```

For buffy coat samples, the expected normal contamination was adapted accordingly:

```
--normal "c(0.9, 0.95, 0.99, 0.995, 0.999, 1)"
```

In the *ichorCNA run script*, custom parameter settings were adjusted from the default cfDNA setting (0.1x) to those recommended for 1x bulk tumor samples in the parameter section for HMM preparation.

CNV profiles and estimated TFx values were generated for all PCa tissue and buffy coat samples. The *in silico* size selection approach was not applied, because sheared gDNA was used as input for lcWGS.

Identification of recurrent CNVs

Recurrent alterations were identified either across all plasma and urine samples with detectable ctDNA or restricted to samples with TFx values above 10%. Copy number states (neutral, loss, gain) in genome-wide 1000-kb bins were merged across all samples and the frequency of their occurrence per corresponding bin were calculated. The copy number states of individual samples were extracted from the *.cna.seg*-files which were obtained as output from the *ichorCNA* algorithm. Analogously, recurrent alterations were assessed in all PCa tissue samples with detectable CNVs, and were compared with results from liquid biopsy analyses.

2.2.14.2 Chromosomal instability analysis

I applied the chromosomal instability analysis (CIA) to the lcWGS data derived from plasma and urinary cfDNA. The magnitude of chromosomal instability was summarized in one synoptic CIA score for each sample. While the TFx values from the *ichorCNA* algorithm were estimated based on large-scale CNVs, the CIA score considered only the most deviating genome-wide 1,000-kb bins. Calculation of the CIA score was performed according to similar approaches from previous publications[400,401].

For each tumor or control sample, log ratio (logR) values in genome-wide 2,601 windows with 1,000-kb size were extracted from the *.cna.seg*-file obtained as output from the *ichorCNA* algorithm. These logR values reflected raw read counts normalized by GC content and mappability, and finally corrected by the median logR value derived from the PoN per genomic window. The logR values of the PoN reflected read counts normalized for GC content and mappability without further correction.

The calculation of the CIA score was based on the genome-wide z-scores for each bin, which were calculated based on the read counts normalized by GC content and mappability without PoN correction. Prior to z-score calculation, the previously applied PoN correction was subtracted from the individual sample logR values to obtain uncorrected logR values for each bin:

$$uncorrected\ logR(bin)_{\text{sample}} = final\ logR(bin)_{\text{sample}} + logR(bin)_{\text{median PoN}}$$

Z-scores for 1,000-kb bins were calculated by comparing normalized counts to the mean logR value of the PoN, also considering the standard deviation (SD) of the logR values:

$$z\text{-}score(bin)_{\text{sample}} = \frac{uncorrected\ logR(bin)_{\text{sample}} - mean\ logR(bin)_{\text{PoN}}}{SD\ logR(bin)_{\text{PoN}}}$$

The 2,601 bins were ranked based on their absolute z-scores (|z-score|), and the 95th- and 99th-percentile of values were extracted for each individual sample. To define the CIA score, the sum of absolute z-scores for bins within the 95th- and 99th-percentile interval was calculated. CIA scores were determined for all plasma and urine samples from tumor patients and controls, respectively.

2.2.15 DNA methylation profiling

I performed genome-wide DNA methylation profiling and assessment of differentially methylated regions (DMRs) on the (cf)MeDIP-seq data derived from plasma and urine samples, as well as from PCa tissue and matched buffy coat samples, respectively. The processing of the DNA methylation data and the determination of DMRs were carried out using the R package *MESA v0.2.2*[343], which represent an extension of the R package *QSEA v.1.16.0*[389]. Both packages were designed for the analysis of (cf)MeDIP-seq data and provide workflows for methylation QC, data normalization, and data manipulation, including CNV correction of the methylation data based on matched lcWGS data and transformation of relative methylation signals into absolute methylation levels (β-values), as well as workflows for DMR analysis. At the time of initial use, the R package MESA had only recently been established, and several minor adjustments were required to ensure its proper application to my dataset. Additionally, I adapted the functions to enable application with both human reference genome assemblies GRCh37/hg19 or GRCh38/hg38, as the R package was initially established only for GRCh38/hg38.

The first step of the methylation analysis workflow was the creation of a QSEAset object based on the pre-processed and aligned BAM files. The QSEAset stored relevant metadata for all samples and included one matrix containing the coverage in genome-wide windows of defined size, as well as a second matrix containing the zygosity for all chromosomes, the genomic regions of the coverage matrix, and information about the reference genome and the analyzed chromosomes (e.g., chromosome length). The QSEAset also stored CNV information in 1-Mb windows, different parameters of the sequencing libraries including normalization factors for effective library size, and parameters of the CpG enrichment analysis, which were needed for the transformation to absolute methylation values. The function `makeQset()`of the R package *MESA v0.2.2* was used with the following parameter settings to create the QSEAset:

```
--mininsertsize 20 --maxinsertsize 1000 --windowsize 300 --fragtype NULL \
--fragmentLength NULL --fragmentSD NULL --cnvwindowsize 1000000 \
--cnvmethod "HMMdefault" \ --coveragemethod "qseaPaired" \
--calcNormFactor True --properPairsOnly False --minmapq 10 \
--badregions "blacklist_encode_hg19_Grange" \
--genome "BSgenome.Hsapiens.UCSC.hg19" \
--chrStyle "UCSC" --minReferenceLength 0 \
--chrSelect "chr1,chr2,chr3,chr4,chr5,chr6,chr7,chr8,chr9,chr10,chr11,chr12,chr13,
chr14,chr15,chr16,chr17,chr18,chr19,chr20,chr21,chr22,chrX,chrY"
```

Genomes were divided into non-overlapping 300-bp bins, and read coverage was calculated as well as CpG density estimated in each genomic window. ENCODE-blacklisted regions for human reference genome hg19[399], sequencing reads with a mapping quality < 10, and windows with less than 10 reads over all analyzed samples were excluded. CNV information in 1-Mb windows were obtained from paired non-enriched lcWGS data using the R package *HMMCopy v.1.32.0*[361] (as part of the R package *QSEA*), and were subsequently used for CNV correction of (cf)MeDIP-seq data. Differences in library size were considered by normalizing counts to trimmed mean of M (TMM) values[402] (as part of the R package *QSEA*), using the library whose upper quartile was closest to the referring mean upper quartile. Genome-wide, relative methylation signals were expressed as normalized reads per kilobase million mapped reads (NRPKM), representing counts in 300-bp windows normalized for CNVs, library size, region length, and zygosity. Additionally, absolute methylation levels were estimated from the relative signals and reported as β-values for each window in each sample. The β-values were scaled from 0 to 1, with 0 being unmethylated and 1 fully methylated. The transformation of relative (cf)MeDIP-seq enrichment signals into absolute methylation levels required the estimation of two components: (1) the so-called offset, which accounted for background read abundance (experimental noise), representing non-methylated regions or regions lacking CpGs and distorting the methylation signal, and (2) the CpG density-dependent read enrichment, approximating the observed enrichment profile with a sigmoidal enrichment function calibrated for each sample by using the assumption of an inverse relationship between methylation and CpG density (“blind calibration”) [389]. The read coverage in genomic windows with known absolute methylation levels were modelled with a linear Poisson distribution, considering the described parameters, the library size, and CNVs. Subsequently, a Bayesian model was employed to estimate absolute methylation levels for all genomic windows.

The QSEAset formed the basis for downstream analyses of the methylation data, including the DMR analysis. Based on the data stored in the QSEAset object, two matrices were generated, containing NRPKMs or β-values, respectively, in all genome-wide 300-bp windows for all samples.

2.2.15.1 Identification of DMRs in liquid biopsy samples

DMRs were determined in plasma and urine samples by comparing all PCa patients or PCa subcohorts with controls, respectively, as well as by comparing the different PCa subcohorts with each other. Two QSEAset objects were created for either all plasma or all urine samples from both PCa patients and controls, as previously described. Subsets of these large QSEAset objects were extracted, depending on the two groups which were compared for the identification of DMRs. The DMR analyses were performed using the function `calculateDMRs()` of the R package *MESA v0.2.2*. Low-coverage regions were defined as genomic 300-bp windows, in which none of the samples from the analyzed dataset harbored ≥2 NRPKMs, and were excluded from the DMR analysis. Absolute methylation levels (β-values) in the remaining windows were compared between two cohorts to identify significant DMRs. The determination of DMRs was based on a generalized linear model (GLM), where read counts were modeled with a negative binomial distribution, and a GLM was fitted for each window with a logarithmic link function. In significance testing, a reduced model distinguishing between two groups (“normal” versus “tumor”) was fitted and the likelihood ratio of the model to a Chi-squared distribution was compared. Benjamini-Hochberg method was used for the correction of multiple testing. The initial transformation from NRPKM to β-values prior to the DMR analysis failed in single genomic windows, if a defined threshold for the minimal number of expected reads for a fully methylated window was not reached (*minEnrichment* = 3, default setting). Consequently, NA-values were assigned to these affected windows. Identified DMRs for which the majority of samples in one or both compared groups contained NA-values were excluded from further analysis. Differences in the methylation levels between the two compared groups were expressed as log fold changes (logFC) of the mean β-values within the identified DMRs, as logFC captures small, yet significant, effects more sensitively that, for example, absolute difference in mean β-values between two groups. Positive logFC values represented hypermethylated regions and negative logFC values indicated hypomethylated regions in one group (e.g., PCa patients) relative to the reference group (e.g., controls). DMRs were annotated for genomic regions (promoter, intron, exon, downstream, distal intergenic regions, 5’UTR, 3’UTR) and CpG-associated landscapes (CpG islands, Shores, Shelf and Open Sea regions), using the function `annotateData()` of the R package *MESA v0.2.2*, with GRCh37/hg19 as human reference genome.

Identified DMRs between two cohorts were overlapped with DMRs detected in another comparison (e.g., metastatic vs. non-metastatic PCa, and metastatic PCa vs. controls) to identify shared regions of aberrant DNA methylation. The function `findOverlapsOfPeaks()` of the R package *ChIPpeakAnno v.3.24.2*[371] was applied to determine overlapping regions between two DMR sets. A permutation test was performed using the R package *regioneR v.1.22.0*[393] to evaluate whether the overlap between two region sets occurred non-randomly. The significance of the observed overlap was assessed by comparing it to a null distribution of overlaps expected under random genomic positioning using the function `permTest()`. The null distribution was generated from 5,000 permutations applying the function `randomizeRegions()`, in which one region set was randomly shuffled across the genome (preserving the original region sizes and chromosome distribution), and the number of overlaps was calculated in each iteration.

2.2.15.2 Genome-wide methylation profiling in PCa tissue samples

I also performed genome-wide DNA methylation profiling for eight PCa tissue samples based on the MeDIP-seq data. Aberrant DNA methylation in PCa tissue was assessed by determining DMRs between PCa tissue and matched buffy coat samples, which served as normal reference. A single shared QSEAset object was created for all PCa tissue and buffy coat samples and was used as input for the DMR analysis based on β-values in genome-wide 300-bp windows, as previously described for the liquid biopsy analyses (see section 2.2.15.1). Significant DMRs were defined with increased stringency based on adjusted p values < 0.01 after correction for multiple testing. Identified DMRs were further filtered based on logFC > 2 or logFC < -2 to select hypermethylated or hypomethylated regions in PCa tissue compared with buffy coat samples, respectively. DMRs were annotated for genomic regions and CpG-associated landscapes, as previously described.

An independent, publicly available MEDIP-seq dataset derived from primary PCa tissue and normal prostate tissue samples, generated by Börno and colleagues[403], was used to validate the findings of my PCa tissue-based DNA methylation analysis, and to identify tumor-informative methylation markers for the application on liquid biopsy samples. The authors performed DMR analysis based on reads per million (RPM) values in genome-wide 500-bp windows using Mann-Whitney-U-test to identify significant differences between PCa tissue (n = 51) and normal prostate tissue samples (n = 53). The Benjamini-Hochberg method was applied to correct for multiple testing (adjusted p value < 0.05). The authors provided a table containing the 100 top-ranked (TOP 100) hypermethylated DMRs in PCa tissue, ranked by the lowest p values among all DMRs identified (Supplementary Data, Table S3a[403]). These 100 genomic regions of 500-bp size were overlapped with the significant, hypermethylated DMRs (logFC > 2; 300 bp-regions) from my PCa tissue-based DMR analysis, applying the function `findOverlapsOfPeaks()` of the R package *ChIPpeakAnno v.3.24.2*. A permutation test was performed using the R package regioneR v.1.22.0[393], as described in the section 2.2.15.1. The shared regions in both datasets were selected as tumor-informative methylation markers for the application on liquid biopsy samples.

2.2.15.3 Assessment of DNA methylation markers in liquid biopsy samples

The matrix containing β-values in genome-wide 300-bp windows for either plasma or urine samples, respectively, from PCa patients and cancer-free controls was filtered for the PCa tissue-derived DNA methylation marker regions. The distribution of β-values in these regions was statistically compared between tumor and control samples. A synoptic methylation score was calculated for each plasma and urine sample from PCa patients and cancer-free controls. The methylation score was defined as the median of all β-values across the 67 PCa tissue-derived DNA methylation marker regions for the respective sample. The distributions of methylation scores in plasma and urinary cfDNA derived from PCa patients and controls were statistically compared.

In addition, PCa tissue-derived DNA methylation markers were assessed in an external liquid biopsy cohort published by Chen and colleagues[354]. The authors performed cfMeDIP-seq on 133 plasma samples from patients with primary localized PCa (n = 30) and mCRPC (n = 103). The plasma samples from localized PCa patients originated from the Canadian Prostate Cancer Genome Network (CPC-GENE) project[404], while the mCRPC samples were gained from three different cohorts (Barrier-, WCDT-, and VPC cohort, respectively) with mCRPC patients undergoing targeted AR inhibition treatment, such as administration of abiraterone and/or enzalutamide, with varying therapy regimens[405,406]. Plasma samples were collected at baseline, and along the treatment course or upon (PSA) progression under treatment, respectively. Raw sequencing data derived from cfMeDIP-seq were obtained from the European Genome-Phenom Archive (EGA) under the accession numbers EGAD00001007972, EGAD00001008711, EGAD00001008712, and EGAD00001008713, containing a pair of R1- and R2-FASTQ files for each sample. Available sequencing data from additional 72 plasma samples derived from mCRPC patients, serving as a validation cohort in the study from Chen *et al.* [354], were excluded from further analysis. FASTQ files from cfMeDIP-seq were subjected to the established *Nextflow* pre-processing and alignment pipeline, as previously described (see section 2.2.11). Identification and removal of duplicate reads was performed using P*icard v.2.27.2's* function `markduplicates --REMOVE_DUPLICATES true`[365], as UMIs were not available for the external cohort. All other processes and parameter settings remained unchanged and were applied in the same manner as for my cfMeDIP-seq data. For subsequent DNA methylation profiling, a QSEAset object was created using the R package *MESA v0.2.2*, containing the external cfMeDIP-seq data from all 133 plasma samples. The same parameter settings as for my liquid biopsy samples were used (see section 2.2.15), but CNVs were estimated based on the methylation-enriched data for subsequent CNV correction (`--cnvmethod "MeCap"`), as lcWGS data were not available. A matrix containing all β-values in genome-wide 300-bp windows was generated and subsequently filtered for the PCa tissue-derived DNA methylation marker regions. Principal component analysis (PCA) and hierarchical clustering of β-values across these marker regions were performed for the external liquid biopsy dataset alone, as well as in combination with my liquid biopsy-based methylation dataset.

2.2.16 Correlation analyses

Spearman's rank correlation was used for correlation analyses and was computed using the function `cor.test(x, y, method = "spearman")` of the R package *stats v.4.0.0*[368]. Spearman rank correlation coefficients (ρ, rho) and the significance level (p value) were reported for each comparison. The strength of the correlation was assessed based on Spearman's ρ (rho), evaluating values of 0.0–0.19 as "very weak", 0.2–0.39 as "weak", 0.4–0.59 as "moderate", 0.6–0.79 as "strong" and 0.8–1.0 as "very strong" correlations[407].

2.2.17 Unsupervised clustering with PCA and hierarchical clustering

Unsupervised clustering based on β-values in selected regions (*i.e.,* PCa tissue-derived methylation marker regions) was performed using (i) hierarchical clustering with the `pheatmap()` function of the R package *pheatmap v.1.0.12*, and (ii) PCA with the `prcom()` function of the R package *stats v.4.0.0*. The first and second principal components, representing the largest and second-largest proportions of variance in the data, respectively, were visualized in PCA plots.

2.2.18 Data manipulation

Data manipulation was performed in R v4.0.0[368] using various R packages, including *dplyr v.1.1.1*[373], *tidyverse v.1.3.1*[395], *readr v.2.1.2*[391], *readxl v.1.3.1*[392], *matrixStats v.0.61.0*[381], and *data.table v.1.14.4*[372].

Transformation of variables to z-scores for standardisation and improved comparability was performed with the function `transform()` of the R package *liver v.1.15*[379].

Manipulation and transformation of genomic data were performed with the R packages *plyranges v1.10.0*[386] and *GenomicRanges v.1.42.0*[374].

2.2.19 Statistical analyses

All statistical analyses were performed in R v4.0.0[368] using the R packages *stats v.4.0.0* and *rstatix v.0.7.2*[394]. Results of the statistical tests were considered significant, if the p value was < 0.05, if not otherwise stated. In case of multiple testing, the p values were adjusted by Benjamini Hochberg method[408], and a significance level of < 0.05 for adjusted p values was used in identifying statistically significant results, if not otherwise stated.

Unpaired Wilcoxon rank sum test or paired Wilcoxon signed rank test were used as non-parametric method to compare two independent (e.g., plasma cfDNA features in tumor vs. control samples) or dependent groups (e.g., cfDNA features in plasma vs. matched urine from same individuals), respectively, without assuming normal distribution of the data. The function `wilcox.test()` of the R package *stats v.4.0.0* was used with parameter setting `paired = FALSE/TRUE` for unpaired/paired Wilcoxon testing, respectively.

The non-parametric Kruskal-Wallis test was applied to detect any significant differences across the medians of multiple groups (e.g., cfDNA features in controls vs. localized PCa patients vs. advanced PCa patients), assuming that the data was not normally distributed. Subsequently, Dunn's post hoc test was complemented to perform pairwise comparisons between the groups, and identify significantly different comparisons. The post hoc test included adjustment for multiple testing based on Benjamini-Hochberg method[408]. The function `kruskal.test()` of the R package *stats v.4.0.0* was used for Kruskal-Wallis testing, and the function `dunn_test()` of the R package *rstatix v.0.7.2* was used for post hoc testing.

Fisher's exact test was applied to compare categorical variables (e.g., smoking status) from the questionnaire assessing lifestyle factors between prostate cancer patients and cancer-free controls. The calculation was performed using the function `fisher.test()` of the R package *stats v.4.0.0*.

The non-parametric, two-sample Kolmogorov-Smirnov test was used to compare the median cumulative frequency distributions of plasma or urinary cfDNA fragmentation between tumor or control samples, in order to assess whether the median distributions of the two cohorts differed significantly. The calculation was performed using the function `ks.test()` of the R package *stats v.4.0.0*. For each comparison between two groups, the testing revealed a p value, as well as the D statistic which represented the maximum vertical distance between the two compared cumulative distributions.

Detectability of ctDNA in plasma and urine samples from localized and advanced PCa patients was investigated based on the assessment of genomic and epigenomic cfDNA features. The limit-of-detection for ctDNA detectability was set at the 95th- or 5th-percentile of the cancer-free control cohort for cfDNA biomarkers which were increased or decreased in tumor samples, respectively. Tumor samples with values beyond these defined ctDNA detection thresholds were considered ctDNA-positive.

2.2.20 Multimodal feature integration with generalized linear modeling

Multivariable models were fitted with GLMs applying the function `glm()` of the R package *stats v.4.0.0* with default settings and a binomial family, corresponding to logistic regression. GLMs were established based on clinical parameters (PSA level, T stage, Gleason score) alone or in combination with either binary ctDNA status (ctDNA positivity, yes/no) or quantitative cfDNA features. Continuous predictor variables were standardized with z-score transformation prior to modeling. The binary outcome variable, encoded as a factor, was defined to distinguish advanced PCa from localized PCa, or PCa patients from PSA-elevated controls. Predicted probabilities from the fitted GLMs were obtained with the function `predict.glm()` of the R package *stats v.4.0.0*. Model discrimination was assessed using the area under the receiver operating characteristic (ROC) curve (AUC). ROC curves were calculated using the function `roc()` of the the R package *pROC v.1.18.2*[387], and AUC estimates, accompanied by their 95th-percentile confidence interval (95% CI), were computed using the functions `auc()` and `ci.auc()` of the same R package, respectively. The AUC of the precision-recall curve and the F1 score were calculated as additional performance metrices. The precision-recall AUC (PR-AUC) was computed using the `pr.curve()` function of the R package *PRROC v.1.3.1*[388]. Predicted probabilities were oriented such that lower scores corresponded to the positive class (which was advanced PCa for the comparison advanced vs. localized PCa, and tumor patients for the comparison tumor vs. controls). For ROC analyses, this orientation was accounted for by specifying `direction = "<"` in the `roc()` function. For the calculation of PR-AUC, prediction scores were transformed by taking the complement of the predicted probability to meet the requirement that higher values indicated the positive class. Optimal probability thresholds were determined from the ROC curves using the function `coords(x = “best”, ret = “threshold”)` of the R package *pROC v.1.18.2*, with default settings, which apply Youden’s J statistic to identify the optimal cut-off. Predicted class labels were generated based on this threshold, with the positive class assigned when predicted probabilities were below it. The F1 score was then computed based on these binary predictions using the `F1_Score()` function of the R package MLmetrics v.1.1.3[383]. Differences in the ROC-AUC between GLMs based on clinical parameters only or combined cfDNA-based models were statistically compared using DeLong’s test for correlated ROC curves, implemented with the function `roc.test(method = "delong")` of the R package *pROC v.1.18.2*. Additionally, likelihood ratio tests were implemented using the function `lrtest()` of the R package *lmtest v.0.9.40*[380] to assess whether integration of cfDNA features resulted in a significantly better model fit compared to clinical models alone. The robustness of the model performance was evaluated using bootstrap resampling implemented with the `boot()` function of the R package *boot v.1.3.28*[370], with default settings. A total of 500 bootstrap iterations were generated with nonparametric, ordinary bootstrap applying case resampling with replacement from the original data. The GLM was refitted in each iteration and the AUC was recalculated. The resulting distribution of 500 bootstrap AUC estimates was summarized by the mean AUC and corresponding 95% CI, which were computed using the function `boot.ci()` of the same R package.

2.2.21 Data visualization

Data visualization was performed using the R packages *ggplot2 v.3.4.2* (box plots, line plots, CNV profiles, volcano plots, bar plots and pie charts), *ggcorrplot v.0.1.4.1* and *ggpubr v.0.6.0*[378] (correlation plots), *plotly v.4.10.2* and *ggfortify v.0.4.17* (PCA plots), *ChIPpeakAnno v.3.24.2* (Venn diagrams), *pheatmap v.1.0.12* (heatmaps for hierarchical clustering), and *RColorBrewer v.1.1-3*[390] (coloring of different graphs). Infographics were prepared in BioRender (BioRender, Toronto, Canada).

All quantitative data for comparisons of two or more groups were visualized in boxplots with the median represented as centre lines and the interquartile range displayed as boxes with Tukey whiskers. Dots represented individual samples within the given groups.

Correlations between two features were visualized with scatter plots using the function `ggscatter()` of the R package *ggpubr v.0.6.0*, and a regression line as well as the correlation coefficients (Spearman's ρ and p value) were added.

Within the cfDNA fragmentation analyses, relative and cumulative frequency distributions of plasma and urinary cfDNA fragment lengths were visualized as line plots for individual samples, as well as for the median frequencies of tumor and control samples, respectively.

For the CNV analyses, genome-wide distribution of copy numbers changes (reported as log2 ratios compared to the copy-number-neutral state) in plasma and urinary cfDNA, as well as in gDNA from PCa tissue and buffy coat samples were visualized as CNV profiles, highlighting regional amplifications and deletions. Recurrent alterations in liquid biopsies and/or in PCa tissue samples were also visualized in CNV profiles, showing the frequencies of the occurrence of genome-wide gains and losses in the analyzed samples. The frequencies were scaled from 0 to 1, where 1 indicates that 100% of the samples harbored a gain or loss in a specific 1,000-kb bin.

For the DNA methylation analyses, identified DMRs in liquid biopsy samples from PCa patients compared with cancer-free controls, as well as DMRs identified between PCa tissue and matched buffy coat samples, were visualized using volcano plots. Each dot in the plot represented one 300-bp window (DMR), with its position determined by the magnitude of the difference (logFC) between the mean β-values of the two groups (x-axis), and the statistical significance of the difference, expressed as negative logarithm of the adjusted p value after correction for multiple testing (y-axis). Significant, hypermethylated DMRs (liquid biopsies: $\log FC > 1$; PCa tissue: $\log FC > 2$) were colored in red, and significant, hypomethylated DMRs (liquid biopsies: $\log FC < -1$; PCa tissue: $\log FC < -2$) were colored in blue. Statistical significance was defined as an adjusted $p < 0.05$ for liquid biopsies and adjusted $p < 0.01$ for PCa tissue analysis.

DMRs were annotated for genomic regions and CpG-associated landscapes, and their distributions were visualized in bar plots and pie charts. Different sets of DMRs in plasma or urinary cfDNA were overlapped to identify shared regions with aberrant DNA methylation. The relationships between the assessed region sets, including the numbers of overlapping and non-overlapping regions, were visualized using Venn diagrams.

ROC curves from comparative GLMs based on clinical parameters alone and combined with the binary ctDNA status or quantitative cfDNA features were displayed in joint graphs. Sensitivity and specificity values were previously determined with the `roc()` function from the R package *pROC v.1.18.2*.

2.2.22 Application of artificial intelligence models

The large language model ChatGPT-4o (OpenAI) was utilized as a supportive tool for R coding, including suggestions for code structure and debugging. All code implementation, statistical analyses, and interpretation of the results were independently performed and verified by me. For the GLM analyses, ChatGPT proposed example code, which was subsequently reviewed and adapted by me. ChatGPT also supported the interpretation of the GLM results, which were critically evaluated and methodologically verified by me. Additionally, ChatGPT was used for text formulation, rephrasing, and shortening of individual sentences across various sections of the PhD thesis, as well as for translating the German summary. All original content, analyses, and scientific arguments were independently developed and written entirely by myself. Elicit[356,409] and Perplexity AI were used as supportive tools for literature searches for the Introduction and Discussion sections.

3 Results

3.1 Patient characteristics

A cohort of 72 men with newly diagnosed localized or advanced PCa, and 36 cancer-free controls were recruited at the Department of Urology at Heidelberg University Hospital (Table 13). For each man, up to 8 ml plasma and 10–20 ml urine were collected as part of routine clinical sampling procedures. Accompanying clinical and imaging (MRI) data, as well as laboratory tests of blood values and urinary status were available for comparative analyses. All PCa patients were enrolled at initial PCa diagnosis, without any previous cancer treatment. The majority of PCa patients was diagnosed at localized stage (75%), whereas one quarter harbored lymph node and/or distant metastases (Table 13). The following results report 73 PCa patients, reflecting 73 available sampling timepoints from tumor patients. One PCa patient was included with two sampling timepoints in a timely distance of 12 months. He was initially diagnosed with localized low-risk PCa (Gleason 6) in the first prostate tissue biopsy, but was upgraded one year later to localized intermediate-risk PCa (Gleason 7b) after a second tissue biopsy. Fifty PCa patients underwent prostatectomy after PCa diagnosis, and the pathological assessment of the prostatectomy tissue revealed that half of them harbored an extended tumor size (12 x pT3a, 14 x pT3b). Furthermore, seven prostatectomy patients were diagnosed with lymph node metastases (N1M0) and two patients harbored distant metastases (N1M1). However, the majority of patients with metastatic PCa at initial diagnosis (7 out of 9) did not undergo prostatectomy, two patients with only lymph node metastases and 12 patients with localized PCa, neither.

The control cohort included 16 men with increased PSA levels (PSA > 2 ng/ml) who underwent PCa screening and harbored no signs of malignancy on prostate biopsy. Furthermore, the control cohort was complemented by men with low PSA levels (PSA < 2 ng/ml) who underwent different examinations due to benign urological conditions (kidney or ureteral stones, benign renal cysts, testicular hydrocele, benign prostatic hyperplasia, phimosis, urinary tract obstruction), including ureterorenoscopy (11 men), resection of a hydrocele testis (2 men), transurethral resection of the prostate (2 men), kidney surgery (3 men), installation of double pigtail stent and circumcision (one man, each). It was confirmed that none of the controls harbored an ongoing or a prior cancer disease over the past five years.

PCa patients were significantly older (median age: 66, range: 49–80) than controls (median age: 58, range: 39–74). Thereby, both localized PCa patients (median age: 66, range: 49–80) and advanced PCa patients (median age: 65, range: 58–75) harbored significantly higher age compared to controls (control vs. localized PCa: p = 2.46e-06, adjusted p = 7.39e-06; control vs. advanced PCa: p = 1.84e-04, adjusted p = 2.76e-04), but no significant differences were seen between localized PCa and advanced PCa patients (p = 0.798, adjusted p = 0.798, Figure A1).

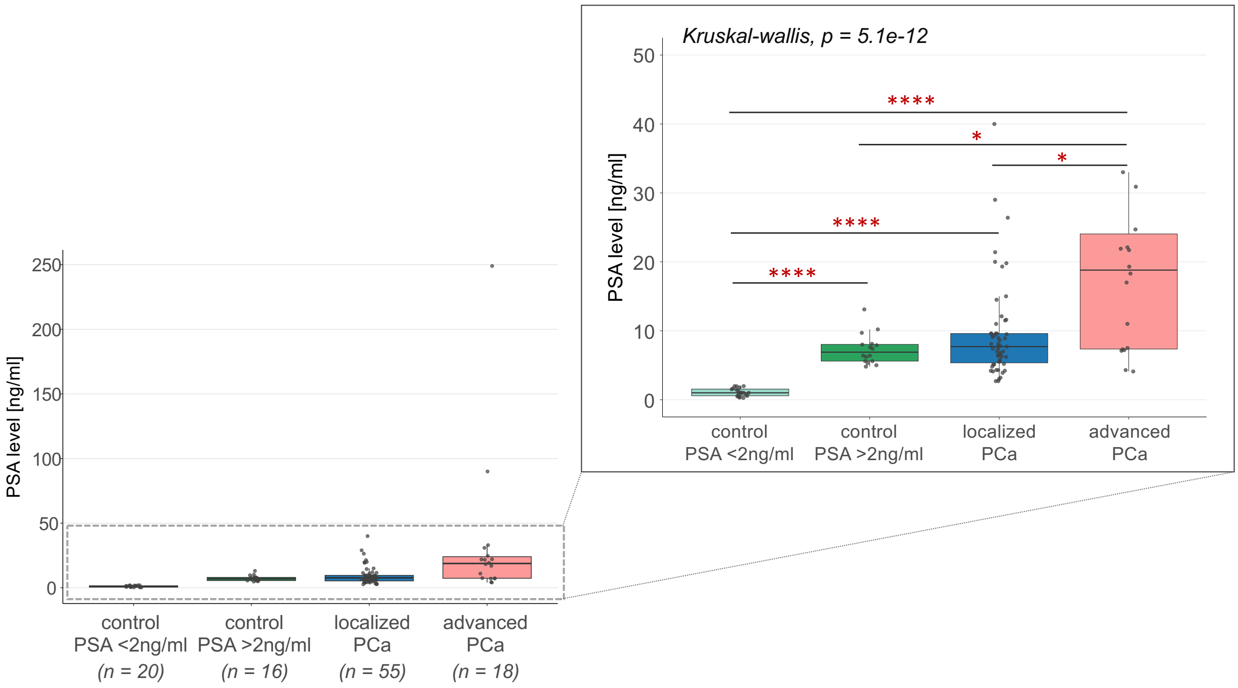
Table 13: Cohort and sample overview.
Overview of the number of PCa patients with localized and advanced disease, as well as cancer-free controls. PCa stages were determined based on UICC (stage I-IV) [11] and D'Amico Risk Classification[97]. Reported Gleason scores were based on prostate biopsy results, or on prostatectomy specimens for PCa patients who underwent prostatectomy. The table additionally summarizes the availability of collected plasma, urine and PCa tissue samples across the cohort. Adapted from Riediger *et al.* [410]

	plasma	**urine**	**PCa tissue^{+}**
localized disease (stage I-III)			
low risk	4	3	0
GS 6 (n = 4)	4	3	0
intermediate risk	42	38*	6
GS 7a (n = 37)	37	33*	6
GS 7b (n = 5)	5	5	0
high-risk	9	9	0
GS 7a/7b (n = 2)	2	2	0
GS 8 (n = 1)	1	1	0
GS 9 (n = 6)	6	6	0
advanced / disseminated disease (stage IV)			
lymph node metastases, N1 M0	9	9	2
GS 7a/7b (n = 4)	4	4	1
GS 8 (n = 1)	1	1	0
GS 9 (n = 4)	4	4	1
distant metastases, M1	9	9	0
GS 7a/7b (n = 3)	3	3	0
GS 8 (n = 3)	3	3	0
GS 9 (n = 2)	2	2	0
GS 10 (n = 1)	1	1	0
control cohort	36	35	0
controls with PSA < 2 ng/ml (n = 20)	20	20	0
controls with PSA > 2 ng/ml (n = 16)	16	15	0

Abbreviations: GS = Gleason score, M0/M1 = absence/presence of distant metastases, N0/N1 = absence/presence of lymph node metastases, n = number; *one patient/sample excluded after QC; $^{+}$matched PCa tissue and buffy coat samples

3.2 PSA measurements and questionnaire-derived PCa risk factors

PSA levels were measured at initial timepoint of diagnosis for PCa patients, and either during PCa screening or at timepoint of sample collection for cancer-free controls. PCa patients harbored significantly higher PSA levels (median PSA level: 8.1 ng/ml, range: 2.7–249 ng/ml) compared to controls (median PSA level: 1.95 ng/ml, range: 0.26–13.1 ng/ml), as determined with statistical testing (p = 4.10e-08). Furthermore, PSA levels were significantly increased in advanced PCa patients (median PSA level: 18.8 ng/ml, range: 4.1–249.0 ng/ml) compared to localized PCa patients (median PSA level: 7.7 ng/ml, range: 2.7–40.0 ng/ml; p = 0.015; adjusted p = 0.015; Figure 9). The control group included 20 men with benign urological diseases and confirmed low PSA levels (median PSA level: 1.0 ng/ml, range: 0.26–1.97 ng/ml), as well as 16 men with increased PSA levels who underwent PCa screening (median PSA level: 6.9 ng/ml, range: 4.8–13.1 ng/ml). The median PSA levels of these two subcohorts within the control group differed significantly from each other (p = 9.79e-06, adjusted p = 1.96e-05, Figure 9). However, both control subgroups showed significantly lower PSA levels compared to advanced PCa patients, and the controls with low PSA levels also harbored significantly lower values compared to localized PCa patients, respectively (Figure 9).



cohort 1	cohort 2	p value	adjusted p value
control, PSA > 2 ng/ml	control, PSA < 2 ng/ml	9.79e-06	1.96e-05
localized PCa	control, PSA < 2 ng/ml	6.34e-10	1.90e-09
localized PCa	control, PSA > 2 ng/ml	0.644	0.644
advanced PCa	control, PSA < 2 ng/ml	2.50e-12	1.50e-11
advanced PCa	control, PSA > 2 ng/ml	0.021	0.025
advanced PCa	localized PCa	0.015	0.022

Figure 9: Distribution of PSA levels in PCa patients and cancer-free controls.
PSA levels measured at the initial timepoint of diagnosis for PCa patients and either at the time of PCa screening or sample collection for controls. The main panel shows the overall distribution of PSA levels, with an additional panel providing a zoomed view, excluding PSA levels > 50 ng/ml. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one individual. The three cohorts were compared using Kruskal-Wallis testing, followed by Dunn's post hoc test. Correction for multiple testing was performed with Benjamini-Hochberg method. Only significant differences are shown in the figures, and a summary of statistical testing results is provided in the table below the figures. Abbreviations: n = number; Significance codes: ****1e-04, ***0.001, **0.01, *0.05

At enrollment, men were asked by a study nurse to complete a survey, covering questions about life style factors, cancer family history (first- and second-degree relatives with PCa), and potential PCa risk factors. These involved consumption of meat and alcohol (days per week), smoking (packyears), physical exercise (hours per week), body mass index (bmi, kg/m^2), infertility or sterilization, and occurrence of urinary tract infection or urinary retention in the last six months.

The survey was completed by 85 men, including 53 PCa patients and 32 cancer-free controls. Incomplete data was available for 21 men, and no data for two cancer-free controls and one PCa patient. Except for the smoking status, no significant differences were seen between PCa patients and controls, or between controls, localized and advanced PCa patients (Table A1, Figure A2). The control cohort involved a significantly higher proportion of smokers ($p = 0.002$, adjusted $p = 0.009$) with a higher number of packyears ($p = 1.03e-04$, adjusted $p = 0.001$) compared to PCa patients.

3.3 Performance assessment of lcWGS and cfMeDIP-seq

3.3.1 Evaluation of sequencing data quality metrics

I performed lcWGS and cfMeDIP-seq for plasma and urine (supernatant) samples from 73 PCa patients and 36 cancer-free controls. Additionally, lcWGS and MeDIP-seq were applied to eight fresh-frozen PCa tissue and matched buffy coat samples from eight PCa patients (Table 13). Plasma samples were available for all men, whereas matched urine samples were missing for four localized PCa patients and one cancer-free control (Table 13).

LcWGS libraries were sequenced to a median of 91 M raw reads (range: 45–141 M) for liquid biopsy samples (plasma and urine) and 77 M raw reads (range: 68–88 M) for tissue and buffy coat samples, resulting in a genome coverage of 2.76x (range: 1.38–4.28x) and 2.32 x (range: 2.07–2.66x), respectively, when considering the size of human genome as 3.3 billion base pairs. After quality filtering and deduplication, in the median 80.3% and 81.8% of the initial sequencing reads remained for lcWGS data of liquid biopsy and tissue/buffy coat samples, finally leading to a median genome coverage of 2.19x (range: 1.07–3.33x) and 1.89x (range: 1.69–2.16x), respectively (Table 14).

(Cf)MeDIP libraries were sequenced to a median of 90 M raw reads (range: 0.075–219 M) for liquid biopsy samples, and 74 M raw reads (range: 50–108 M) for tissue and buffy coat samples (Table 14). A genome-wide coverage was not calculated, since the sequencing reads only covered regions containing 5-methylcytosines which were enriched with the DNA methylation immunoprecipitation procedure. (Cf)MeDIP libraries involved higher duplicate rates in the raw sequencing data compared to matched lcWGS data (Table A2). After quality filtering and deduplication, in the median 66.8% and 72.4% of the initial sequencing reads remained for (cf)MeDIP-seq data of liquid biopsy and tissue/buffy coat samples, resulting in a median of 60 M (range: 0.053–137 M) and 54 M final reads (range: 34–80 M), respectively (Table 14). I excluded one urine sample from a localized PCa patient after sequencing QC, due to a very low number of sequencing reads (74,626 reads) in the cfMeDIP-seq data, and the corresponding lcWGS data was equally excluded.

Differences between lcWGS and (cf)MeDIP-seq data were also seen with regard to the GC content. The final lcWGS data of both liquid biopsy and PCa tissue/buffy coat samples harbored a median GC content of 41–42%, representing the known genome-wide, average GC content of 41% in human genome[411], while the GC content of final (cf)MeDIP-seq data was increased to a median of 46% (Table A2).

Table 14: Raw and filtered sequencing reads in lcWGS and (cf)MeDIP-seq data.
Summary of QC parameters for lcWGS and (cf)MeDIP-seq data derived from plasma and urinary cfDNA, as well as from gDNA of PCa tissue and matched buffy coat samples. Reported metrics include the total number of sequencing. reads before and after quality filtering and deduplication, the total number of filtered reads, and the proportion of final sequencing reads retained following deduplication and quality filtering (%). Adapted from Riediger *et al.* [410]

	total raw reads, M	total filtered reads, M	total final reads, M	final reads, %
cfDNA, lcWGS	91.0 [45.4 –141.2]	18.1 [9.8 –31.4]	72.4 [35.5 –109.8]	80.28 [70.16–84.56]
plasma	88.2 [56.3 –117.0	16.9 [9.9 –23.6]	71.2 [45.3 –95.2]	81.32 [72.54–84.56]
urine	92.3 [45.4 –141.2]	19.2 [9.8 –31.4]	73.2 [35.5 –109.8]	79.51 [70.16–83.36]
cfDNA, cfMeDIP-seq	90.1 [74,626–218.9]	30.7 [21,268–81.7]	60.2 [53,358–137.2]	66.78 [38.55–76.01]
plasma	91.6 [70.2 –200.5]	29.0 [20.2 –64.0]	62.4 [36.2 –136.4]	68.24 [50.82–76.01]
urine	88.4 [74,626–218.9]	31.8 [21,268–81.7]	56.2 [53,358–137.2]	64.04 [38.55–75.45]
gDNA, lcWGS	76.6 [68.2 –87.9]	14.0 [12.2 –16.7]	62.4 [55.8 –71.2]	81.76 [80.78–82.59]
PCa tissue	78.4 [68.2 –83.7]	14.7 [12.5 –16.0]	63.5 [55.8 –68.0]	81.43 [80.78–82.01]
buffy coat	73.3 [70.1 –87.9]	13.3 [12.2 –16.7]	60.1 [57.9 –71.2]	82.07 [81–82.59]
gDNA, MeDIP-seq	74.4 [49.7 –108.0]	20.5 [15.8 –28.1]	54.1 [33.9 –80.0]	72.38 [67.44–74.73]
PCa tissue	73.3 [49.7 –80.2]	20.7 [15.8 –26.1]	52.6 [33.9 –57.7]	70.58 [67.44–74.73]
buffy coat	74.7 [68.3 –108.0]	20.5 [17.8 –28.1]	54.3 [50.3 –80.0]	73.02 [72.07–74.02]

Abbreviations: cfDNA = cell-free DNA, gDNA = genomic DNA, lcWGS = low-coverage whole-genome sequencing, M = Million, (cf)MeDIP-seq = (cell-free) methylated DNA immunoprecipitation sequencing, QC = quality control

3.3.2 Quality assessment of DNA methylation enrichment

I conducted the DNA methylation QC to assess the efficiency of the DNA methylation enrichment and to confirm the quality of the (cf)MeDIP-seq dataset for subsequent analyses. The entire DNA methylation QC workflow consisted of two parts: a laboratory assessment following the library preparation procedure, and a bioinformatic evaluation of the (cf)MeDIP-seq data.

3.3.2.1 Laboratory methylation QC

I performed the initial methylation QC assessment after the preparation of (cf)MeDIP-seq and lcWGS libraries with cfDNA or gDNA input, respectively, quantifying 5mC- and 5C-spike-ins with qPCR. I calculated the recovery from 5mC- and 5C-spike-ins based on the qPCR results, and determined the specificity of the DNA methylation enrichment based on the ratio of the 5mC- and 5C-spike-in recoveries. Distinct differences in the recoveries of 5mC- and 5C-spike-ins were visible in all (cf)MeDIP-seq libraries, suggesting appropriate enrichment of methylated DNA (Figure 10A and B). The median 5mC-spike-in recovery was 78% in plasma and urine samples and 82% in PCa tissue and buffy coat samples, whereas the median 5C-spike-in recovery was distinctly lower in both liquid biopsy samples (0.3%) and PCa tissue/buffy coat samples (1.1%; Figure 10A and B; Table A3). Four plasma and four urine samples harbored a calculated 5mC-spike-in recovery above 100%, which might be associated with deviations in the qPCR efficacy that was set to 100% for all calculations. There was no significant difference in the 5mC-spike-in recoveries between cfDNA and gDNA samples (Figure 10A). However, values for the 5C-spike-in recoveries differed significantly between the two groups (Figure 10B), and consequently, gDNA libraries also harbored significantly lower specificities compared to cfDNA libraries (Figure 10C). Plasma and urine samples reached a median specificity of 0.996, whereas the median specificity of PCa tissue and buffy coat samples was at 0.987 (Table A3). The three QC metrics did not differ significantly between cfMeDIP-seq libraries from tumor patients and controls (Figure A3). No samples were excluded after the laboratory methylation QC.

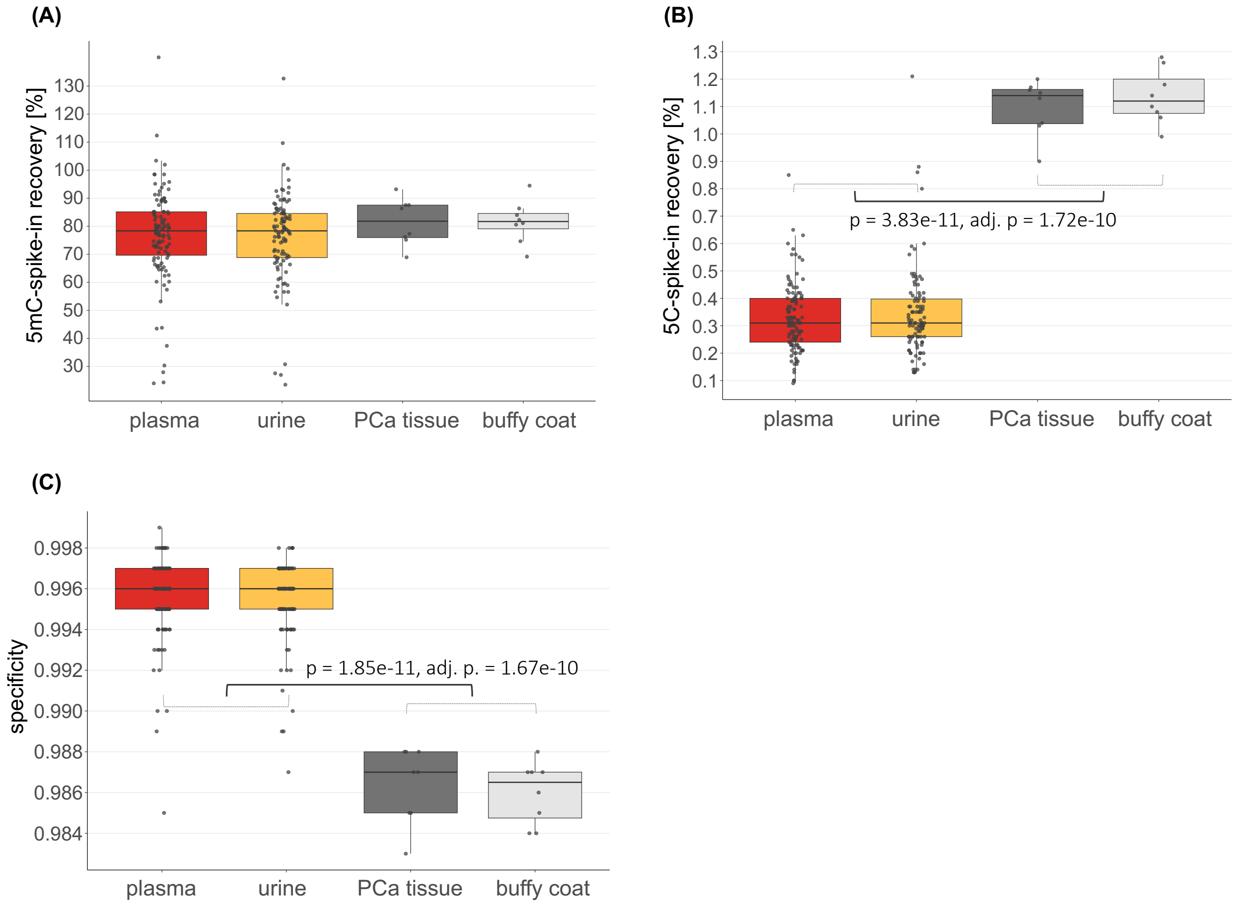


Figure 10: Methylation QC metrics derived from qPCR quantification of 5mC- and 5C-spike-ins.
(A) Distribution of the recovery of the methylated (5mC-) spike-in across plasma, urine, PCa tissue and buffy coat samples. **(B)** Distribution of the recovery of the unmethylated (5C-) spike-in across plasma, urine, PCa tissue and buffy coat samples. **(C)** Distribution of the specificity of the methylation enrichment across plasma, urine, PCa tissue and buffy coat samples. **(A–C)** Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one sample. Statistical comparisons between results derived from plasma and urinary cfDNA and those derived from gDNA from PCa tissue or buffy coat samples were performed using the Wilcoxon rank sum test. Correction for multiple testing was performed using the Benjamini-Hochberg method. Only significant differences are shown. Abbreviations: adj. p. = adjusted p value, 5C = unmethylated, 5mC = methylated.

3.3.2.2 Bioinformatic methylation QC

The bioinformatic methylation QC of the (cf)MeDIP-seq data involved the assessment of three metrics (*i.e.*, saturation, CpG coverage, and methylation enrichment score) which reflected different quality characteristics of the methylation data, as outlined in the following.

Saturation analysis

The saturation analysis evaluated whether the (cf)MeDIP-seq data contained enough short reads to achieve saturated and reproducible methylation profiles with regard to the reference genome. I aimed to achieve a saturation value > 0.9 for sufficient quality, following the suggestions from Shen and colleagues[333]. All liquid biopsy, PCa tissue and buffy coat samples reached this threshold (Figure 11A, Table A4), and no sample had to be excluded because of bad saturation quality.

Median saturation was significantly increased for MeDIP-seq libraries based on sheared gDNA (PCa tissue and buffy coat samples) compared to cfMeDIP-seq libraries based on plasma or urinary cfDNA (p-value = 6.83e-05, adjusted p = 1.23e-04; Figure 11A). There was no significant difference between matched PCa tissue and buffy coat samples (p = 0.547). However, saturation was significantly different in cfMeDIP-seq libraries from all 102 urine samples compared to all 109 plasma samples (p = 9.69e-18; adjusted p = 7.75e-17), as well as between matched plasma and urine samples (p = 9.59e-15; adjusted p= 7.67e-14). There was no significant difference in the saturation between tumor and control samples, in either plasma or urine (Figure A4).

Methylation enrichment score

I calculated the methylation enrichment score to assess the representation of CpGs in the sequencing data from methylation-enriched (cf)DNA fragments compared to CpGs in the reference genome, thereby evaluating the efficacy of the methylation enrichment. The assessment of the methylation enrichment score did not result in the exclusion of any liquid biopsy, PCa tissue or buffy coat samples.

The median enrichment score was 2.42 (range: 1.88–3.18) in cfMeDIP-seq data from plasma samples, and 2.25 (range: 1.82–2.84) in sequencing data from urine samples (Figure 11B; Table A4). For PCa tissue and matched buffy coat samples, the median enrichment score was 2.41 (range: 2.34–2.62) and 2.31 (range: 2.23–2.58), respectively (Figure 11B; Table A4). There was no significant difference between the enrichment score values from MEDIP-seq data derived from sheared gDNA compared to cfMeDIP-seq data derived from plasma or urinary cfDNA (Figure 11B). Equally, there was no significant difference between matched PCa tissue and buffy coat samples (p = 0.195). However, the methylation enrichment score was significantly increased in cfMeDIP-seq data from all plasma (n=109) compared to all urine samples (n = 102; p = 2.46e-11; adjusted p = 9.84e-11), and it also significantly differed between matched plasma and urine samples (n = 102, each; p = 1.58e-10, adjusted p = 6.32e-10). There was no significant difference in the methylation enrichment scores in plasma and urinary cfMeDIP-seq data between tumor and control samples (Figure A4).

The median enrichment score values from lcWGS data were distinctly decreased compared to the corresponding (cf)MeDIP-seq data. LcWGS data from plasma and urine samples both showed a median enrichment score of 1.14, while lcWGS data from matched PCa tissue and buffy coat samples showed median enrichment scores of 0.98 and 0.97, respectively (Table A4). These pronounced differences in the methylation enrichment scores between matched (cf)MeDIP-seq and lcWGS data were consistent with differences observed in laboratory QC metrics (i.e., 5mC- and 5C-recovery, specificity) of the corresponding libraries, equally reflecting the methylation enrichment efficacy.

CpG coverage

I conducted the assessment of the CpG coverage in the (cf)MeDIP-seq data to assess efficacy and specificity of the methylation enrichment, since DNA methylation occurs predominantly at CpG nucleotides[185]. Thereby, I investigated three parameters for each sample (Figure 11). The proportion of CpGs that were covered by more than five reads and the proportion of covered CpGs from all present CpGs in the reference genome were informative about the methylation enrichment efficacy. The proportion of reads without the CpG pattern was used as an indicator of the specificity of the methylation enrichment. The assessment of the CpG coverage did not result in the exclusion of any liquid biopsy, PCa tissue or buffy coat samples.

CfMeDIP-seq data from plasma and urinary cfDNA covered in the median 78.6% (range: 54.4–88.5%) and 80.1% (range: 61.1–93.1%) of CpGs from the human genome, respectively (Figure 11C; Table A4). Of these total CpGs, 37.2% (range: 13.8–57.7%) in plasma and 34.9% (range: 11.2–65.1%) in urine were covered by more than five sequencing reads (Figure 11D; Table A4). In contrast, a median of 3.6% and 5.1% of the sequencing reads did not contain CpGs in plasma and urinary cfMeDIP-seq data, respectively, while minimum and maximum values ranged from 2–14% (Figure 11E, Table A4). For all three parameters, no significant differences between tumor and control samples were identified in either plasma or urine, but results differed significantly between plasma and urinary cfDNA (Figure A4). Corresponding lcWGS data from plasma and urine samples harbored more than 20% of reads without CpG pattern, whereas the proportion of CpGs covered by more than five reads was in the median below 10% (Table A4).

MeDIP-seq data from PCa tissue and buffy coat samples showed significantly lower proportions of reads without CpG pattern (PCa tissue, median: 2.3%, range: 1.5–2.5%; buffy coat, median: 1.7%, range 1.4–3.1%), and of all covered CpGs from human genome (PCa tissue, median: 75.5%, range 69.9–77.4%; buffy coat, median: 76.8%, range: 74.2–81%) compared with liquid biopsy samples (Figure 11C and E, Table A4). The proportion of CpGs covered by more than five sequencing reads did not differ significantly from plasma and urinary cfMeDIP-seq data (PCa tissue, median: 34.7%, range: 24.7–37.7%; buffy coat, median: 37.0%, range: 34.7–46.9%; Figure 11D, Table A4).

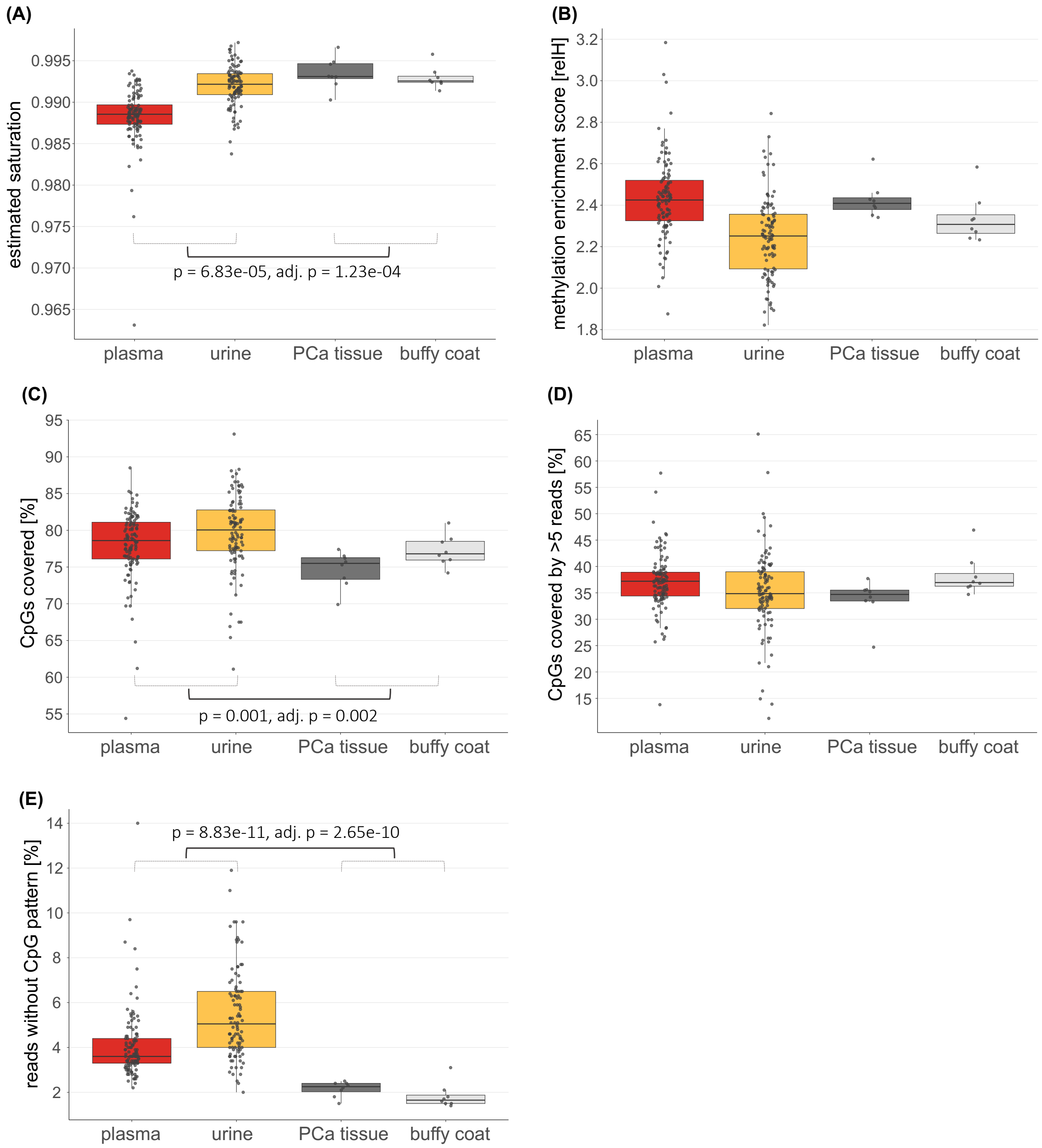


Figure 11: Bioinformatic methylation QC metrics derived from (cf)MeDIP-seq data.
(A) Distribution of the estimated saturation, derived from Pearson correlation coefficients obtained in the saturation analysis, across plasma, urine, PCa tissue and buffy coat samples. **(B)** Distribution of the relative methylation enrichment score (relH) across plasma, urine, PCa tissue and buffy coat samples. **(C)** Distribution of the proportion of covered CpGs relative to all CpGs present in the human genome (CpGs covered) across plasma, urine, PCa tissue and buffy coat samples. **(D)** Distribution of the proportion of CpGs covered by more than five reads across plasma, urine, PCa tissue and buffy coat samples. **(E)** Distribution of the proportion of reads lacking a CpG pattern across plasma, urine, PCa tissue and buffy coat samples. **(A–E)** Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one sample. Statistical comparisons between results derived from plasma and urinary cfDNA and those derived from gDNA from PCa tissue or buffy coat samples were performed using the Wilcoxon rank sum test. Correction for multiple testing was performed using the Benjamini-Hochberg method. Only significant differences are shown.

3.4 Insert size–based analysis of plasma and urinary cfDNA fragmentation

I assessed plasma and urinary cfDNA fragmentation based on the distribution of insert sizes extracted from paired-end lcWGS and cfMeDIP-seq data. Characteristic differences were evident between plasma and urinary cfDNA fragmentation, as well as between the insert size distributions of lcWGS or cfMeDIP-seq data (Figure 12). Additionally, plasma and urinary cfDNA fragmentation profiles differed between samples from cancer-free controls, localized PCa and advanced PCa patients (Figure 12).

3.4.1 Plasma and urinary cfDNA exhibited characteristic fragment size distributions

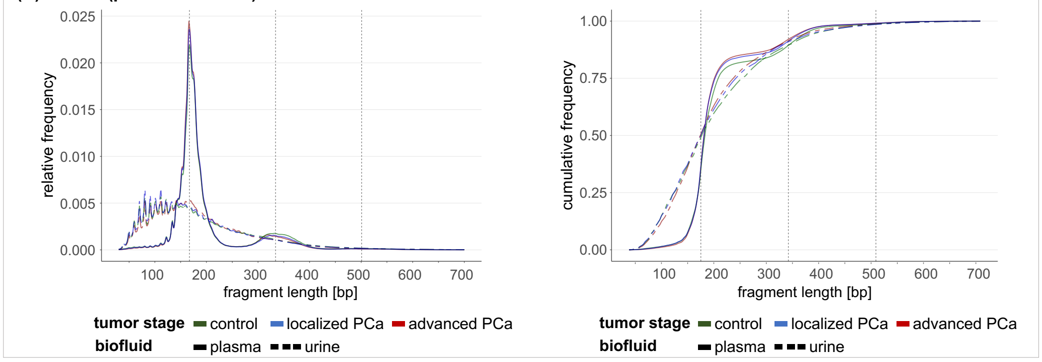
Plasma cfDNA

Plasma cfDNA fragmentation profiles based on the relative frequency distributions of insert sizes derived from lcWGS and cfMeDIP-seq data harbored a prominent peak at ~167 bp fragment length which is known to be associated with the DNA wrapped around the nucleosome plus linker DNA[246]. Additionally, a smaller second peak at ~334 bp and a marginal third peak at ~501 bp fragment lengths were seen (Figure 12A and B), displaying multiples of the first peak. Plasma cfDNA fragments shorter than 150 bp harbored a modest oscillation pattern with 10-bp periodicity. These characteristic fragmentation profiles were evident in both lcWGS and cfMeDIP-seq data (Figure 12C). In cfMeDIP-seq, the height of the first peak was decreased, while the second laddering peak was increased compared to lcWGS. The cumulative frequency distributions of plasma cfDNA fragment lengths from both lcWGS and cfMeDIP-seq data showed characteristics of a sigmoid curve, with an intermediate plateau starting at ~200–220 bp. The curve increased again at 300 bp until it reached a final plateau at ~400 bp fragment length. For lcWGS, the cumulative frequency at the first plateau was already 0.8–0.9 (Figure 12C). In cfMeDIP-seq, the curve was initially shifted by ~10 bp toward longer fragment sizes, but also reached the first plateau at ~200 bp fragment length, with a lower cumulative frequency of 0.55–0.65. The second increase after the plateau showed a steeper slope in cfMeDIP-seq compared to lcWGS (Figure 12C).

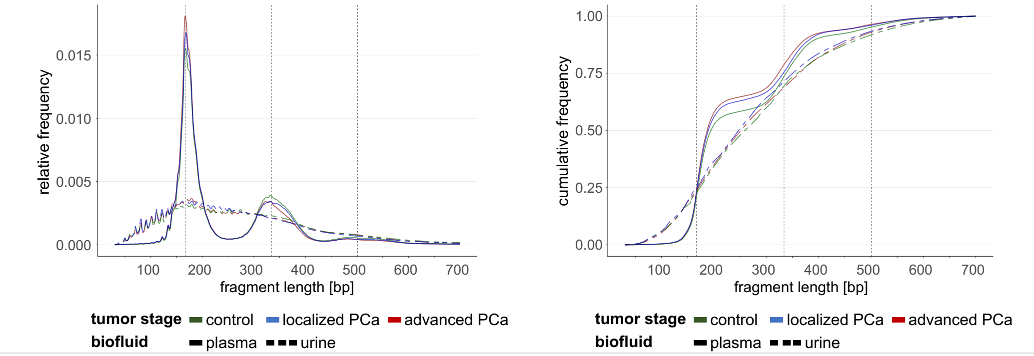
Urinary cfDNA

Urinary cfDNA fragmentation profiles based on the relative frequency distributions of insert sizes derived from lcWGS and cfMeDIP-seq data harbored an elongated peak which ranged from ~30–450 bp, but did not show a prominent peak at ~167 bp known from plasma cfDNA (Figure 12A and B). Urinary cfDNA harbored a high proportion of short fragments < 100–150 bp and a distinct oscillation pattern with 10-bp periodicity. Furthermore, the oscillation pattern was extended to urinary cfDNA fragments with 150–300 bp length. For cfMeDIP-seq, the relative frequency distribution was broader and flatter compared with lcWGS, extending to 500–600 bp fragment length (Figure 12D). A slight increase was visible at ~334 bp, which resembled the increased second peak in cfMeDIP-seq data from plasma samples. The cumulative frequency distributions of urinary cfDNA fragment lengths from both lcWGS and cfMeDIP-seq data also showed characteristics of a sigmoid curve, but lacking an intermediate plateau (Figure 12D). For lcWGS data, the steepest slope occurred at ~30–167 bp, and the cumulative frequency reached 0.5 at 167 bp fragment length. In comparison, the cumulative frequency of plasma cfDNA fragmentation based on lcWGS data was 0.375 at the same fragment length, and a cumulative frequency of 0.5 was reached at ~170–175 bp. For cfMeDIP-seq data from urine samples, the cumulative frequency distribution was flatter and the sigmoidal increase was more elongated compared to lcWGS (Figure 12D). Plasma and urinary cfDNA fragmentation profiles based on cfMeDIP-seq converged at 167 bp fragment length, both exhibiting a cumulative frequency of 0.25 (Figure 12B).

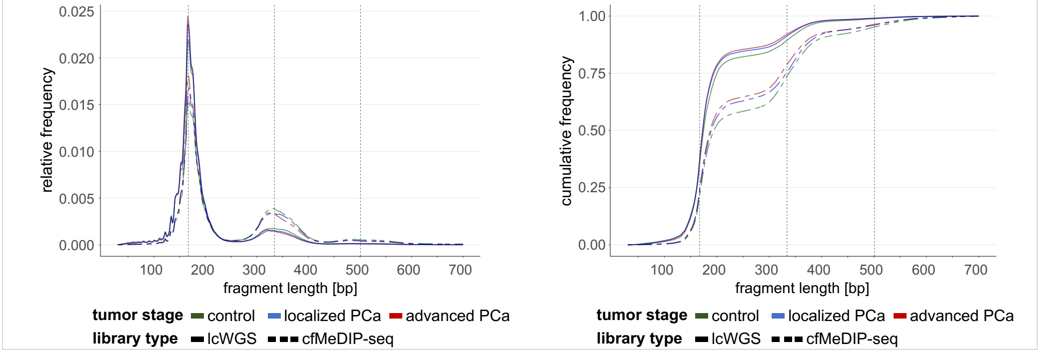
(A) lcWGS (plasma vs. urine)



(B) cfMeDIP-seq (plasma vs. urine)



(C) plasma (lcWGS vs. cfMeDIP-seq)



(D) urine (lcWGS vs. cfMeDIP-seq)

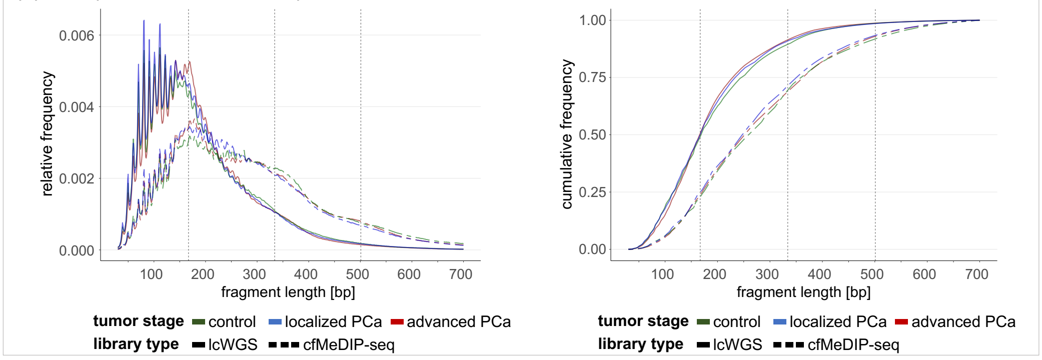


Figure 12: Comparison of plasma and urinary cfDNA fragmentation profiles derived from lcWGS or cfMeDIP-seq data.

(A-D) CfDNA fragmentation profiles shown as group-wise median profiles across samples from localized PCa patients, advanced PCa patients, and controls. Y-axis: relative frequencies (left panels) and cumulative frequencies (right panels) of cfDNA fragments of a given length (x-axis), relative to all fragments within the 30–700 bp range. Vertical dotted grey lines indicate 167 bp and its multiples, 334 bp (2x 167 bp) and 501 bp (3x 167 bp). **(A–B)** Comparison of plasma and urinary cfDNA fragmentation profiles derived from **(A)** lcWGS data and **(B)** cfMeDIP-seq data, shown in a single plot (plasma: solid line; urine: dashed line). **(C–D)** Comparison of cfDNA fragmentation profiles derived from lcWGS and cfMeDIP-seq data in **(C)** plasma and **(D)** urine, shown in a single plot (lcWGS: solid line; cfMeDIP-seq: dashed line).

3.4.2 Plasma and urinary cfDNA fragmentation profiles differed between PCa patients and controls

Next, I evaluated plasma and urinary cfDNA fragmentation patterns in tumor and control samples. The subsequent analyses focused on lcWGS data, which represented the complete cfDNA fragment length distributions without methylation-based enrichment. Plasma and urinary cfDNA fragmentation profiles from cancer-free controls, localized, and advanced PCa patients harbored the characteristics described above, but also revealed differences among the three groups.

Plasma cfDNA

In plasma cfDNA fragmentation profiles with a typical peak at ~167 bp based on the relative frequency distributions of fragment sizes, tumor samples harbored a 1.08-fold higher peak compared to controls (Figure 13A; Table A5). Advanced PCa patients showed the highest peak (maximum relative frequency (max. rel. freq.) = 0.025 at 166 bp) which was in the median 1.12-fold higher as the one from controls (max. rel. freq. = 0.022 at 167 bp), and 1.04-fold higher compared to localized PCa patients (max. rel. freq. = 0.024 at 167 bp). In contrast, control samples harbored a higher second and third peak (max. rel. freq. = 0.0018 [332 bp] and 0.00019 [415 bp]) compared to both localized PCa patients (max. rel. freq. = 0.00157 [330 bp] and 0.00016 [415 bp]) and advanced PCa patients (max. rel. freq. = 0.00161 [320 bp] and 0.00014 [415 bp]; Figure 13B; Table A5). Comparing the heights of the first and the second peak in the plasma cfDNA fragmentation profiles of control samples, the first peak was in the median 12.38-fold higher than the second peak. For tumor samples, this difference was further pronounced (15.24-fold). The first peak was located at 166 bp and 167 bp in plasma samples of advanced PCa patients, as well as controls and localized PCa patients, respectively. The location of the second peak was progressively shifted toward shorter fragment lengths in tumor samples, from 332 bp in controls to 330 bp in localized PCa and 320 bp in advanced PCa, relative to the theoretically expected size of 334 bp (2x 167 bp). The fragment length of 250 bp corresponded to the theoretical vertex between the first peak at 167 bp and the second peak at ~334 bp. Tumor samples harbored a higher proportion of cfDNA fragments < 250 bp, relative to the entire fragment length range (30–700 bp), compared with control samples. This was also visible in the cumulative frequency distributions, where 85.4% and 84.6% of all fragments were covered within the fragment range of 30–250 bp in plasma samples from advanced and localized PCa patients, respectively, whereas 82.2% were covered in the same range in control samples (Figure 13A). The cumulative frequency distributions of tumor and control samples diverged near 167 bp, with the median curve of control samples exceeding that of tumor samples in the 30–150 bp range, but increasing more slowly beyond 167 bp (Figure 13A).

Differences of plasma cfDNA fragmentation in tumor and control samples were also reflected in the statistical comparison of the median cumulative frequency distributions using Kolmogorov-Smirnov testing. The comparison revealed significant differences between the two groups, when assessing the whole fragment length range (30–700 bp: D statistic = 0.091, p = 0.008), as well as when separately considering shorter fragments associated with the first peak (30–250 bp: D statistic = 0.167, p = 0.004) or longer fragments (250–700 bp: D statistic = 0.113, p = 0.006), respectively. No significant difference between the two groups were observed for cumulative frequency distributions within the 30–150 bp fragment range, associated with the 10-bp periodic oscillation pattern (D statistic = 0.066, p = 0.954).

The plasma cfDNA fragmentation analysis initially assessed distributions of relative and cumulative frequencies which were calculated using the entire fragment length range (30–700 bp) as the reference. To evaluate whether this approach potentially masked differences in the distributions of shorter or longer cfDNA fragments between tumor and control samples, I performed additional analyses for the subranges 30–250 bp and 250–700 bp. The subrange of 30–250 bp included the entire first peak up to its theoretical vertex (~250 bp), prior to the onset of the second peak at approximately 334 bp. The second subrange covered longer fragments (250–700 bp), including the second peak and a potential third peak, as multiples of the first one. Relative and cumulative frequencies of all fragment lengths were re-calculated in relation to the respective subranges. When considering the subrange of 30–250 bp, the height of the first peak increased for both tumor and control samples (Figure A5). This increase was stronger for controls compared to PCa patients, while the first peak was still higher in tumor samples, and the differences in the distributions of the two groups were less pronounced. The cumulative frequency distributions equally showed that the curves of tumor and control samples approached each other. Kolmogorov-Smirnov testing revealed no significant differences between the median cumulative frequency distributions of tumor and control samples anymore (30–250 bp: D statistic = 0.054, p = 0.900). In the subrange analysis of 250–700 bp fragment length, the height of the second peak equally increased for both tumor and control samples (Table A5; Figure A5). This increase was more pronounced compared to the overall increase of the first peak observed in the 30–250 bp subrange (30–250 bp: max. rel. freq. < 2-fold increase; 250–700 bp: max. rel. freq. > 5-fold increase). The third peak also became more prominent in the subrange analysis of 250–700 bp (Figure A5). Comparing tumor and control samples, the height of the second peak increased less in control samples. As a result, tumor samples harbored a similar or slightly higher peak (1.002-fold) than control samples, opposite to pattern observed across the full fragment length range (30–700 bp). Similar to the subrange for shorter cfDNA fragments, cumulative frequency distributions of tumor and control samples approached each other, and Kolmogorov-Smirnov testing revealed no significant differences between the median distributions of the two groups (250–700 bp: D statistics = 0.049, p = 0.655).

In summary, plasma cfDNA fragmentation profiles were characterized by a prominent peak at ~167 bp, as well as a smaller second and third peak at multiple fragment lengths of the first one. Additionally, an oscillation pattern with 10-bp periodicity within the 30–150 bp fragment length range was visible. The single fragmentation profiles of tumor samples deviated from the median relative frequency distribution of controls, but also varied among each other. Focusing on the prominent peak at ~167 bp in plasma cfDNA, most of the tumor samples harbored a higher summit compared to the median of controls (Figure 13C). Plasma samples from advanced PCa patients showed in the median the highest peak of all three groups, but there were also single plasma samples from localized PCa patients (n = 4) with a prominent peak that exceeded the one from advanced PCa patients (Figure 13C).

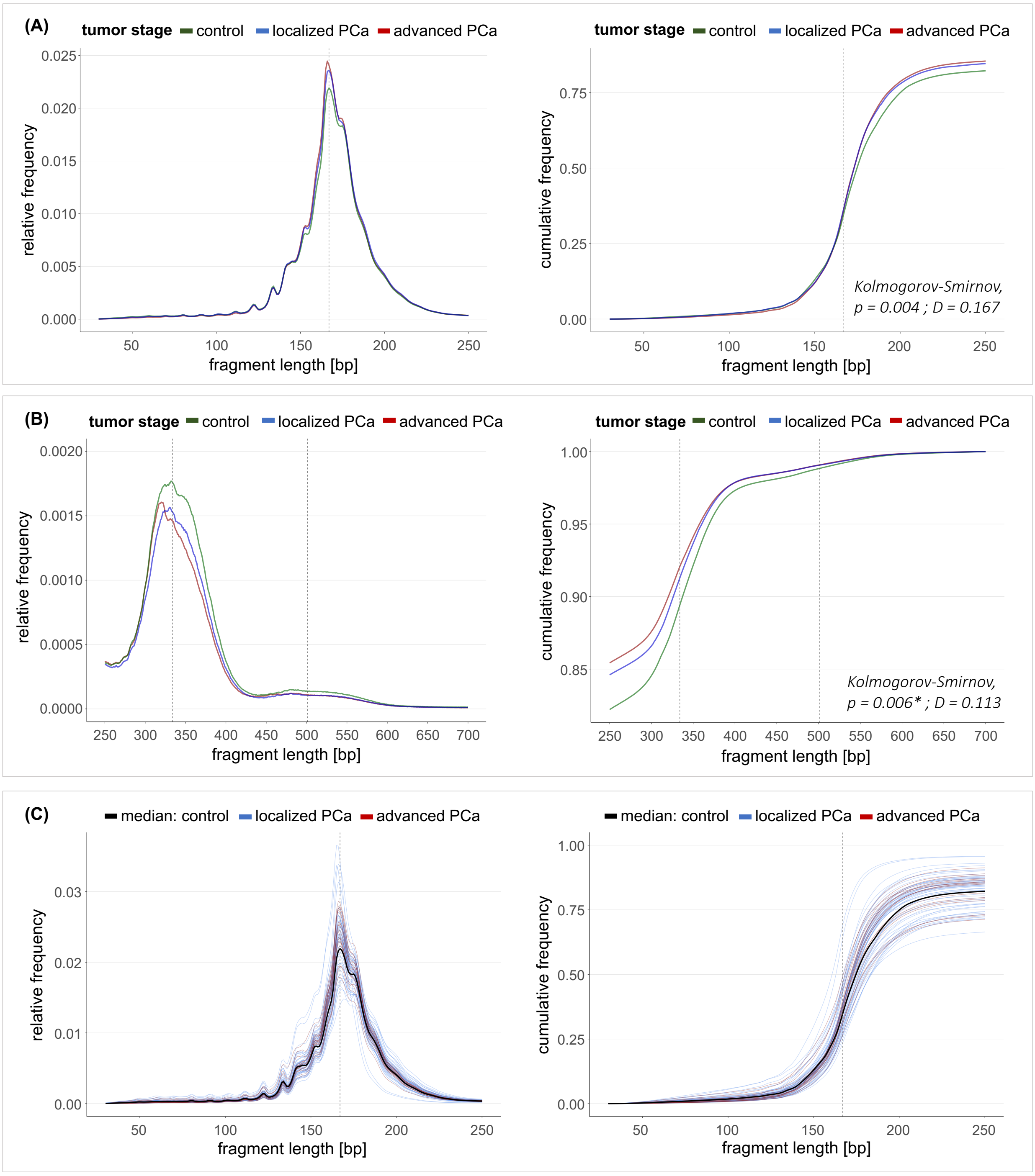


Figure 13: Relative and cumulative frequency distributions of plasma cfDNA fragmentation.
(A–B) Plasma cfDNA fragmentation profiles shown as group-wise median profiles across all samples from localized PCa patients, advanced PCa patients, and controls. Right panel: Median cumulative distributions of tumor samples and control samples were compared with Kolmogorov-Smirnov testing. **(A)** Fragment length distributions within the 30–250 bp range. **(B)** Fragment length distributions within the 250–700 bp range. **(C)** Individual plasma cfDNA fragmentation profiles of localized PCa and advanced PCa patients shown alongside the median fragmentation profile of all controls within the 30–250 bp range. **(A–C)** Y-axis: relative frequencies (left panels) and cumulative frequencies (right panels) of cfDNA fragments with a given length (x-axis) relative to all fragments within the 30–700 bp range. Vertical dotted grey lines indicate 167 bp and its multiples, 334 bp (2x 167 bp) and 501 bp (3x 167 bp). Abbreviations: D = distance (D statistic); *approximate p values in the presence of ties (Kolmogorov-Smirnov testing). Adapted from Riediger *et al.*[410]

Urinary cfDNA

Urinary cfDNA fragmentation profiles exhibited an elongated peak extending to fragment lengths of 400 bp or longer, without distinct laddering peaks. However, a small peak at ~167 bp fragment length, overlapping with the elongated peak, was evident in the median relative frequency distribution of urine samples from advanced PCa patients (Figure 14A). The median relative frequency distributions of urine samples from both localized PCa patients and controls showed no distinct peak, but revealed local maxima associated with the oscillation pattern. Within the 150–170 bp fragment range, the median relative frequency distributions of the three groups showed the greatest divergence, also reflected by their maximum relative frequency values. Advanced PCa samples exhibited the highest maximum relative frequency (max. rel. freq. = 0.0053 at 169 bp), which was 1.11-fold higher than control samples (max. rel. freq. = 0.0047 at 160 bp) and 1.05-fold higher than localized PCa samples (max. rel. freq. = 0.0050 at 151 bp). At ~200–210 bp, the distributions of all three cohorts began to converge again. In the fragment length range from 150–300 bp, a similar oscillation pattern as for smaller fragments < 150 bp, albeit with lower amplitude, was visible for all three cohorts (Figure 14A and B).

With regard to the cumulative frequency distributions, all three cohorts displayed similar courses with minor deviations. As observed in plasma, the proportion of cfDNA fragments < 250 bp relative to all fragments (30–700 bp) was higher in tumor than in control samples (Figure 14A). For urine samples from advanced and localized PCa patients, 79.9% and 78.8% of all fragments, respectively, were covered within the 30–250 bp fragment range, compared with 75.7% in control samples (Figure 14A). Up to approximately 150 bp, the cumulative frequency curves of urine samples from localized PCa patients and controls progressed closely together, whereas the curve for advanced PCa samples trended lower (Figure 14A). Near 150 bp, curves for advanced PCa patients and controls intersected, and the curve for control samples subsequently trended below those from localized and advanced PCa samples, which remained more closely aligned (Figure 14A). From approximately 400 bp fragment length onward, the median cumulative frequency curve of control samples gradually approached those of the tumor samples (Figure 14B).

In contrast to plasma cfDNA, Kolmogorov-Smirnov testing revealed no significant differences between the cumulative frequency distributions of tumor and control samples across any of the evaluated fragment length ranges (30–700 bp: D statistic = 0.024, p = 0.991; 30–150 bp: D statistic = 0.033, p = 1; 30–250 bp: D statistic = 0.068, p = 0.689; 30–400 bp: D statistic = 0.043, p = 0.881; 250–700 bp: D statistic = 0.036, p = 0.939).

I equally performed subrange analyses for urinary cfDNA fragmentation, assessing fragments within the ranges 30–250 bp, 30–400 bp and 250–700 bp. The subrange of 30–400 bp was additionally evaluated, as it enclosed the fragment lengths of the entire elongated peak observed in urinary cfDNA. When considering the subranges of 30–250 bp and 30–400 bp, relative frequency distributions from tumor and control samples remained largely consistent with those observed in the assessment of the entire fragment length range (Figure A6). Comparing the three different cohorts, the median relative frequency distributions from controls and localized PCa samples approached each other, whereas the distribution from advanced PCa samples showed greater deviation, particularly evident by a more pronounced maximum at 162 bp. Consistent with the full-range analysis, Kolmogorov-Smirnov testing revealed no significant differences between the cumulative frequency distributions of tumor and control samples (30–150 bp: D statistic = 0.050, p = 0.998; 30–250 bp: D statistic = 0.027, p = 1). For the subrange of longer urinary cfDNA fragments (250–700 bp), relative frequency distributions from tumor and control samples also remained largely consistent with those observed in the assessment of the entire fragment length range. Cumulative frequency curves showed reduced deviation between tumor and control samples, and Kolmogorov–Smirnov testing revealed no significant difference between the median distributions of the two cohorts (250–700 bp: D statistic = 0.022, p = 1).

In summary, urinary cfDNA fragmentation profiles were characterized by an elongated peak, proceeding until 400–500 bp, as well as a distinct oscillation pattern within the 30–150 bp fragment length range and for longer fragments until ~300 bp. The fragmentation profiles of tumor samples deviated from the median relative frequency distribution of controls, and also showed variation among individual tumor samples (Figure 14C). In urinary cfDNA from advanced PCa patients, the median relative frequency distribution showed an additional peak at ~167 bp that overlapped with the elongated peak. This feature was not prominently observed in the median distributions of urine samples from localized PCa patients or controls. However, individual urine samples from localized PCa patients and four control samples also displayed a peak at ~167 bp, with varying prominence.

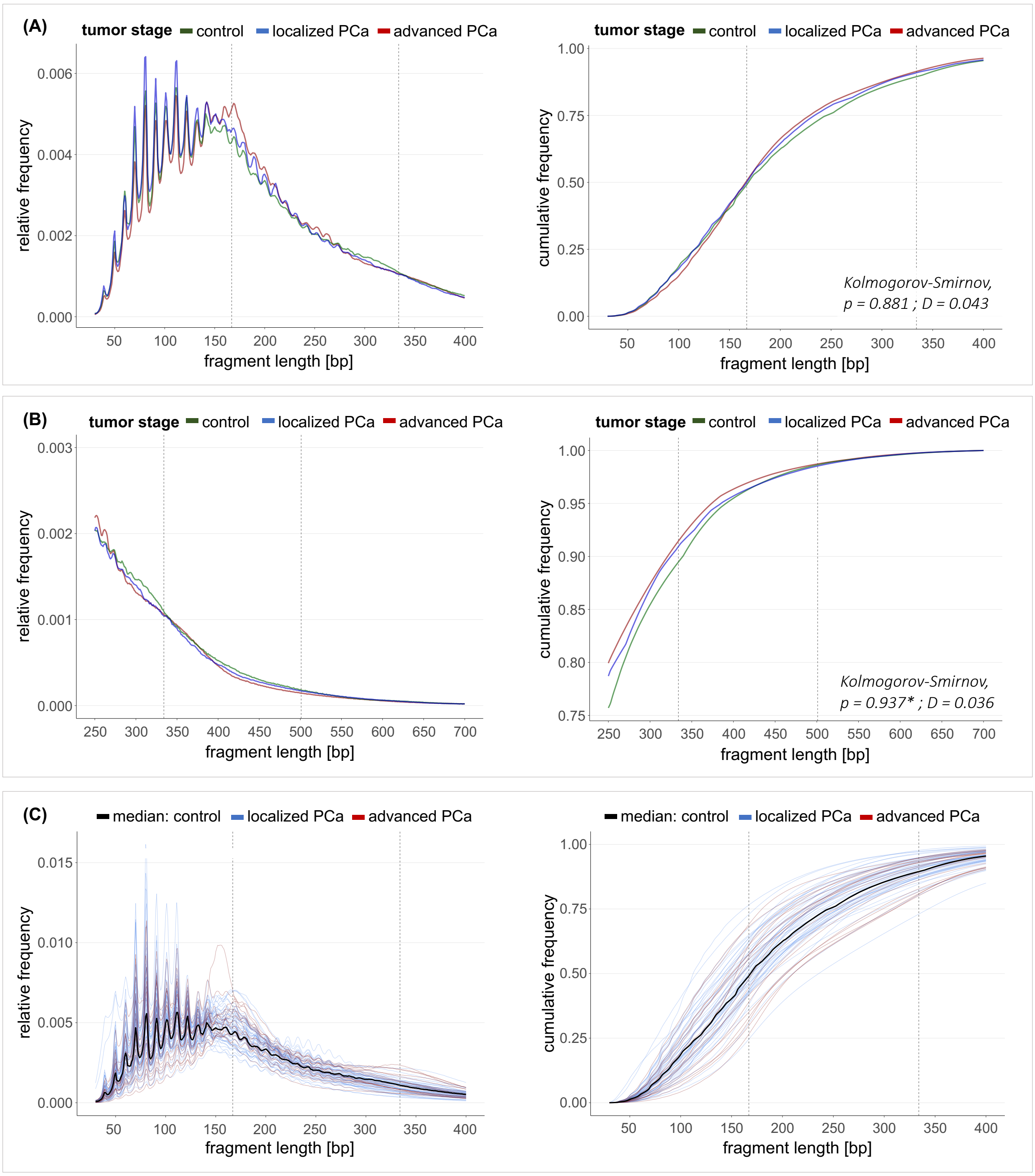


Figure 14: Relative and cumulative frequency distributions of urinary cfDNA fragmentation.
(A–B) Urinary cfDNA fragmentation profiles shown as group-wise median profiles across all samples from localized PCa patients, advanced PCa patients, and controls. Right panel: Median cumulative distributions of tumor samples and control samples were compared with Kolmogorov-Smirnov testing. **(A)** Fragment length distributions within the 30–400 bp range. **(B)** Fragment length distributions within the 250–700 bp range. **(C)** Individual urinary cfDNA fragmentation profiles of localized PCa and advanced PCa patients shown alongside the median fragmentation profile of all controls within the 30–400 bp range. **(A–C)** Y-axis: relative frequencies (left panels) and cumulative frequencies (right panels) of cfDNA fragments with a given length (x-axis) relative to all fragments within the 30–700 bp range. Vertical dotted grey lines indicate 167 bp and its multiples, 334 bp (2x 167 bp) and 501 bp (3x 167 bp). *approximate p values in the presence of ties (Kolmogorov-Smirnov testing). Adapted from Riediger *et al.*[410]

3.4.3 Quantitative metrics reflected characteristics of plasma and urinary cfDNA fragmentation

I extracted different quantitative features from plasma and urinary cfDNA fragmentation distributions derived from lcWGS data for comparative assessment between tumor and control samples. These included descriptive statistical parameters (e.g., maximum relative frequency, mean fragment length, modal fragment length), the proportion of different fragment length ranges compared to entire fragment length range (30–700 bp), ratios between pairs of fragment length range proportions, and the assessment of the 10-bp periodic oscillation pattern in 30–150 bp and 150–300 bp fragment length ranges.

3.4.3.1 Descriptive statistical parameters

The assessment of descriptive statistical parameters derived from plasma and urinary cfDNA fragmentation included the calculation of the modal, median and mean fragment lengths from the relative frequency distributions. In addition, the maximum relative frequency was determined, along with the corresponding fragment length and the cumulative frequency. Further analyses included the determination of the fragment length at a cumulative frequency of 0.5, as well as the relative and cumulative frequencies at fragment lengths of 167 bp and 334 bp, respectively.

Plasma cfDNA

For plasma cfDNA fragmentation, tumor samples showed a higher maximum relative frequency than controls, with a modal fragment length of 167 bp observed in both localized and advanced PCa patients, corresponding to the typical fragmentation peak (Table A6). The cumulative frequency at the modal fragment length was likewise higher in PCa patients, whereas the fragment length at a cumulative frequency of 0.5 was shorter in PCa patients (174 bp) than in controls (176 bp). Control samples harbored a longer mean fragment length (205 bp) compared with PCa patients overall (198 bp), with advanced PCa patients displaying the shortest mean fragment length (196 bp). Despite these observed differences between PCa patients and controls, none of the descriptive statistical parameters of plasma cfDNA fragmentation differed significantly between the two groups (Table A6).

Urinary cfDNA

For urinary cfDNA fragmentation, PCa patients and controls harbored a similar maximum relative frequency (0.007), which was located at the modal fragment length of 111 bp in both groups, whereas the modal fragment length was increased to 148 bp in the subcohort of advanced PCa patients (Table A6). In the majority of samples, the maximum frequency corresponded to a pronounced local maximum within the 10-bp oscillation pattern, rather than to a peak at ~167 bp, as observed in plasma cfDNA. Comparison of minimum and maximum modal fragment lengths across controls, localized PCa and advanced PCa patients revealed similar ranges in all three cohorts. Urinary cfDNA from control samples showed a higher cumulative frequency at the modal fragment length (0.228) than tumor samples (0.205), while the subgroup of advanced PCa patients harbored a distinctly higher value compared to both localized PCa patients and controls (0.325). The mean fragment length decreased from controls (194 bp) to localized PCa patients (188 bp) and further to advanced PCa patients (184 bp). In statistical testing, the relative frequency at 167 bp fragment length was significantly higher in PCa patients compared to controls ($p = 0.036$), but this difference did not remain significant after correction for multiple testing across all evaluated fragmentation features (adjusted $p = 0.681$). All other descriptive statistical parameters of urinary cfDNA fragmentation equally showed no significant differences between tumor and control samples (Table A6).

3.4.3.2 Proportions of fragment length ranges

The evaluation of the relative frequency distributions of plasma and urinary cfDNA fragmentation revealed differences in the fragment length profiles between PCa patients and controls (compare with Section 3.4.2). To further assess these differences quantitatively, I defined several fragment length ranges, and calculated their proportions relative to all fragments within the 30–700 bp range. The selected fragment length ranges involved both short and longer cfDNA fragments, as well as fragment lengths associated with the first and second peak at ~167 bp and ~334 bp (±10 bp), respectively. Proportions of fragment length ranges in plasma or urinary cfDNA were statistically compared between tumor and control samples using the Wilcoxon rank sum test, and among all three groups using the Kruskal-Wallis test followed by Dunn’s post hoc test.

Plasma cfDNA

For plasma cfDNA fragmentation, tumor samples showed lower proportions of short fragments within the 30–60 bp, 30–100 bp, and 30–150 bp fragment length ranges compared with control samples (Figure 15A; Table A6). The proportion of fragments within the 30–60 bp range (P30–60 bp) was significantly lower in PCa patients compared with controls, but this difference did not remain significant after correction for multiple testing across all fragmentation features (p = 0.030, adjusted p = 0.170; Table A6). In contrast, all fragment length ranges which were overlapping with the first peak at ~167 bp or its ascending and descending regions (P30–180 bp, P163–169 bp, P160–180 bp, P180–220 bp, P150–300 bp) revealed higher proportions in tumor compared with control samples (Figure 15A). For the 150–300 bp fragment length range, the proportion was significantly higher in PCa patients than in controls, but the difference also did not remain significant after correction for multiple testing (p = 0.041, adjusted p = 0.170; Table A6). Fragment length ranges associated with the second peak at ~334 bp showed lower proportions in plasma samples from PCa patients compared with control samples. The statistical comparison of all three cohorts (controls, localized PCa, and advanced PCa) revealed no significant differences for any of the assessed fragment length proportions (Table A7).

Urinary cfDNA

For urinary cfDNA fragmentation, similar trends to those in plasma cfDNA were evident. Control samples harbored higher proportions of short fragment length ranges (P30–60 bp and P30–100 bp), as well as long fragment length ranges (P324–344 bp, P250–320 bp, P250–420 bp, P420–700 bp; Figure 15B; Table A6). In contrast, urine samples from PCa patients showed higher proportions for fragment length ranges which were associated with the prominent peak at ~167bp in plasma cfDNA (P30–180 bp, P163–169 bp, P160–180 bp, P180–220 bp, P150–300 bp; Figure 15B). For the 163–169 bp fragment length range, proportions were significantly higher in tumor compared with control samples, although the difference did not remain significant after correction for multiple testing (p = 0.044, adjusted p = 0.681; Table A6). However, the comparison between advanced PCa patients and controls for the same fragment length range remained significant (p = 0.012, adjusted p = 0.036; Table A7). Similarly, the proportion of the 160–180 bp fragment length range (P160–180 bp) showed significant differences in the comparison of all three groups (p = 0.045), equally driven by the comparison between advanced PCa patients and controls (p = 0.013, adjusted p = 0.039; Table A7).

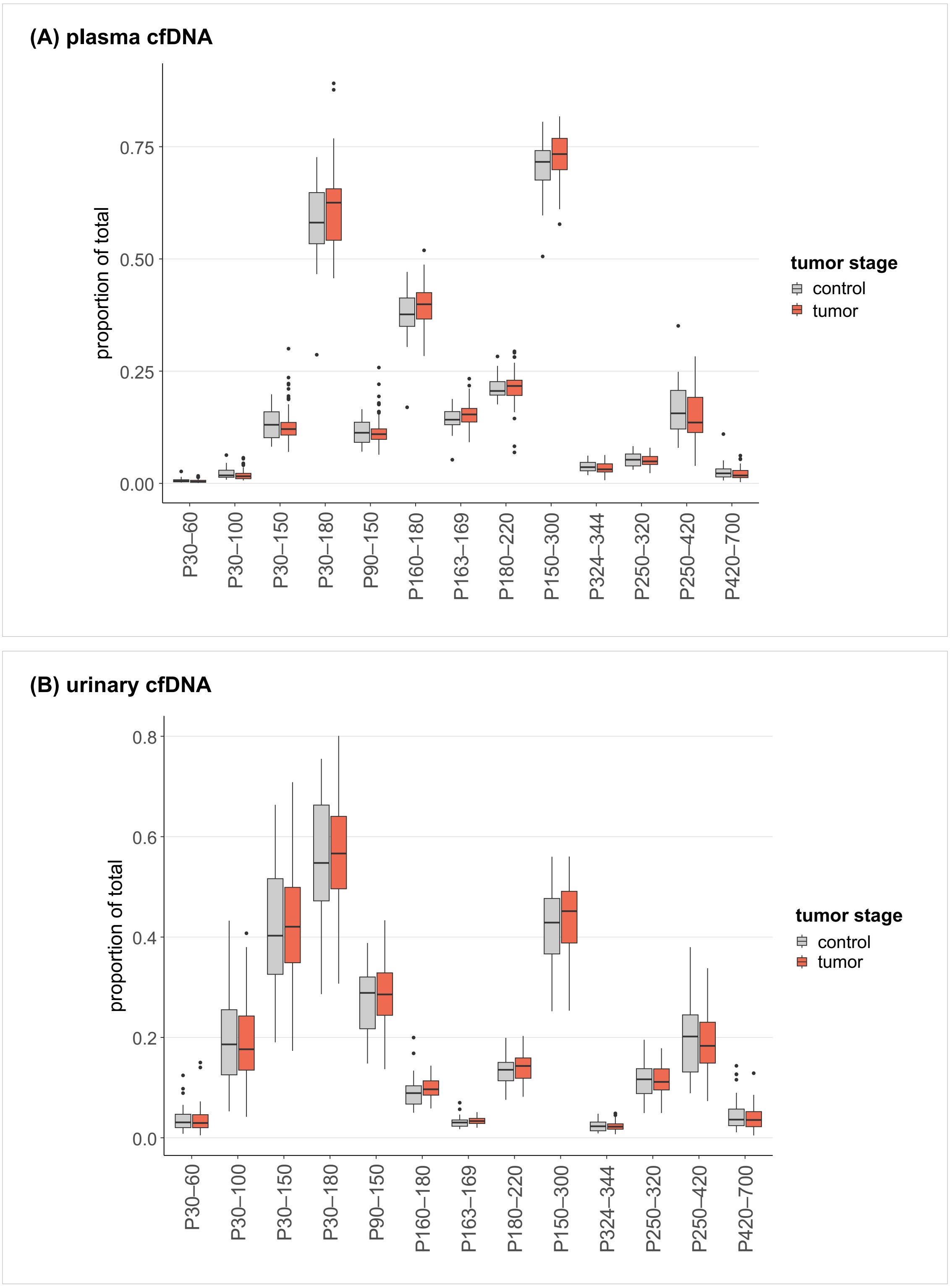


Figure 15: Proportions of selected fragment length ranges in plasma and urinary cfDNA.
Proportions of fragment length ranges were calculated relative to all cfDNA fragments within the 30–700 bp range, as derived from lcWGS data. **(A)** Comparison of plasma cfDNA fragment length proportions between tumor and control samples. **(B)** Comparison of urinary cfDNA fragment length proportions between tumor and control samples. **(A–B)** Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Dots represent outlying samples. Adapted from Riediger *et al.*[410]

3.4.3.3 Ratios between proportions of fragment length ranges

In addition to the analyses of individual fragment length range proportions, I calculated pairwise ratios between selected proportions to assess differences between short and long fragments, as well as between the theoretical first peak at ~167 bp and shorter fragments (< 100–150 bp). Ratios of fragment length range proportions in plasma and urinary cfDNA were statistically compared between tumor and control samples. Overall, the assessment of these ratios was consistent with the results obtained from the analyses of individual fragment length range proportions.

Plasma cfDNA

For plasma cfDNA fragmentation, control samples harbored a higher proportion of short fragments (< 100–150 bp) and a lower proportion of fragment length ranges associated with the peak at ~167 bp (see Section 3.4.3.2). Accordingly, ratios between short fragment length ranges (< 150 bp) and fragment length ranges overlapping with the first peak (160–180 bp, 163–169 bp, 150–300 bp) were higher in controls than in PCa patients (Figure 16A; Table A6). These differences were statistically significant for the ratios P30–150 bp/P163–169 bp ($p = 0.041$, adjusted $p = 0.170$), P30–100 bp/P160–180 bp ($p = 0.021$, adjusted $p = 0.170$) and P30–100 bp/P163–169 bp ($p = 0.017$, adjusted $p = 0.170$), but did not remain significant after correction for multiple testing (Table A6). For all three ratios, the observed differences were driven by advanced PCa patients, harboring significantly lower values compared with controls (Table A7). Conversely, PCa patients showed a higher, but non-significant, ratio between fragment length ranges associated with the first peak and the second peak (*i.e.*, P160–180 bp/P250–420 bp, $p = 0.181$).

Urinary cfDNA

For urinary cfDNA fragmentation, similar trends in the analyzed ratios were observed as for plasma cfDNA. Ratios between short fragment length ranges (< 150 bp) and fragment length ranges associated with the first peak in plasma cfDNA were higher in controls than in tumor samples, whereas the ratio P160–180 bp/P250–420 bp was again higher in PCa patients (Figure 16B; Table A6). None of the analyzed ratios differed significantly between tumor and control samples or among controls, localized PCa and advanced PCa patients (Table A6 and Table A7).

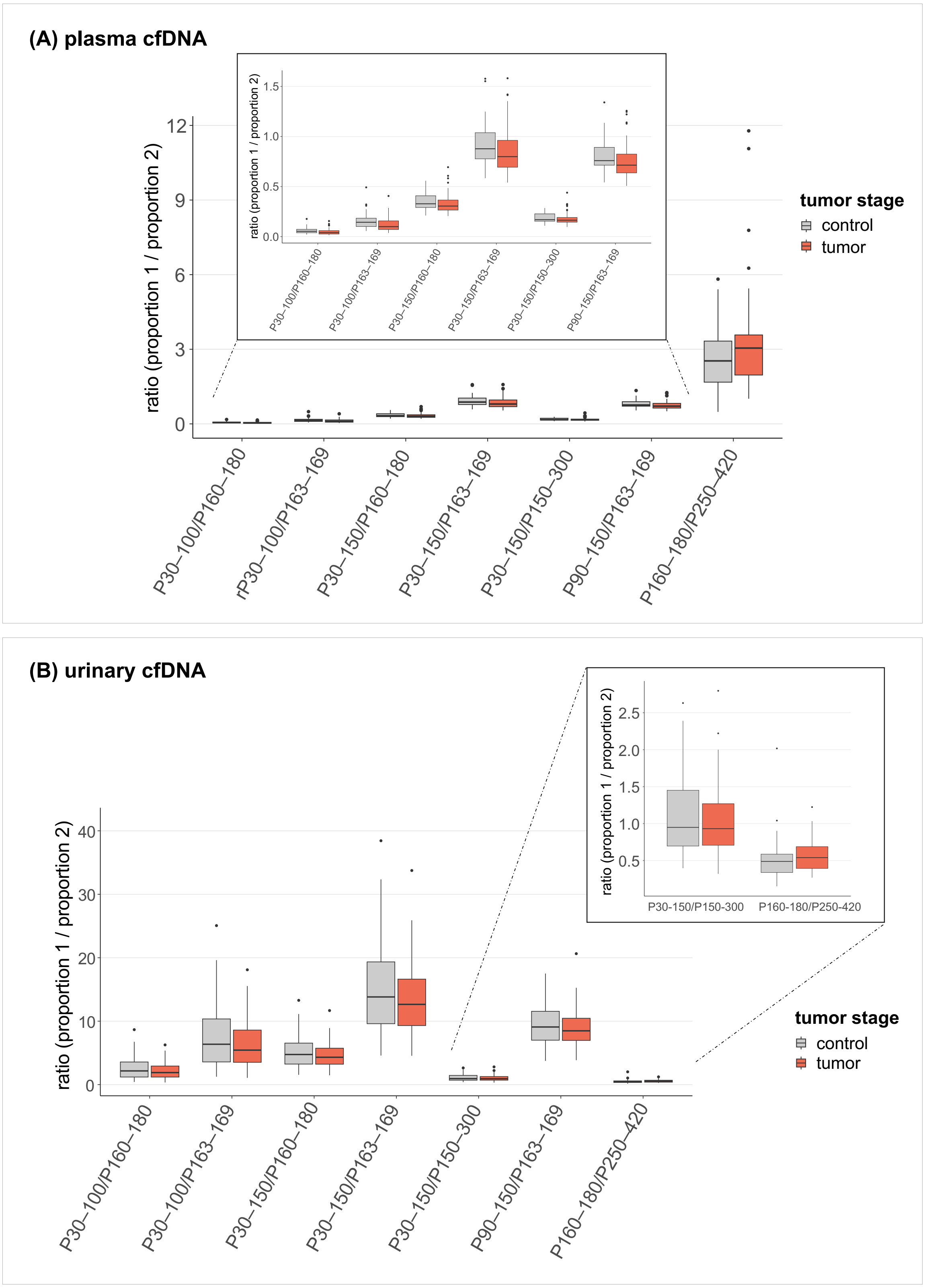


Figure 16: Pairwise ratios between proportions of selected fragment length ranges in plasma and urinary cfDNA. Pairwise ratios were calculated between fragment length range proportions derived from lcWGS data, with each proportion calculated relative to all cfDNA fragments within the 30–700 bp range. **(A)** Comparison of pairwise ratios of fragment length range proportions in plasma cfDNA between tumor and control samples. **(B)** Comparison of pairwise ratios of fragment length range proportions in urinary cfDNA between tumor and control samples. **(A–B)** Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Dots represent outlying samples. The additional boxes display zoomed windows. Adapted from Riediger *et al.*[410]

3.4.3.4 10-bp periodic oscillation pattern

The relative frequency distributions of plasma and urinary cfDNA fragmentation exhibited a periodic oscillation pattern within the 30–150 bp fragment length range, characterized by local maxima and minima at approximately 10-bp intervals (Figure 17A; Figure 18A). In urinary cfDNA, this 10-bp oscillation pattern was more pronounced than in plasma cfDNA and extended to fragment lengths of 150–300 bp. I calculated the average positions of local maxima and minima, and determined a synoptic oscillation score for each plasma and urine sample based on the differences in the amplitudes of local maxima and minima.

<u>Plasma cfDNA</u>

In plasma cfDNA, local maxima were located at 51, 61, 71, 81, 92, 102, 112, 122, 133, and 144 bp, whereas the positions of local minima followed approximately 3–5 bp subsequently at 54, 66, 75, 86, 97, 107, 116, 127, 136, and 145 bp. The proportion of plasma cfDNA fragments within the 30–150 bp range decreased from controls (0.131) to localized PCa (0.121) and advanced PCa patients (0.119; Table A6). Accordingly, tumor samples showed a less pronounced 10-bp oscillation pattern in plasma cfDNA compared with control samples (Figure 17A), which was reflected by a progressive decrease in the oscillation score from controls to localized PCa and advanced PCa patients (Figure 17B; Table A6). The oscillation score was significantly lower in plasma samples from PCa patients than in controls, but this difference did not remain significant after correction for multiple testing ($p = 0.020$, adjusted $p = 0.170$; Figure 17B; Table A6). Comparison among all three cohorts further revealed significant differences between localized PCa patients and controls ($p = 0.048$), as well as between advanced PCa patients and controls ($p = 0.028$), although the observed differences likewise did not remain significant after correction for multiple testing (adjusted $p = 0.072$ for both; Table A7).

Within the 150–300 bp fragment length range, no clear 10-bp oscillation pattern was observed in plasma cfDNA (Figure 13A and B). Between one and up to seven local maxima and minima were distributed across this fragment length range, with variable distances between them (Table A6). Consequently, this fragment length range was not included in further analyses of the 10-bp oscillation pattern in plasma cfDNA.

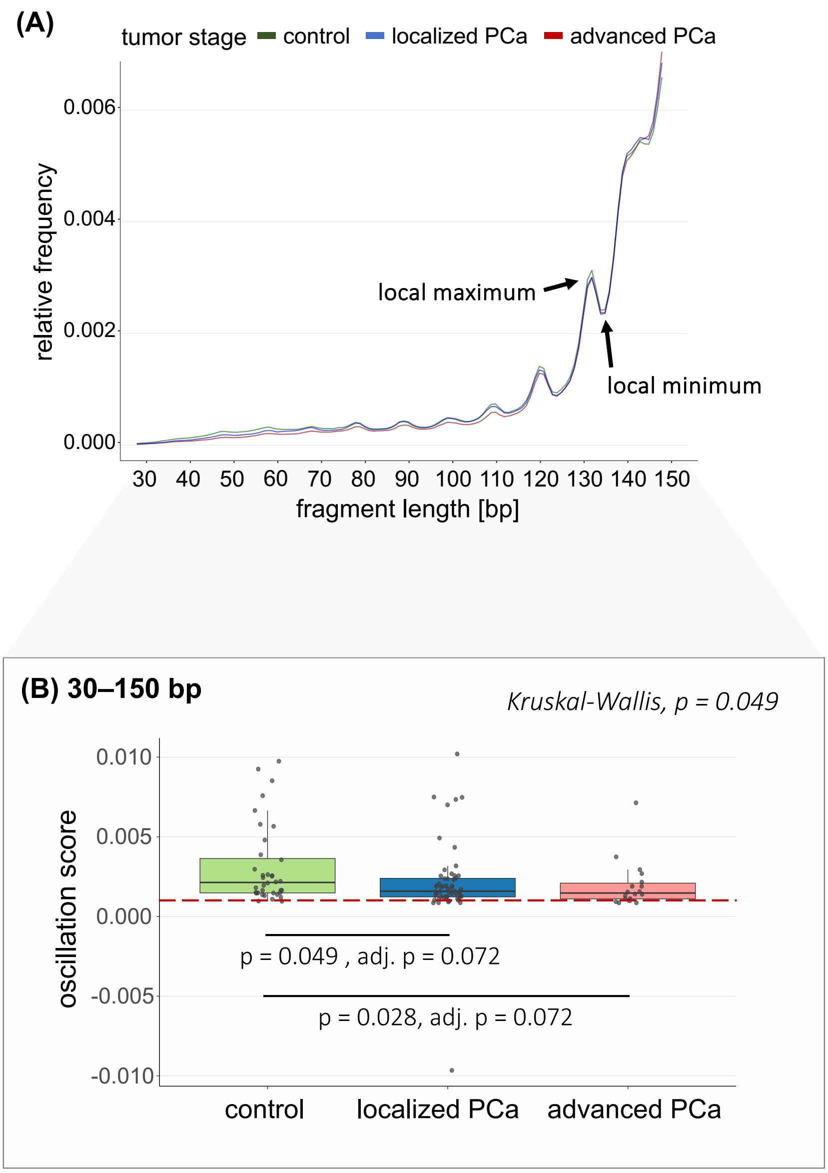


Figure 17: Oscillation pattern with 10-bp periodicity in plasma cfDNA fragmentation.
(A) Plasma cfDNA fragmentation profiles within the 30–150 bp fragment length range, derived from lcWGS data, exhibiting an oscillation pattern with local maxima and minima at 10-bp intervals. Median fragmentation profiles are shown across all samples from localized PCa patients, advanced PCa patients, and controls. Y-axis: relative frequencies of cfDNA fragments with a given length (x-axis) relative to all fragments within the 30–700 bp range. **(B)** Distribution of 10-bp oscillation scores across plasma samples from localized PCa patients, advanced PCa patients, and controls. Oscillation scores were calculated based on the deviation between the sum of the heights of all local maxima and the sum of the depths of all local minima within the 30–150 bp fragment length range. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one sample. The three cohorts were compared using Kruskal-Wallis testing, followed by Dunn's post hoc test. Correction for multiple testing was performed using the Benjamini-Hochberg method. Only significant differences are shown. The red dashed line indicates the ctDNA detection threshold, defined as the 5th-percentile of values derived from control samples. Adapted from Riediger *et al.*[410]

Urinary cfDNA

In urinary cfDNA, local maxima were located at 41, 52, 62, 72, 83, 93, 103, 114, 124, 133, 142, and 148 bp, while local minima were slightly shifted toward longer fragment lengths and were located at 45, 56, 66, 77, 87, 97, 108, 118, 128, 137, and 146 bp. The proportion of urinary cfDNA fragments within the 30–150 bp range was lower in controls (0.403) than in PCa patients (0.421; Table A6). The highest oscillation score was observed in samples from localized PCa patients, whereas control samples harbored a lower score, and samples from advanced PCa patients the lowest oscillation score (Figure 18B, Table A6). Statistical comparison revealed no significant differences in the oscillation scores between PCa patients and controls in urinary cfDNA (Table A6 and Table A7).

In contrast to plasma cfDNA, a distinct 10-bp oscillation pattern was visible in urinary cfDNA within the 150–300 bp fragment length range (Figure 18C). Local maxima were located at 159, 175, 196, 210, 220, 228, 237, 242, 250, 258, 267, 275, 283, and 291 (293) bp, and local minima were located at 165, 183, 201, 208, 216, 225, 233, 241, 248, 257, 266, 274, 283, 291, and 297 bp. The proportion of fragments within the 150–300 bp range was higher in tumor than in control samples, and was the highest in samples from advanced PCa patients (Figure 18C; Table A6). The oscillation scores were similar between samples from localized PCa patients and controls, and were increased in samples from advanced PCa patients (Figure 18C). However, statistical analyses revealed no significant differences between tumor and control samples, nor among controls, localized PCa and advanced PCa patients (Table A6 and Table A7).

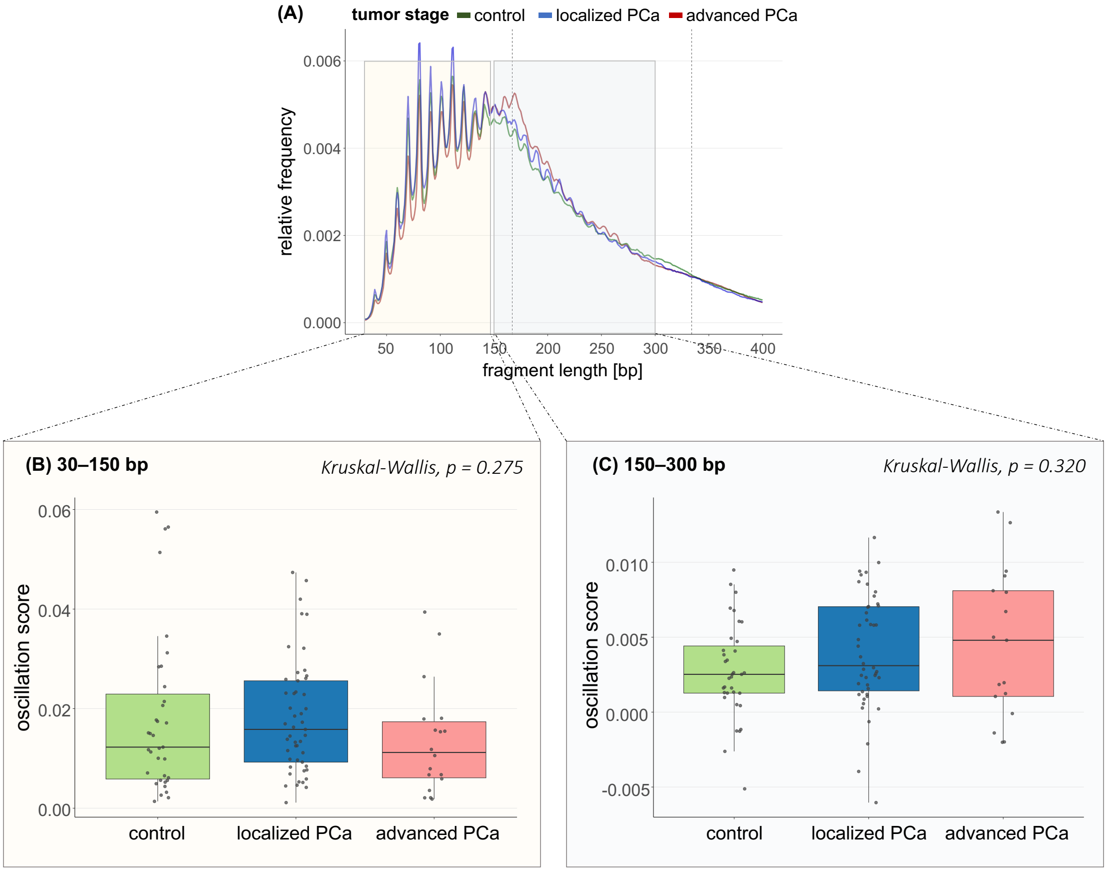


Figure 18: Oscillation pattern with 10-bp periodicity in urinary cfDNA fragmentation.
(A) Urinary cfDNA fragmentation profiles within the 30–400 bp fragment length range, derived from lcWGS data, exhibiting an oscillation pattern with local maxima and minima at 10-bp intervals. Median fragmentation profiles are shown across all samples from localized PCa patients, advanced PCa patients, and controls. Y-axis: relative frequencies of cfDNA fragments with a given length (x-axis) relative to all fragments within the 30–700 bp range. **(B–C)** Distribution of 10-bp oscillation scores across plasma samples from localized PCa patients, advanced PCa patients, and controls. Oscillation scores were calculated based on the deviation between the sum of the heights of all local maxima and the sum of the depths of all local minima, within the **(B)** 30–150 bp and **(C)** 150–300 bp fragment length ranges. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one sample. The three cohorts were compared using Kruskal-Wallis testing, followed by Dunn's post hoc test. Correction for multiple testing was performed using the Benjamini-Hochberg method. Only significant differences are shown. Adapted from Riediger *et al.*[410]

3.4.4 CfDNA fragmentation features were informative for ctDNA detection

Based on the previous cfDNA fragmentation analyses, I extracted one informative feature for plasma and urinary cfDNA to determine the applicability for ctDNA detection in tumor samples.

Plasma cfDNA

For plasma cfDNA, I selected the 10-bp oscillation score in the 30–150 bp fragment length range, given the pronounced differences in the scores observed between tumor and control samples (Figure 17B). Additionally, the potential of this fragmentation feature for tumor characterization was already described in previous studies[213,274]. Control samples harbored higher oscillation scores compared to tumor samples. Consequently, I defined the positivity threshold for ctDNA detection as the 5th-percentile of the results obtained from all control samples. Tumor samples with values below this threshold (0.00105) were determined as ctDNA-positive. Twelve PCa patients harbored an oscillation score below the defined threshold (Table 15; Table A8), including four patients with metastatic PCa, and eight patients with localized PCa (7 x intermediate-risk, 1 x high-risk).

Urinary cfDNA

For urinary cfDNA, I selected the proportion of fragments within the 163–169 bp range relative to all fragments (P163–169 bp; Figure 19A), as this metric showed significant differences between tumor and control samples, particularly between advanced PCa patients and controls (Figure 19B; Table A6 and Table A7). Tumor samples harbored higher values than controls. Consequently, I defined the positivity threshold for ctDNA detection as the 95th-percentile of the results obtained from all control samples. Tumor samples with results above this threshold (0.0496) were determined as ctDNA-positive. One tumor patient with lymph-node positive, advanced PCa harbored a value above the defined threshold (Table 15; Table A8). The maximum values for P163–169 bp of the entire tumor and control cohort were identified in two control samples. Their results were 3.5 and 2.3 SD above the mean of all controls, respectively. Tumor samples showed values intermediate between these extreme measurements and the remaining controls. Accordingly, thresholds derived from the control distribution differed depending on the influence of one or both extreme values, under which additional seven or even 11 tumor samples, respectively, would have exceeded the threshold (Table A8).

Table 15: CtDNA detection in plasma and urinary cfDNA based on fragmentation features.
Number of localized and advanced PCa patients with detectable ctDNA (plasma: oscillation score below 5th-percentile of values from controls; urine: P163–169 bp above 95th-percentile of values from controls) in plasma or urine separately, for matched plasma and urine samples (ctDNA detected in both biofluids from the same patient), and for complementary ctDNA detection in either plasma or urine. 68 PCa patients with matched plasma and urine samples; 4 localized PCa patients with missing urine samples.

tumor stage	analyzed liquid biopsy sample	# ctDNA-positive patients
localized PCa	plasma, n = 55	8 (15%)
	urine, n = 50	0 (0%)
	matched plasma + urine, n = 50	0 (0%)
	plasma or urine, n = 55	8 (15%)
advanced PCa	plasma, n = 18	4 (22%)
	urine, n = 18	1 (6%)
	matched plasma + urine, n = 18	0 (0%)
	plasma or urine, n = 18	4 (22%)

Abbreviations: ctDNA = circulating tumor-derived DNA, P163–169 bp = proportion of fragments with 163–169 bp length

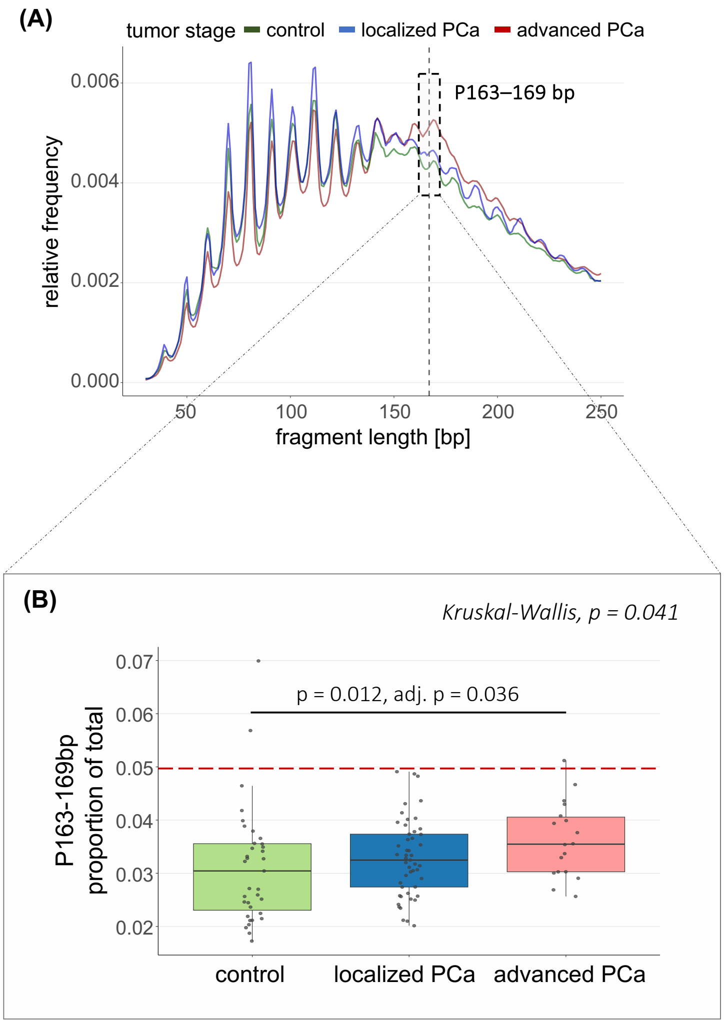


Figure 19: CtDNA detection in PCa samples based on urinary cfDNA fragmentation.
(A) Urinary cfDNA fragmentation profiles within the 30–250 bp fragment length range, derived from lcWGS data. Median fragmentation profiles are shown across all samples from localized PCa patients, advanced PCa patients, and controls. The box indicates the 163–169 bp fragment length range. Y-axis: relative frequencies of cfDNA fragments with a given length (x-axis) relative to all fragments within the 30–700 bp range. **(B)** Distribution of the proportion of fragments within the 163-169 bp range (P163–169 bp) across urine samples from localized PCa patients, advanced PCa patients, and controls. The proportion P163–169 bp was calculated relative to all cfDNA fragments within the 30–700 bp range. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one sample. The three cohorts were compared using Kruskal-Wallis testing, followed by Dunn's post hoc test. Correction for multiple testing was performed using the Benjamini-Hochberg method. Only significant differences are shown. The red dashed line indicates the ctDNA detection threshold, defined as the 95th-percentile of values derived from control samples. Abbreviations: P163–169 bp = proportions of fragments within the range 163–169 bp. Adapted from Riediger *et al.*[410]

3.5 Genome-wide CNV profiling in liquid biopsies and PCa tissue

I assessed genome-wide CNVs in 1000-kb windows using lcWGS and cfMeDIP-seq data from plasma and urinary cfDNA samples of PCa patients and controls, applying the *ichorCNA* algorithm. The sequencing reads were downsampled to 20 M paired (40 M total) reads to ensure uniformity across samples. No sample had to be excluded because of insufficient number of sequencing reads after downsampling. The sequencing data from cancer-free controls was used as copy-number-neutral reference for the tumor samples, establishing a PoN (plasma: n = 36; urine: n = 35). This PoN was also used as reference for individual control samples, excluding the analyzed control sample from the respective PoN (plasma: n = 35; urine: n = 34).

The CNV analysis using the *ichorCNA* algorithm revealed genome-wide CNV profiles displaying chromosomal gains, losses and copy-number-neutral regions for each sample, as well as an estimated TFx based on the identified genomic alterations. I used the estimated TFx to determine the presence of ctDNA in the tumor samples, with a detection threshold defined based on the 95th-percentile of the results obtained from all control samples.

In addition to the CNV analyses of all downsampled sequencing reads, I performed *in silico* size selection on the lcWGS data, and subsequently repeated the CNV analysis using size-selected, downsampled sequencing reads. This approach aimed to enrich for ctDNA and thereby improve the sensitivity of the CNV analysis. For plasma cfDNA, I size-selected sequencing reads to an insert size of 90–150 bp, since it was previously shown that *in silico* size selection for this fragment length range enriched for ctDNA[204,274]. For urinary cfDNA, I assessed 15 different fragment length ranges, since the application of *in silico* size selection prior to the CNV analysis in urinary cfDNA has not been previously reported in literature. Size-selected sequencing reads were downsampled to 2 M paired (4 M total) reads, due to the reduced number of remaining sequencing reads after *in silico* size selection. For plasma cfDNA, no samples had to be excluded because of insufficient number of sequencing reads. For urinary cfDNA, all samples harbored sufficient numbers of sequencing reads after the *in silico* size selection for the different insert size ranges. One urine sample did not reach 4 M total reads after the size selection for the 20–110 bp fragment length range, but yielded 3.8 M total reads, corresponding to a deviation of 5.1% from the target read number. Given this minor deviation, the sample was retained for subsequent analyses.

3.5.1 Genome-wide CNV analysis indicated ctDNA presence in plasma cfDNA

3.5.1.1 CNV analysis based on lcWGS and cfMeDIP-seq data

<u>LcWGS data</u>

Genome-wide CNV analysis based on lcWGS data revealed detectable genomic alterations in a small number of plasma samples. Both PCa patients and controls harbored a median estimated TFx of 0.0 (Table A9), as 63.89% of control samples (23 / 36) and 65.75% of tumor samples (48 / 73) had an estimated TFx value of 0.0. Statistical analyses revealed no significant differences in TFx values between tumor and control samples (p = 0.973), nor among controls, localized PCa and advanced PCa patients (p = 0.681; Figure 20).

The ctDNA positivity threshold was set to 0.018 based on the 95th-percentile of the results obtained from all control samples. Five tumor samples were classified as ctDNA-positive based on estimated TFx values above this threshold (range: 0.019–0.464; Figure 20). Two ctDNA-positive plasma samples from metastatic PCa patients demonstrated distinctly elevated TFx values (0.464 and 0.160), whereas three tumor samples from two patients with localized high-risk PCa and one patient with localized intermediate-risk PCa showed slightly increased TFx values above the threshold (range: 0.019–0.024).

The maximum TFx value of controls (0.0415) was 2.27-fold higher than the 95th-percentile threshold (0.018), indicating the presence of outliers in the control cohort. The two ctDNA-positive samples from metastatic PCa patients with distinctly elevated TFxs exceeded the values of the two control samples with the highest and second highest TFx. The CNV profile of one of these control samples demonstrated a systematic loss of chromosome 19 (Figure A7), a finding which has previously been reported by the authors of the *ichorCNA* algorithm to potentially bias the CNV analysis[286]. In the second control sample, the corresponding urine sample likewise showed the second highest TFx value of all controls in the CNV analysis based on lcWGS data from urinary cfDNA. However, detectable alterations were not concordant between matched plasma and urine samples (Figure A7).

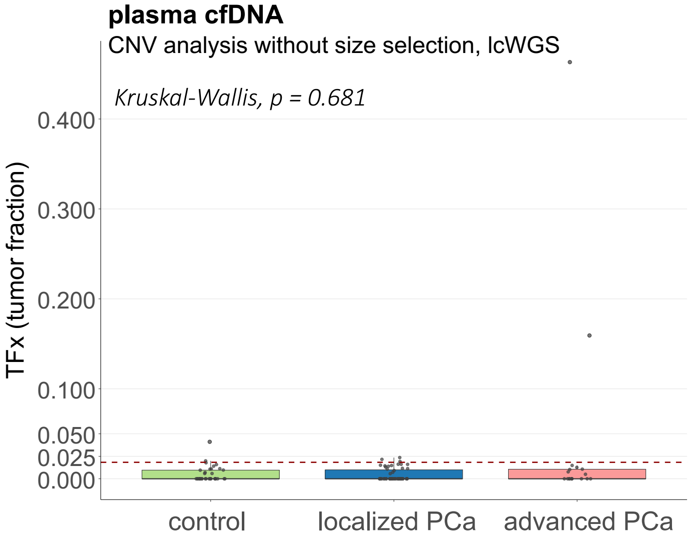


Figure 20: CNV analysis without size selection in lcWGS data from plasma cfDNA.
Distribution of the estimated TFx across plasma samples from localized PCa patients, advanced PCa patients, and controls. TFx values were estimated based on genome-wide CNV profiles. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one sample. The three cohorts were compared using Kruskal-Wallis testing, followed by Dunn's post hoc test. Correction for multiple testing was performed using the Benjamini-Hochberg method. Only significant differences are shown. The red dashed line indicates the ctDNA detection threshold, defined as the 95th-percentile of values derived from control samples. Adapted from Riediger *et al.*[410]

CfMeDIP-seq data

I equally applied the genome-wide CNV analysis on the cfMeDIP-seq data from plasma cfDNA, in order to assess whether genomic alterations could be identified based on methylation-enriched data. The assessment revealed that the estimated TFx values based on the cfMeDIP-seq data for both tumor and control samples were increased compared to the results obtained from matched lcWGS data. Only a small number of tumor samples harbored a lower or comparable TFx value based on the cfMeDIP-seq data compared with the lcWGS data, while the majority of tumor samples showed an increased TFx, also in case where the CNV analysis based on lcWGS data revealed a TFx value of 0.0 (Figure 21A). Estimated TFx values derived from corresponding lcWGS and cfMeDIP-seq data showed a weak correlation (Spearman's Rho, $\rho = 0.319$, $p < 0.05$; Figure A8).

The maximum TFx of control samples increased to 0.057. The ctDNA detection threshold based on the 95th-percentile of the results obtained from all control samples (0.037) was equally higher compared to the threshold derived from lcWGS data. Plasma samples from PCa patients harbored a slightly lower median TFx (median TFx = 0.014) compared to controls (median TFx = 0.015), while both medians were below the detection threshold. The median TFx of the PCa cohort was driven by the low median TFx of localized PCa patients (median TFx = 0.013), who comprised the majority of PCa patients, whereas advanced PCa patients harbored an increased median TFx (median TFx = 0.022) compared to controls. Statistical analyses revealed no significant differences in TFx values between tumor and control samples, nor among controls, localized PCa and advanced PCa patients ($p = 0.134$; Figure 21B).

Despite the higher ctDNA detection threshold in the CNV analysis based on the cfMeDIP-seq data, a greater number of tumor samples (n = 11) harbored a TFx value above the threshold (range: 0.037–0.468) compared with the analysis based on the lcWGS data. The two metastatic PCa patients who showed distinctly elevated TFx values in the lcWGS-based analysis harbored comparable TFx values based on the cfMeDIP-seq data (Figure 21B). Two out of three additional ctDNA-positive tumor samples in the lcWGS-based analysis were not classified as ctDNA-positive based on the cfMeDIP-seq analysis, whereas the third ctDNA-positive sample remained positive and harbored an increased TFx compared with the lcWGS-based analysis. Eight additional tumor samples demonstrated increased TFx values relative to the lcWGS-based analysis and exceeded the ctDNA positivity threshold in the CNV analysis based on the cfMeDIP-seq data (Figure 21A and B).

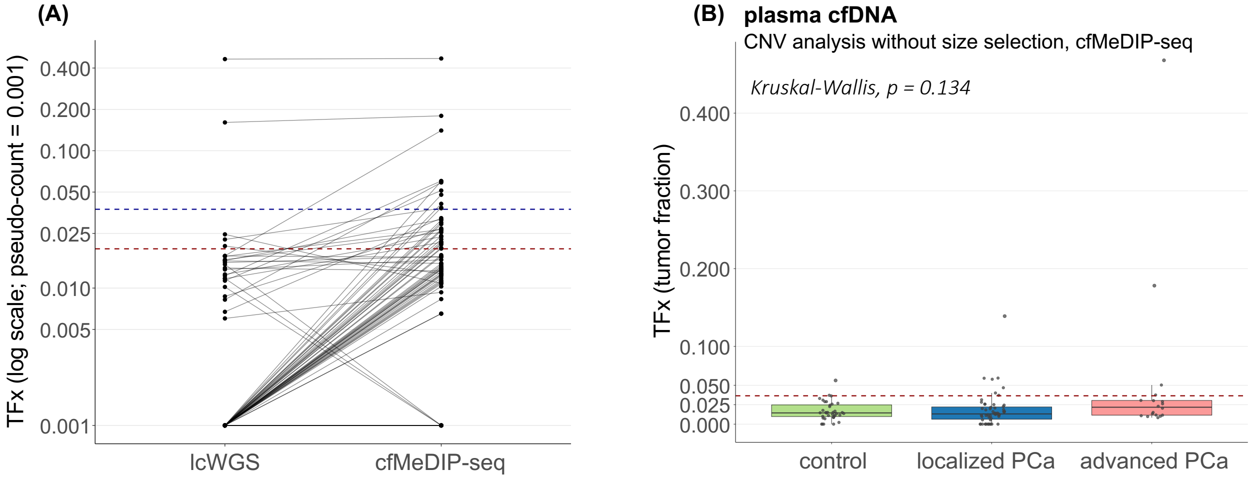


Figure 21: CNV analysis without size selection in cfMeDIP-seq data from plasma cfDNA.
(A) Comparison of matched TFx values in tumor samples derived from lcWGS data and cfMeDIP-seq data. TFx values were estimated based on genome-wide CNV profiles. TFx values are shown on a log-transformed y-axis, and a pseudo-count of 0.001 was added for visualization. Red and blue dashed lines indicate the ctDNA detection threshold derived from lcWGS data and cfMeDIP-seq data, respectively, defined as the 95th-percentile of values derived from control samples. **(B)** Distribution of the estimated TFx across plasma samples from localized PCa patients, advanced PCa patients, and controls. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one sample. The three cohorts were compared using Kruskal-Wallis testing, followed by Dunn's post hoc test. Correction for multiple testing was performed using the Benjamini-Hochberg method. Only significant differences are shown. The red dashed line indicates the ctDNA detection threshold, defined as the 95th-percentile of values derived from control samples.

3.5.1.2 CNV analysis based on *in silico* size-selected lcWGS data

I applied *in silico* size selection (90–150 bp) to lcWGS-derived sequencing reads from plasma cfDNA with the aim to enhance the sensitivity of the CNV analysis through ctDNA enrichment in the tumor samples, as previously demonstrated by Mouliere *et al.* [204,274].
Estimated TFx values for both tumor and control samples were significantly higher in the CNV analysis with *in silico* size selection than in the analysis without size selection (tumor, p = 2.07e-07; control, p = 1.41e-04; Table A9). The median TFx of all tumor samples increased more (median absolute increase: 0.009) compared to the median TFx of controls (median absolute increase: 0.005), and the median TFx of plasma samples from advanced PCa patients increased the most through the size selection (median absolute increase: 0.010; Figure 22A; Table A9). A smaller number of tumor and control samples harbored a TFx of 0.0 (control: 19.44% (n = 7), tumor: 21.92% (n = 16)) compared with the CNV analysis without size selection. Both localized and advanced PCa samples demonstrated a higher, but not significantly increased, median TFx compared to control samples (Figure 22B).

Furthermore, the approach revealed a greater number of tumor samples with detectable ctDNA in the CNV analysis. Nine tumor samples harbored a TFx value above the detection threshold (range: 0.025–0.577), which was also defined based on the 95th-percentile of results derived from control samples. The TFx increase based on the size selection was also evident in the two plasma samples from metastatic PCa patients which showed elevated TFx values in the CNV analysis without size selection, with TFx values increasing from 0.160 to 0.211 (+ 32%) and 0.464 to 0.577 (+ 24%), respectively (Figure 22C; Figure A9). Three tumor samples with ctDNA-positive result in the CNV analysis without size selection were ctDNA-negative based on the CNV analysis with size selection, while seven additional PCa samples turned ctDNA-positive (Figure 22C; Figure A9).

The ctDNA positivity threshold (0.023) was 1.25-fold higher compared to the threshold derived from the CNV analysis without size selection. The maximum TFx observed among all control samples (0.056) was 1.35-fold higher and exceeded the 95th-percentile threshold by 2.44-fold, consistent with the findings in the CNV analysis without size selection. However, a different control sample showed the highest TFx following size selection. The CNV profile of this sample also demonstrated a systematic decrease in chromosome 19 (Figure A7). The control sample with the second highest TFx after size selection corresponded to the control sample with the highest TFx of all controls in the CNV analysis without size selection, whereas the control sample with the previously second highest TFx no longer showed an elevated TFx value.

In summary, *in silico* size selection of sequencing reads with an insert size of 90–150 bp prior to CNV analysis resulted in increased TFx values across all plasma samples and a higher number of ctDNA-positive PCa samples.

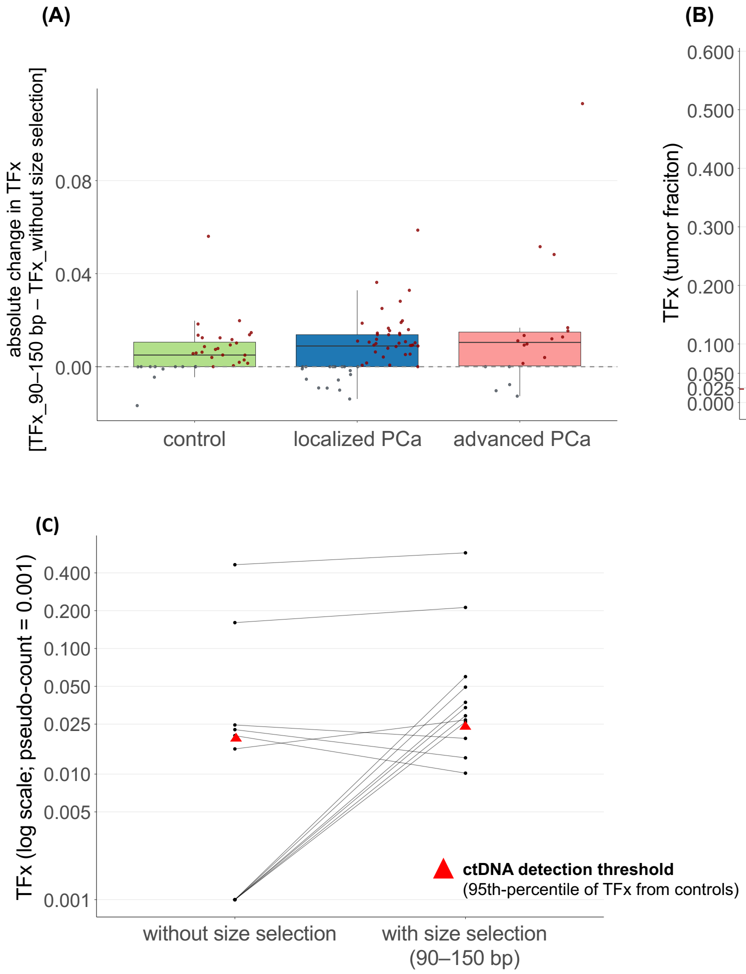


Figure 22: CNV analysis with *in silico* size selection in lcWGS data from plasma cfDNA.
(A) Distribution of the absolute change in estimated TFx values following *in silico* size selection for 90–150 bp fragments across plasma samples from localized PCa patients, advanced PCa patients, and controls. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one sample. Samples showing increased TFx values after size selection are highlighted in red, whereas samples with decreased TFx values are highlighted in gray. **(B)** Distribution of the estimated TFx values across plasma samples from localized PCa patients, advanced PCa patients, and controls. TFx values were estimated based on genome-wide CNV profiles derived from the CNV analysis with size selection. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one sample. The three cohorts were compared using Kruskal-Wallis testing, followed by Dunn's post hoc test. Correction for multiple testing was performed using the Benjamini-Hochberg method. Only significant differences are shown. The red dashed line indicates the ctDNA detection threshold, defined as the 95th-percentile of values derived from control samples. **(C)** Comparison of matched TFx values in tumor samples derived from the CNV analyses of lcWGS data without size selection and with *in silico* size selection for 90–150 bp fragments. TFx values are shown on a log-transformed y-axis, and a pseudo-count of 0.001 was added for visualization. Red triangles indicate the corresponding ctDNA detection thresholds, defined as the 95th-percentile of values derived from control samples. Adapted from Riediger *et al.*[410]

3.5.2 Genome-wide CNV analysis indicated ctDNA presence in urinary cfDNA

3.5.2.1 CNV analysis based on lcWGS and cfMeDIP-seq data

LcWGS data

The genome-wide CNV analysis based on lcWGS data from urinary cfDNA likewise revealed detectable genomic alterations in urine samples, exceeding the number of positive plasma samples. Urine samples from PCa patients harbored a higher median TFx (0.0169) compared with controls (0.0134; Table A9). When comparing all three cohorts, urine samples from advanced PCa patients demonstrated a significantly higher median TFx compared with control samples (0.0255; $p = 0.004$), whereas the median TFx in samples from localized PCa patients was slightly increased (0.0141; $p = 0.168$; Figure 23). Advanced PCa samples also harbored a higher median TFx than localized PCa samples, but this difference did not reach statistical significance ($p = 0.058$). In urine, 31.43% of control samples ($n = 11$) and 16.42% of tumor samples ($n = 11$) harbored a TFx value of 0.0, representing lower proportions than observed in plasma cfDNA.

The ctDNA positivity threshold was set to 0.0412 based on the 95th-percentile of the results obtained from all control samples. Eight tumor samples were classified as ctDNA-positive based on estimated TFx values above this threshold (range: 0.0498–0.678; Figure 23; Figure A10). These ctDNA-positive tumor samples originated from three patients with advanced PCa (2x N1M0, 1x M1) and five patients with localized intermediate-risk PCa. The maximum TFx observed among control samples (0.064) was 1.56-fold higher compared to the 95th-percentile threshold, indicating the presence of outliers within the control cohort. Both maximum TFx and the 95th-percentile value of control samples were higher in urine than in plasma. The control sample with the highest TFx (0.064) was not obtained from the same control as the outlying plasma sample in the lcWGS-based CNV analysis without size selection. The control sample with the second highest TFx (0.055) corresponded to the control sample with the second highest TFx value in plasma cfDNA (Figure A7). Both control samples demonstrated a distinct peak at ~160–170 bp in their urinary cfDNA fragmentation profiles, consistent with the patterns observed in the matched plasma cfDNA samples (Figure A11).

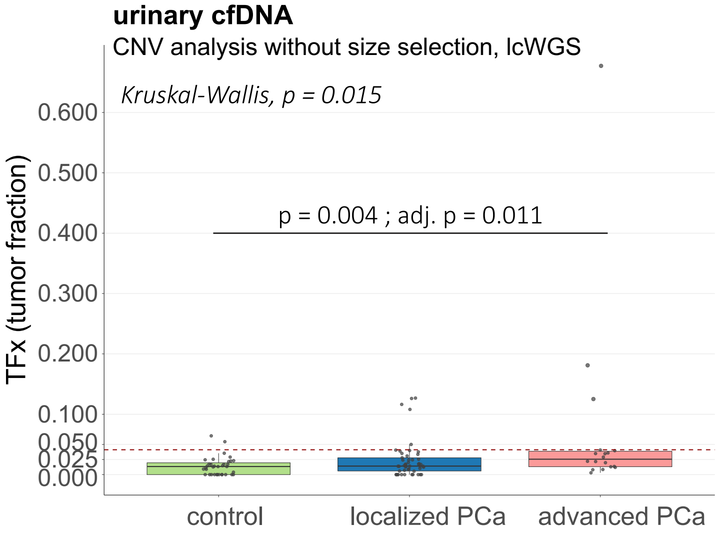


Figure 23: CNV analysis without size selection in lcWGS data from urinary cfDNA.
Distribution of the estimated TFx across urine samples from localized PCa patients, advanced PCa patients, and controls. TFx values were estimated based on genome-wide CNV profiles. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one sample. The three cohorts were compared using Kruskal-Wallis testing, followed by Dunn's post hoc test. Correction for multiple testing was performed using the Benjamini-Hochberg method. Only significant differences are shown. The red dashed line indicates the ctDNA detection threshold, defined as the 95th-percentile of values derived from control samples. Adapted from Riediger *et al.*[410]

CfMeDIP-seq data

I equally applied the genome-wide CNV analysis on the cfMeDIP-seq data from urinary cfDNA, analogous to the analysis performed for plasma cfDNA.

Similar to the plasma analysis, estimated TFx values derived from cfMeDIP-seq data for both tumor and control samples were higher than those obtained from matched lcWGS data, but this increasing trend was less consistent in urine than in plasma (Figure 24A). Several urine samples showed comparable or lower TFx in the CNV analysis based on cfMeDIP-seq data compared to lcWGS-based analysis, and the maximum TFx in tumor samples was also lower in the cfMeDIP-seq data (Figure 24A). Estimated TFx values derived from corresponding lcWGS and cfMeDIP-seq data demonstrated a moderate correlation across tumor samples (Spearman's $\rho = 0.615$, $p < 0.05$; Figure A12).

In the CNV analysis based on cfMeDIP-seq data, tumor samples harbored a higher median TFx (0.023) compared to controls (0.021; Table A9). Urine samples from advanced PCa patients showed the highest median TFx (0.027), whereas the median TFx in urine samples from localized PCa patients (0.022) was slightly increased compared with control samples (Figure 24B; Table A9). Statistical comparison revealed no significant differences among the cohorts.

The ctDNA detection threshold based on the 95th-percentile of values derived from control samples (0.091), was distinctly higher in the CNV analysis based on cfMeDIP-seq data than in the lcWGS-based analysis. Fewer tumor samples ($n = 6$) harbored TFx values which exceeded this threshold (range: 0.098–0.158), in contrast to the findings in plasma cfDNA. These ctDNA-positive tumor samples originated from two advanced PCa patients (N1M0) and four patients with localized intermediate-risk PCa. All six samples were ctDNA-positive the lcWGS-based CNV analysis, but their TFx values remained similar or decreased in the analysis based on the cfMeDIP-seq data (Figure 24A).

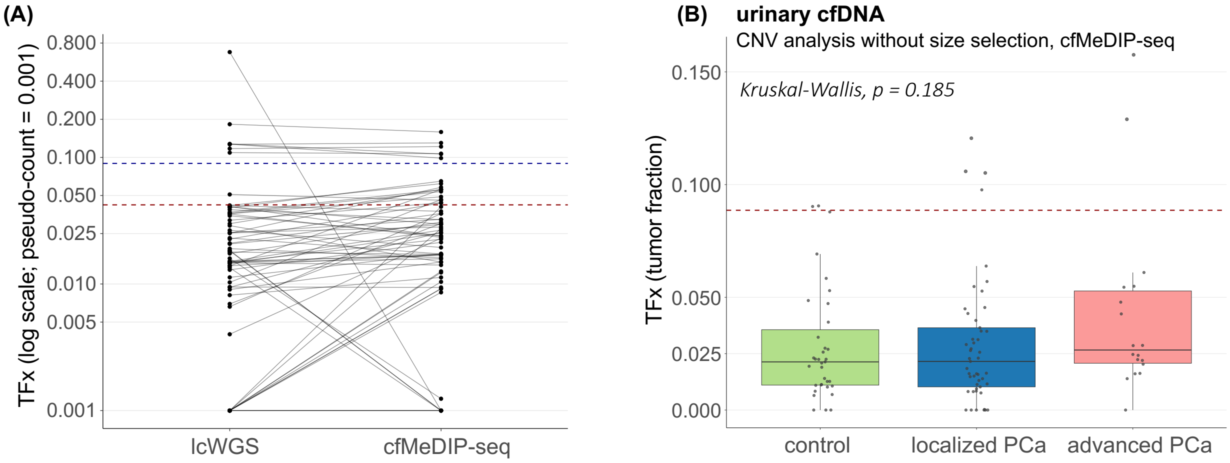


Figure 24: CNV analysis without size selection in cfMeDIP-seq data from urinary cfDNA.
(A) Comparison of matched TFx values in tumor samples derived from lcWGS data and cfMeDIP-seq data. TFx values were estimated based on genome-wide CNV profiles. TFx values are shown on a log-transformed y-axis, and a pseudo-count of 0.001 was added for visualization. Red and blue dashed lines indicate the ctDNA detection threshold derived from lcWGS data and cfMeDIP-seq data, respectively, defined as the 95th-percentile of values derived from control samples. **(B)** Distribution of the estimated TFx across urine samples from localized PCa patients, advanced PCa patients, and controls. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one sample. The three cohorts were compared using Kruskal-Wallis testing, followed by Dunn's post hoc test. Correction for multiple testing was performed using Benjamini Hochberg method. Only significant differences are shown. The red dashed line indicates the ctDNA detection threshold, defined as the 95th-percentile of values derived from control samples.

3.5.2.2 CNV analysis based on *in silico* size-selected lcWGS data

I applied *in silico* size selection to lcWGS-derived sequencing reads from urinary cfDNA with the aim to enrich for ctDNA in the tumor samples. In this context, I investigated 15 iterated fragment length ranges to identify the appropriate range providing the greatest improvement in the CNV analysis (Figure 25). For this purpose, I compared the resulting TFx values for tumor and controls samples, as well as the number of ctDNA-positive tumor samples with TFx values above the respective ctDNA detection threshold (95th-percentile of TFx values from control samples) between the 15 CNV analyses following size selection for different fragment length ranges.

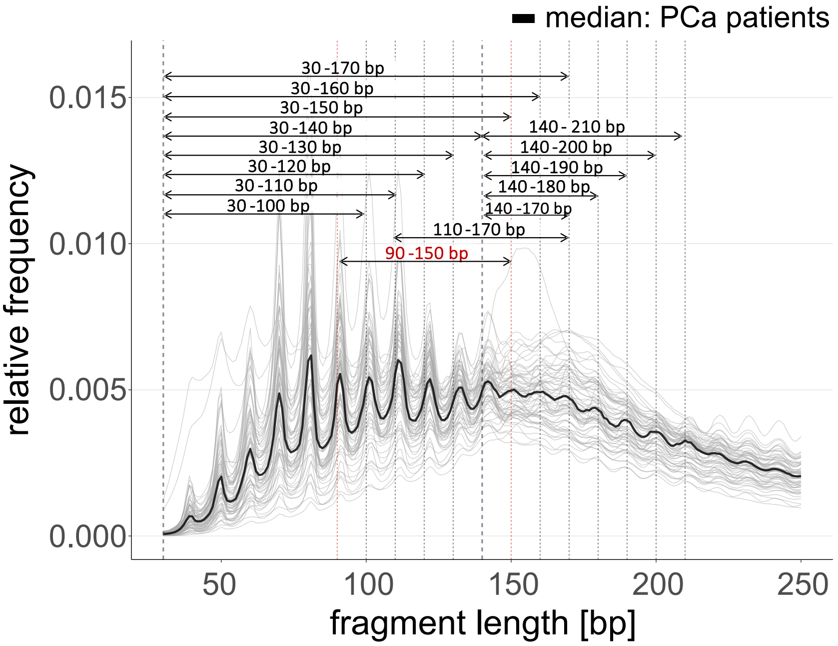


Figure 25: *In silico* size selection for 15 different fragment length ranges prior to CNV analysis of urinary cfDNA. Urinary cfDNA fragmentation profiles from localized and advanced PCa patients are shown. Gray curves represent the relative frequency distributions of urinary cfDNA fragments in individual tumor samples within the 30–250 bp fragment length range. The black line indicates the median distribution across all tumor samples. Black and red arrows denote the selected fragment length ranges, with the red arrow highlighting the 90–150 bp range, which was applied for size selection in the CNV analysis of plasma cfDNA.

Comparison of TFx values derived from urine samples of both PCa patients and controls across CNV analyses with *in silico* size selection for different fragment length ranges revealed that the highest median TFx values were obtained when selecting for longer fragments. The highest-ranking fragment length range was 140–200 bp (median TFx = 0.017), followed by 140–190 bp, 140–170 bp, 140–180 bp, 140–210 bp (median TFx range: 0.0168–0.0170). The same trend was also observed for the tumor samples alone, for which the CNV analyses after size selections for longer fragment length ranges equally revealed the highest TFx values (Figure 26A, Table 16). *In silico* size selection for fragments < 140 bp resulted in similar or lower median TFx values in tumor samples compared with the CNV analysis without size selection (Figure 26A). Median TFx values of tumor samples remained below the respective ctDNA detection thresholds across all analyses. However, size selections consistently reduced the absolute difference between the ctDNA detection threshold and the median TFx of tumor samples compared with CNV analysis without size selection (Figure A13). The CNV analysis without size selection revealed the highest ctDNA detection threshold, whereas the threshold decreased for each of the CNV analyses with *in silico* size selection, independent of the chosen fragment length range (Table 16). Furthermore, the CNV analysis without size selection identified the lowest number of ctDNA-positive tumor samples (n = 8). Size selection for longer fragments resulted in higher ctDNA detection thresholds but fewer ctDNA-positive tumor samples compared with size selection for shorter fragments, which reduced the detection thresholds and increased the number of ctDNA-positive samples (Table 16). CNV analyses with size selection for the fragment length ranges 20–120 bp, 20–140 bp and 20–160 bp resulted in the highest number of ctDNA-positive tumor samples (n = 20), and the size selection for 20–160 bp and 20–140 bp also revealed the lowest and third-lowest detection threshold among all analyses, respectively.

Table 16: CNV analysis following *in silico* size selection across 15 fragment length ranges in urinary cfDNA. Summary of CNV analysis results for urinary cfDNA based on lcWGS following *in silico* size selection across 15 different fragment length ranges. Reported metrics include the median estimated TFx in tumor and control samples, the ctDNA detection threshold defined as the 95th-percentile of TFx values derived from control samples, and the number of ctDNA-positive tumor samples based on this threshold. Metrics are shown for the CNV analysis without size selection and for CNV analyses with size selection across the different fragment length ranges.

size selection	threshold: 95th-percentile	# positive tumor samples	PCa patients (median TFx)	controls (median TFx)
without sizeselection	0.0412	8	0.017	0.013
140–210 bp	0.0376	11	0.017	0.014
140–200 bp	0.0372	12	0.017	0.014
140–170 bp	0.0341	12	0.018	0.015
140–190 bp	0.0320	14	0.017	0.014
140–180 bp	0.0312	12	0.019	0.014
110–170 bp	0.0310	13	0.016	0.014
20–170 bp	0.0308	11	0.017	0.013
20–100 bp	0.0291	15	0.016	0.015
90–150 bp	0.0288	10	0.015	0.013
20–120 bp	0.0257	20	0.016	0.012
20–130 bp	0.0256	17	0.016	0.011
20–110 bp	0.0253	16	0.016	0.012
20–140 bp	0.0239	20	0.016	0.012
20–150 bp	0.0231	18	0.015	0.012
20–160 bp	0.0218	20	0.016	0.010

Abbreviations: TFx = tumor fraction

The CNV analyses revealed that size selection for different fragment length ranges led to changes (decrease/increase) in estimated TFx values. Size selection for longer fragments (110–170 bp and > 140 bp) resulted in similar or slightly increased TFx values compared to the CNV analysis without size selection for both tumor and control samples, whereas size selection for shorter fragments resulted in unchanged or decreased TFx values (Figure 26B). In control samples, no significant differences in the TFx values were observed between analyses with and without size selection for any fragment length ranges (Table A10). For tumor samples, CNV analyses following size selection for the fragment length ranges 20–110 bp ($p = 0.039$; adjusted $p = 0.146$), 20–150 bp ($p = 0.034$; adjusted $p = 0.146$), 20–170 bp ($p= 0.035$; adjusted $p = 0.146$), and 90–150 bp ($p = 0.007$; adjusted $p = 0.098$) resulted in significant differences in the median TFx compared with the CNV analysis without size selection (Table A10). However, these significant differences represented a decline in the median TFx values after size selection (Figure 26B), and the results did not remain significant after correction for multiple testing.

Further analyses on the magnitude and the direction of changes in TFx values revealed that *in silico* size selection for longer fragments resulted in a stronger increase in the TFx values for control samples compared with tumor samples (Table A10). For CNV analyses with size selection for shorter fragment length ranges (20–100 bp to 20–170 bp, and 90–150 bp), TFx values in tumor samples decreased. When restricting the analysis to tumor samples with TFx values above the respective ctDNA detection threshold in the CNV analyses with size selection for different fragment length ranges, and comparing these corresponding TFx values across CNV analyses, no consistent pattern was observed (Figure A13).

Overall, *in silico* size selection across different fragment length ranges showed heterogeneous effects on TFx values. No fragment length range fulfilled both criteria of consistently increasing TFx values in urinary cfDNA from PCa patients compared with CNV analysis without size selection and leading to a higher number of ctDNA-positive patients due to a greater increase in TFx values in tumor samples than in controls.

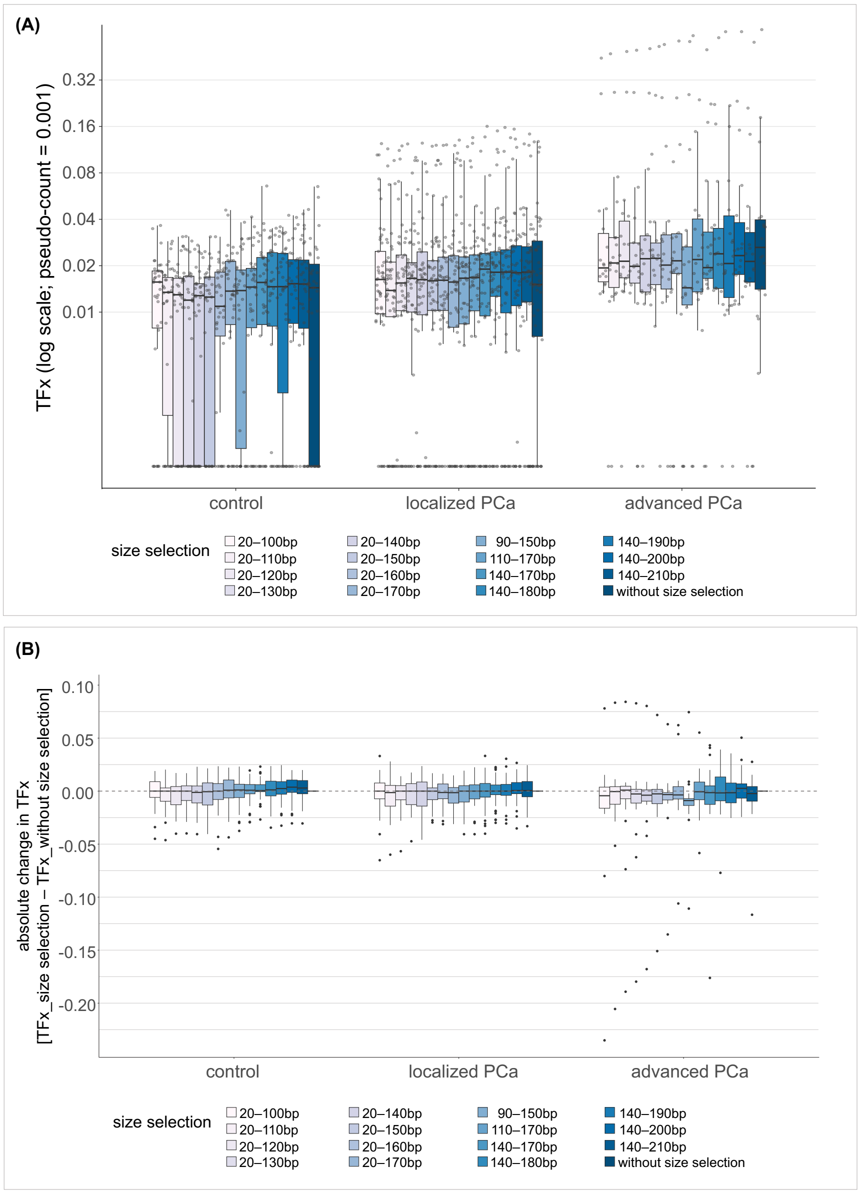


Figure 26: CNV analysis with *in silico* size selection in lcWGS data from urinary cfDNA.
(A) Distribution of the estimated TFx values across urine samples from localized PCa patients, advanced PCa patients, and controls. TFx values were estimated based on genome-wide CNV profiles derived from the CNV analysis after size selection for different fragment length ranges. TFx values are shown on a log-transformed y-axis, and a pseudo-count of 0.001 was added for visualization. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one sample. **(B)** Distribution of the absolute change in estimated TFx values following *in silico* size selection for 15 different fragment length ranges across plasma samples from localized PCa patients, advanced PCa patients and controls. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Dots represent outlying samples. Adapted from Riediger *et al.*[410]

3.5.3 Plasma and urinary cfDNA provided complementary CNV information

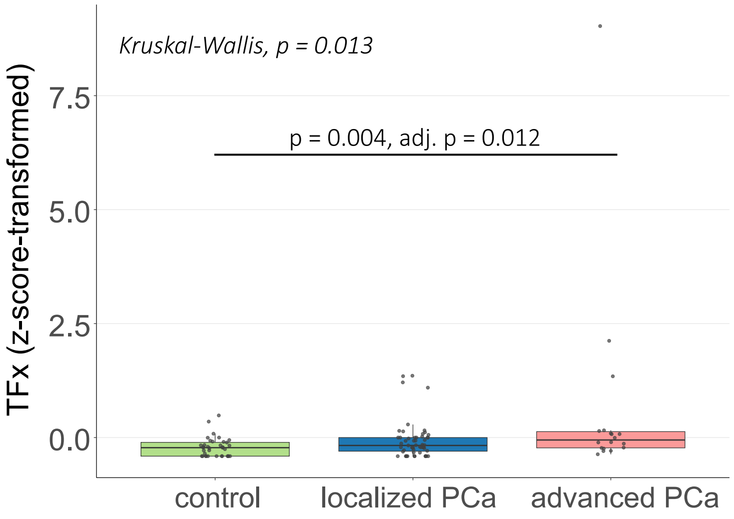
I performed a complementary analysis of CNV data derived from matched plasma and urine samples from PCa patients to assess the presence of common genomic alterations and to compare the resulting TFx values with regard to ctDNA positivity. This comparative analysis revealed both concordant and complementary information in the corresponding plasma and urinary cfDNA.

The CNV analyses without size selection did not identify any pairs of matched plasma and urine samples with detectable CNVs and increased TFx values above the respective ctDNA detection thresholds. Two plasma samples from PCa patients presented distinct genomic alterations, reflected by elevated TFx values > 10%. However, the corresponding urine samples harbored TFx values below the detection threshold, and no genomic alterations could be detected based on their CNV profiles. Following *in silico* size selection for 90–150 bp fragment length in plasma cfDNA, four samples turned ctDNA-positive with increased TFx values above the ctDNA detection threshold, for which the corresponding urine samples also harbored detectable ctDNA based on the CNV analysis without size selection (Table 17; Table A11 and Figure A14). Reversely, CNV analysis in urinary cfDNA identified four samples with elevated TFx values > 10%, whereas the corresponding plasma samples revealed TFx values below the respective ctDNA detection thresholds in both CNV analyses without and with size selection (Table A11 and Figure A14).

Table 17: CtDNA detection in plasma and urinary cfDNA based on the estimated TFx derived from the CNV analysis. Number of localized and advanced PCa patients with detectable ctDNA (TFx above 95th-percentile of values from controls) in plasma or urine separately, for matched plasma and urine samples (ctDNA detected in both biofluids from the same patient), and for complementary ctDNA detection in either plasma or urine. 68 PCa patients with matched plasma and urine samples; 4 localized PCa patients with missing urine samples.

tumor stage	analyzed liquid biopsy sample	# ctDNA-positive patients
localized PCa	plasma, n = 55	5 (9%)
	urine, n = 50	5 (10%)
	matched plasma + urine, n = 50	2 (4%)
	plasma or urine, n = 55	8 (15%)
advanced PCa	plasma, n = 18	4 (22%)
	urine, n = 18	3 (17%)
	matched plasma + urine, n = 18	2 (11%)
	plasma or urine, n = 18	5 (28%)

In order to capture the strongest molecular signal per individual, the higher of the z-score-transformed TFx values from the two biofluids was determined. An increasing trend of these highest per-individual values was seen from controls to localized PCa to advanced PCa patients, with significantly increased values in advanced PCa patients compared to controls (p = 0.004, adjusted p = 0.012; Figure 27). In localized PCa patients, 55% of the patients (n = 30) harbored the highest signal in plasma, whereas for advanced PCa patients, the highest values were more frequently detected in urinary cfDNA (67%, n = 12; Figure 27).



	plasma	urine
localized PCa	30 (55%)	25 (45%)
advanced PCa	6 (33%)	12 (67%)

Figure 27: Highest molecular signal based on TFx values across matched plasma and urine samples. Distribution of highest per-individual TFx values across localized PCa patients, advanced PCa patients, and controls. For each individual, the higher of the z-score-transformed TFx values from matched plasma and urine samples was used to capture the strongest molecular signal across biofluids. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one individual. The three cohorts were compared using Kruskal-Wallis testing, followed by Dunn's post hoc test. Correction for multiple testing was performed using the Benjamini-Hochberg method. Only significant differences are shown. The table adjacent to the plot indicates the proportion of PCa patients in whom the highest signal was detected in plasma or urine, respectively. Adapted from Riediger *et al.*[410]

Overall, ctDNA detection rates based on the presence of CNVs and corresponding increased TFx values were moderate to low (Table 17). For localized PCa patients, a positive tumor signal was identified in 9–10% of plasma and urine samples, while the detection rate was doubled, but still moderate, for advanced PCa patients (plasma: 22%; urine: 17%, Table 17). The proportion of matched plasma and urine samples with detectable ctDNA was also low with 4% and 11% in localized and advanced PCa patients, respectively. The complementary analysis of plasma and urine samples improved ctDNA detection rates for both localized PCa (15%) and advanced PCa patients (28%).

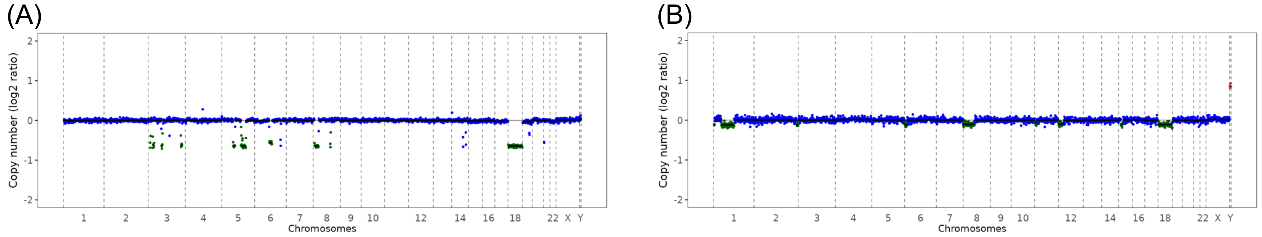
3.5.4 Genomic analysis identified heterogeneous CNV profiles in primary PCa tissue

I additionally performed CNV analysis on sheared gDNA from eight fresh-frozen primary PCa tissue samples, which were obtained during prostatectomy. The tumor cell content of these samples ranged from 40% to 90% (Figure 28), as determined by pathological assessment. I performed the analysis on lcWGS data using the *ichorCNA* algorithm, with genome-wide CNVs assessed in 1000-kb windows. I used eight matched buffy coat samples as copy-number-neutral reference, and additional CNV analysis of these buffy coat samples confirmed the absence of genomic alterations (Figure A15).

Five out of eight PCa tissue samples presented detectable CNVs (Figure 28). Four of these samples harbored predominantly or exclusively deletions (Figure 28A–D). One PCa tissue sample showed pronounced genomic instability compared with the other samples and demonstrated both deletions and amplifications (Figure 28E). The corresponding prostate tumor was advanced, with expansion into the seminal vesicle (pT3b) and lymph node metastases (pN1), and harbored a Gleason 9 pattern. Among the other four PCa tissue samples with detectable CNVs, two were restricted to the prostate gland (pT2c) without lymph node metastases (cN0), whereas the other two were locally advanced, with expansion beyond the prostate capsule or into the seminal vesicle (pT3a/b). One of these two samples additionally harbored lymph node metastases (pN1; Figure 28B), but demonstrated fewer

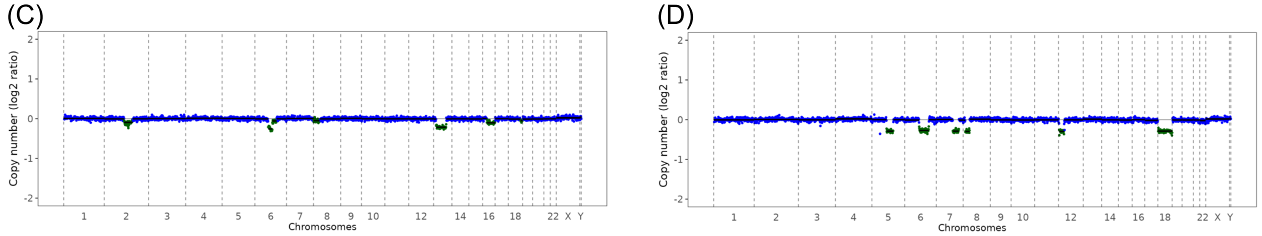
CNVs compared to the advanced PCa with Gleason 9 pattern (Figure 28E). All four PCa tissue samples showed a Gleason 7a pattern. Three additional PCa tissue samples revealed no detectable CNVs despite sufficient tumor cell content (Figure 28F–H). These samples originated from locally advanced PCa with Gleason 7a pattern and expansion beyond the prostate capsule or into the seminal vesicle (pT3a/b), but without lymph node involvement (pN0 or cN0).

In summary, the CNV analysis of eight primary PCa tissue samples identified three different groups: PCa tissues with extensive CNVs (Figure 28E), PCa tissues with moderate CNVs (Figure 28A–D) and PCa tissues with no detectable CNVs (Figure 28F–H).



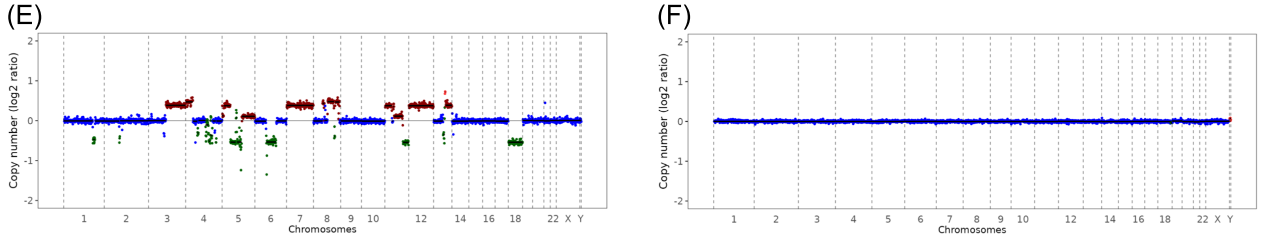
Gleason 7a; pT2c pNx (cN0 cM0) L0 V0 Pn1;
80% tumor cell content

Gleason 7a; pT3b pN1 (1/21) cM0 L1 V0 Pn1;
60% tumor cell content



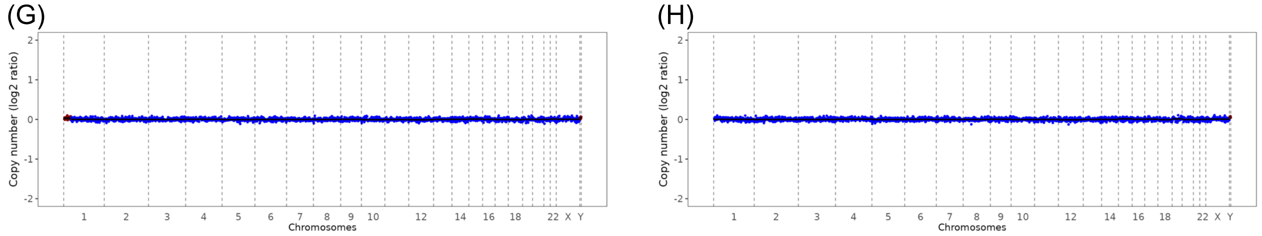
Gleason 7a; pT2c pNx (cN0 cM0) L0 V0 Pn1;
70% tumor cell content

Gleason 7a; pT3a pNx (cN0 cM0) L0 V0 Pn1;
70% tumor cell content



Gleason 9; pT3b pN1 (2/22) cM0 L1 V0 Pn1;
70% tumor cell content

Gleason 7a; pT3b pNx (cN0 cM0) V0 Pn0;
50% tumor cell content



Gleason 7a; pT3a pN0 (0/8) cM0 L1 V0 Pn1;
90% tumor cell content

Gleason 7a; pT3a pNx (cN0 cM0) L0 V0 Pn1;
40% tumor cell content

Figure 28: Genome-wide CNV profiling of gDNA from fresh-frozen PCa tissue samples.
(A–H) Individual CNV profiles of eight PCa tissue samples derived from lcWGS data. CNV profiles display chromosomal gains (red) and losses (green) relative to the copy-number-neutral state (blue). **(A–E)** PCa tissue samples with detectable genomic alterations. **(F–H)** PCa tissue samples without detectable genomic alterations. Abbreviations: cM = clinical assessment of distant metastases, cN/pN = clinical/pathological assessment of lymph node metastases, c/pT = clinical/pathological assessment of the primary tumor's extent, ctDNA = circulating tumor-derived DNA, L0/L1 = absence/presence of lymphatic invasion, M0/M1 = absence/presence of distant metastases, N0/N1 = absence/presence of lymph node metastases, Nx = unknown lymph node status, Pn0/Pn1 = absence/presence of perineural invasion, V0/V1 = absence/presence of venous invasion. Adapted from Riediger *et al.* [410]

The five PCa tissue samples with detectable CNVs harbored recurrent genomic alterations that predominantly presented as deletions, whereas one sample also showed several amplifications (Figure 29). These amplifications were located on chromosomes 3, 5, 7, 8, 11, 12 and 13. Recurrent deletions occurring in 40–50% or more of the tissue samples were identified on chromosomes 5p, 6p, 8p, 12p, 13q and 18.

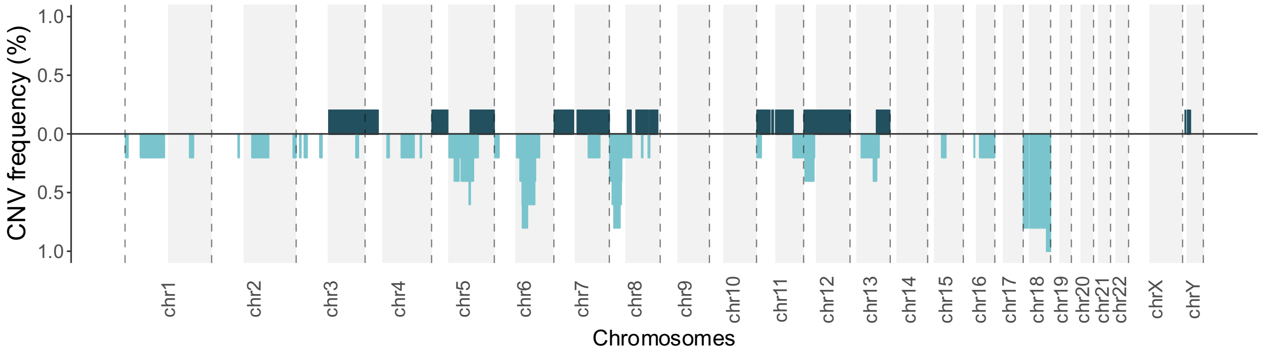


Figure 29: Recurrent CNVs in PCa tissue samples.
Summary of recurrent amplifications and deletions across five PCa tissue samples with detectable CNVs. The CNV profile displays chromosomal gains (dark green) and losses (light green) relative to the copy-number-neutral state. The y-axis indicates the proportion of PCa tissue samples with a detected copy number state (frequency) at each chromosomal position shown on the x-axis. Areas shaded in gray represent the q-arm of the respective chromosome. Adapted from Riediger *et al.*[410]

3.5.5 Recurrent CNVs in liquid biopsies were concordant with PCa tissue profiles

I subsequently examined the presence of shared CNVs across liquid biopsy samples and assessed their concordance with CNV profiles derived from PCa tissue samples.

Two plasma and seven urine samples with detectable CNVs and increased TFx values (> 10%), identified in the CNV analyses with *in silico* size selection for 90–150 bp plasma cfDNA fragments and without size selection of urinary cfDNA fragments, respectively, demonstrated recurrent genomic alterations within plasma or urinary cfDNA and across both biofluids (Figure 30). Samples with increased TFx values were selected to ensure sufficient ctDNA content for an informative assessment of recurrent alterations, as CNV profiles from samples with lower ctDNA content revealed more heterogeneity (Figure A16).

Among the two plasma samples with TFx > 10%, both samples showed amplifications on chromosome 3 and deletions on chromosomes 6, 8, 15, 16, and 18, whereas additional alterations were observed in only one of the two plasma samples (Figure 30A). Furthermore, amplifications on chromosomes 9q, 12q, and 19q, as well as deletions on chromosomes 1p, 2, 4, 8p, 12q, 17, and 19q, were detected in plasma, but were less frequent or not detectable in urine samples. In urinary cfDNA, more recurrent alterations were identified, but with lower individual frequencies, based on the higher number of urine samples with TFx > 10% (n = 7; Figure 30B).

Among plasma and urine samples, recurrent gains were observed on chromosomes 1q, 3, 6, 7, 8q, and 16, while recurrent losses affected chromosomes 5q, 6q, 10, 13q, 15q, 16q, and 18 (Figure 30C). When combining results from plasma and urine samples with TFx > 10% (n = 9 liquid biopsy samples), the reported recurrent alterations were present in more than 40% of samples (Figure 30C).

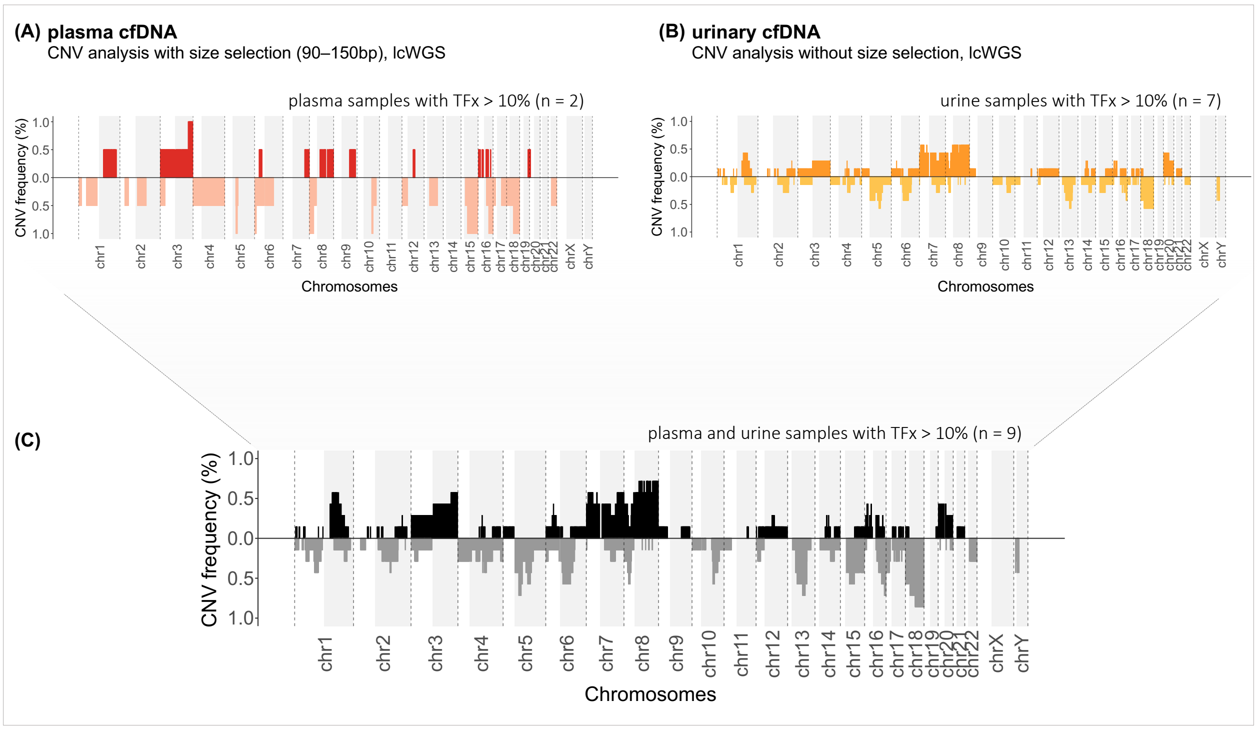


Figure 30: Recurrent CNVs in plasma and urinary cfDNA from PCa patients.
Summary of recurrent amplifications and deletions across plasma (**A**, n = 2) and urine (**B**, n = 7) samples with detectable CNVs and TFx values > 10%. CNV analyses were performed on the lcWGS data applying *in silico* size selection for 90–150 bp fragments in plasma cfDNA and no size selection in urinary cfDNA. **(C)** Recurrent CNVs across plasma and urine samples combined. **(A–C)** CNV profiles display chromosomal gains (dark color) and losses (light color) relative to the copy-number-neutral state. The y-axis indicates the proportion of PCa tissue samples with a detected copy number state (frequency) at each chromosomal position shown on the x-axis. Panels **(A)** and **(B)** show frequencies across two plasma and seven urine samples, respectively, whereas panel **(C)** depicts frequencies across all nine liquid biopsy samples combined. Areas shaded in gray represent the q-arm of the respective chromosome.

Recurrent CNVs identified in liquid biopsy samples were subsequently compared with the findings from the CNV analysis of eight primary PCa tissue samples. The recurrent alterations detected in plasma and urine samples with TFx > 10% were concordant with CNVs identified in five out of eight PCa tissue samples. All plasma and urine samples with detectable CNVs presented both deletions and amplifications (Figure 30), whereas four PCa tissue samples with detectable CNVs harbored only deletions (Figure 28A–D) and one tissue sample additionally showed amplifications (Figure 28E).

Common deletions occurring in more than 40% of both PCa tissue and liquid biopsy samples were found on chromosomes 5q, 6q, 8p, 13q, and 18 (Figure 31). Additional common deletions present in at least 20% of samples were located on chromosomes 1p, 2q, 4q, 7q, 12p, and 16q. Common amplifications present in at least 40% of liquid biopsy samples on chromosomes 3q, 5p, 7, and 8q were also detected in the single PCa tissue sample that harbored amplifications (Figure 31).

The five PCa tissue samples with detectable CNVs were also compared with available matched plasma and urine samples from the same PCa patients. However, all corresponding liquid biopsy samples showed TFx values below the ctDNA detection threshold, and therefore, CNVs detected in PCa tissue could not be reliably compared with CNV profiles derived from plasma and urine samples (Figure A17).

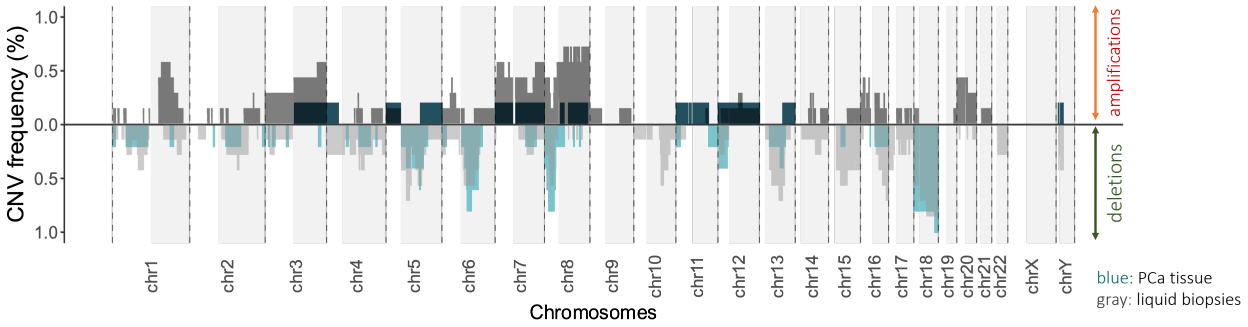


Figure 31: Recurrent CNVs in liquid biopsies and PCa tissue.
Summary of recurrent amplifications and deletions across five PCa tissue samples with detectable CNVs and across liquid biopsy samples comprising two plasma and seven urine samples with detectable CNVs and TFx values > 10%. Recurrent CNVs in PCa tissue samples (green) and liquid biopsy samples (gray) are shown as overlaid frequency profiles. The CNV profile displays chromosomal gains (dark color) and losses (light color) relative to the copy-number-neutral state. The y-axis indicates the proportion of samples with a detected copy-number state at each chromosomal position on the x-axis, calculated separately for PCa tissue samples (n = 5) and liquid biopsy samples (n = 9). Areas shaded in gray represent the q-arm of the respective chromosome. Adapted from Riediger *et al.*[410]

3.6 Chromosomal instability analysis in plasma and urine revealed complementary genomic information

I performed the CIA as complementary genomic assessment of chromosomal instability in plasma and urinary cfDNA in addition to the CNV analysis. Based on the CIA, I determined a synoptic score as a second genomic cfDNA feature, alongside the estimated TFx obtained with the *ichorCNA* algorithm, to reflect genome-wide chromosomal instability in plasma and urinary cfDNA. The CIA score considered genome-wide windows with the highest deviation from the copy-number-neutral state. I calculated the CIA score based on z-score-transformed read counts in genome-wide 1000-kb windows, which were normalized for mappability, GC content and reads from the copy-number-neutral reference, following approaches described in previous publications[400,401].

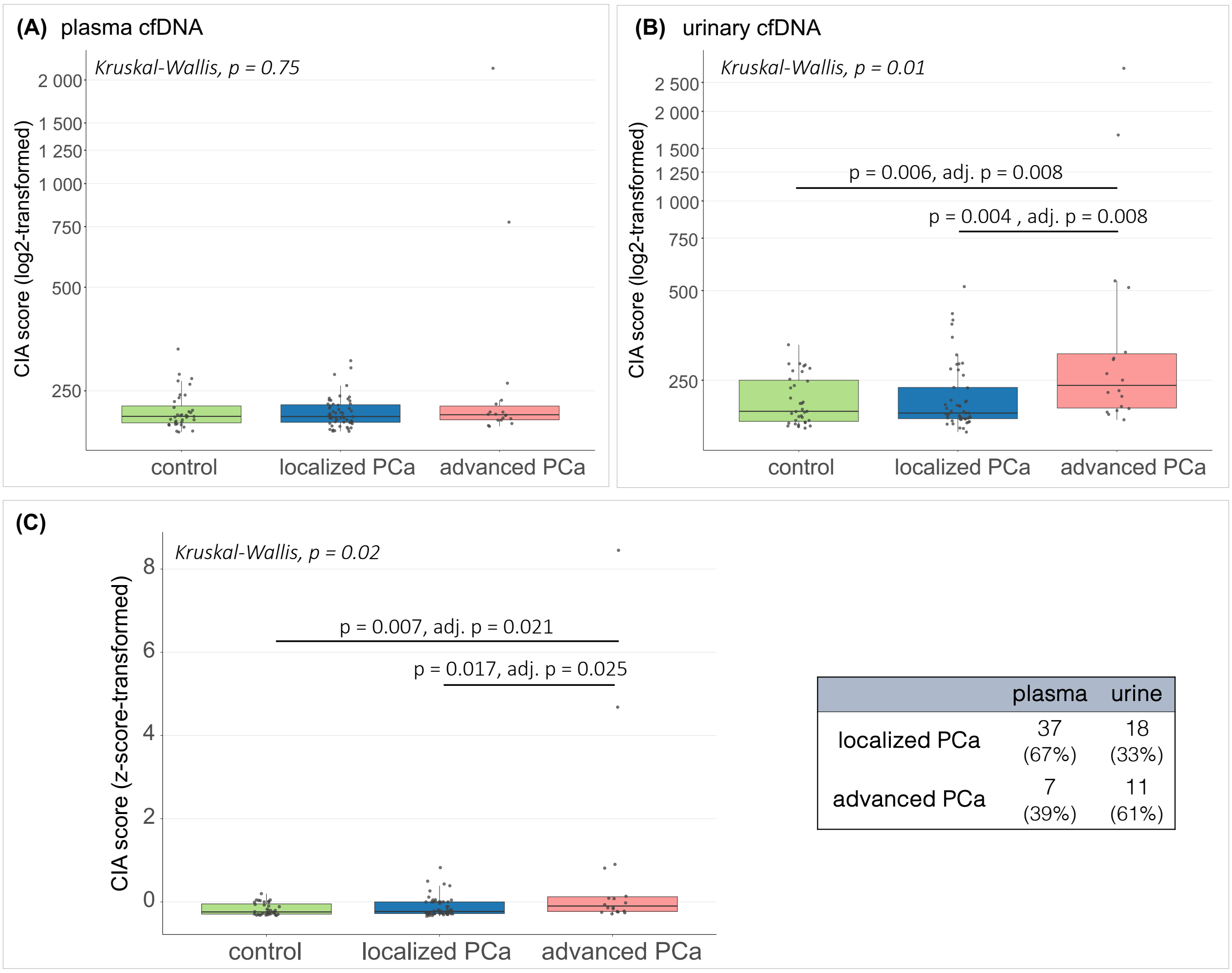
Tumor samples harbored higher median CIA scores than control samples, in both plasma and urinary cfDNA (Figure 32A and B).

In plasma cfDNA, the median CIA score of tumor samples (median: 212.1, range: 190.7–2162.3) was 1.01-fold higher than the median CIA score of control samples (median: 210.8, range: 189.7–330.6). Plasma samples from localized PCa patients showed a median CIA score (median: 210.5, range: 190.7–305.9) comparable to controls, whereas the median CIA score of advanced PCa samples was increased (median: 213.1, range: 197.1–2162.3; Figure 32A). Statistical testing revealed no significant differences among the three cohorts (p = 0.75).

In urinary cfDNA, the median CIA score of tumor samples (median: 204.1, range: 167.4–2788.4) was 1.04-fold higher than the median CIA score of control samples (median: 196.6; range: 172.6–328.9). This trend was primarily driven by urine samples from advanced PCa patients, showing significantly higher CIA scores (median: 240.4, range: 184.1–2788.4) compared with both controls (p = 0.006, adjusted p = 0.008) and localized PCa patients (p = 0.004, adjusted p = 0.008). Urine samples from localized PCa patients showed a slightly lower median CIA score (median: 194.0, range: 167.4–515.7) than controls (Figure 32B).

I next assessed the relationship between CIA scores and the corresponding TFx values derived from the CNV analysis without size selection in plasma and urinary cfDNA from PCa patients. Moderate correlations were observed in plasma (Spearman's ρ = 0.44), whereas correlations were stronger in urine (Spearman's ρ = 0.76; Figure A18).

In order to capture the strongest molecular signal per individual, the higher of the z-score-transformed CIA scores from the two biofluids was determined. A slightly increasing trend in these highest per-individual values was observed from controls to localized PCa to advanced PCa patients, with significantly higher values in advanced PCa patients compared to both localized PCa patients (p = 0.017, adjusted p = 0.025) and controls (p = 0.007, adjusted p = 0.021; Figure 32C). Consistent with the CNV analysis, localized PCa harbored the highest CIA score values more frequently in plasma (67%, n = 37), whereas for advanced PCa, the highest values were more often detected in urinary cfDNA (61%, n = 11; Figure 32C).



	plasma	urine
localized PCa	37 (67%)	18 (33%)
advanced PCa	7 (39%)	11 (61%)

Figure 32: Overview of CIA scores derived from genome-wide assessment of chromosomal instability in plasma and urinary cfDNA.
(A–B) Distribution of CIA scores across **(A)** plasma and **(B)** urine samples from localized PCa patients, advanced PCa patients, and controls. CIA scores are shown on a log-transformed y-axis. **(C)** Distribution of highest per-individual CIA score values across localized PCa patients, advanced PCa patients, and controls. For each individual, the higher of the z-score-transformed CIA score values from matched plasma and urine samples was used to capture the strongest molecular signal across biofluids. The table adjacent to the plot indicates the proportion of PCa patients in whom the highest signal was detected in plasma or urine, respectively. **(A–C)** Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one individual. The three cohorts were compared using Kruskal-Wallis testing, followed by Dunn's post hoc test. Correction for multiple testing was performed using the Benjamini-Hochberg method. Only significant differences are shown. Abbreviations: CIA = chromosomal instability analysis. Adapted from Riediger *et al.*[410]

I further evaluated the CIA score for ctDNA detection, defining the 95th-percentile of CIA score values obtained from control samples as ctDNA detection threshold. Urine samples presented a higher ctDNA detection threshold (283.96) compared with plasma samples (273.36). This assessment identified ctDNA positivity in five plasma and 14 urine samples from PCa patients (Table 18; Table A12).

The two plasma samples with the highest CIA scores in the tumor cohort also showed the highest TFx values based on the CNV analysis. Both plasma samples originated from metastatic PCa patients (M1). Three additional ctDNA-positive plasma samples with CIA score values above the detection threshold originated from patients with intermediate- or high-risk localized PCa. All three plasma samples revealed TFx value of 0.0 in the CNV analysis without size selection, but two of them were ctDNA-positive after *in silico* size selection for 90–150 bp plasma cfDNA fragments. I therefore investigated, whether *in silico* size selection for 90–150 bp plasma cfDNA fragments also improved ctDNA detection based on the CIA in plasma cfDNA. Following size selection, only four out of the five ctDNA-positive plasma samples remained with a CIA score above the ctDNA detection threshold, and the CIA scores were decreased compared to the analysis without size selection (Table A12).

The 14 ctDNA-positive urine samples with CIA score values above the ctDNA detection threshold for urinary cfDNA originated from seven patients with advanced PCa (3x M1, 3x N1M0) and seven patients with localized PCa (1x high-risk, 6x intermediate-risk). Five urine samples with the highest CIA scores in the tumor cohort also harbored high TFx values in the CNV analysis of urinary cfDNA. One additional urine sample with a TFx > 10% showed a CIA score slightly above the ctDNA detection threshold. Overall, six out of eight ctDNA-positive urine samples in the CNV analysis also harbored increased CIA scores above the ctDNA detection threshold, while two samples showed increased TFx values without elevated CIA scores. In one of these two cases, the matched plasma sample was ctDNA-positive based on the CIA in plasma cfDNA.

One metastatic PCa patient harbored matched ctDNA-positive plasma and urine samples based on increased CIA scores (Table 18). In this case, the plasma sample belonged to those with the highest TFx values in the CNV analysis both with and without size selection, whereas the corresponding urine sample was not ctDNA-positive based on the CNV analysis.

In summary, ctDNA detection rates based on CIA scores exceeding the respective detection thresholds were moderate to low for both plasma and urinary cfDNA (Table 18). Compared to the CNV analysis, higher detection rates were achieved for urine samples from both localized and advanced PCa patients, whereas the ctDNA detection rates were halved for plasma samples. For localized PCa patients, ctDNA positivity was identified in 5% and 14% of plasma and urine samples, respectively. For advanced PCa patients, the detection rate was doubled in plasma samples (11%) and 2.8-fold higher in urine samples (39%; Table 18). The proportion of matched plasma and urine samples with detectable ctDNA was low, with only one corresponding pair from a metastatic PCa patient. However, the complementary analysis of plasma and urine samples resulted in increased detection rates for both localized (18%) and advanced PCa patients (44%).

Table 18: CtDNA detection in plasma and urinary cfDNA based on the CIA score.
Number of localized and advanced PCa patients with detectable ctDNA (CIA score above 95th-percentile of values from controls) in plasma or urine separately, for matched plasma and urine samples (ctDNA detected in both biofluids from the same patient), and for complementary ctDNA detection in either plasma or urine. 68 PCa patients with matched plasma and urine samples; 4 localized PCa patients with missing urine samples.

tumor stage	analyzed liquid biopsy sample	# ctDNA-positive patients
localized PCa	**plasma**, n = 55	3 (5%)
	urine, n = 50	7 (14%)
	matched plasma + urine, n = 50	0 (0%)
	plasma or urine, n = 55	10 (18%)
advanced PCa	**plasma**, n = 18	2 (11%)
	urine, n = 18	7 (39%)
	matched plasma + urine, n = 18	1 (6%)
	plasma or urine, n = 18	8 (44%)

Abbreviations: CIA = chromosomal instability analysis

3.7 Genome-wide DNA methylation profiling in liquid biopsies and PCa tissue

3.7.1 Genome-wide identification of DMRs

3.7.1.1 DMRs in plasma and urinary cfDNA distinguished metastatic PCa from non-metastatic disease and controls

I performed genome-wide differential methylation analyses on cfMeDIP-seq data derived from both plasma and urinary cfDNA to investigate differences in methylation status between PCa patients and controls, as well as among PCa subgroups. To this end, I assessed differences in absolute methylation levels, expressed as β-values, across 10,279,967 genomic windows of 300 bp in size to identify DMRs among the cohorts. After filtering out low-coverage regions and retaining only windows with > 2 NRPKMs in at least one sample per comparison, a median of 10.4% (range: 7.8–11.6%) of the initial 300-bp windows remained for subsequent DMR analyses.

In both plasma and urinary cfDNA, no DMRs were detected when comparing all PCa patients with controls, in most comparisons between individual PCa subgroups and controls, or among PCa subgroups (Table A13). However, when metastatic PCa patients (advanced PCa, M1) were compared with either controls or PCa patients without distant metastases (N0/N1 M0), DMRs were identified in both plasma and urinary cfDNA (Table 19, Table A13). *In silico* size selection of plasma cfDNA fragments within the 90–150 bp range, analogous to the CNV analysis, did not improve methylation analyses in plasma samples and resulted in no detectable DMRs across all comparisons. The observed methylation differences were predominantly driven by a small number of plasma or urine samples which showed distinct methylation patterns within the identified DMRs. These samples corresponded to ctDNA-positive tumor samples in the CNV analysis. When the analysis was repeated without these samples, the number of resulting DMRs was substantially reduced (Table A14).

Table 19: Results from the DMR analysis in plasma and urinary cfDNA from metastatic PCa patients compared with controls or PCa patients without distant metastases.
Summary of results from differential methylation analyses performed on plasma and urinary cfDNA from metastatic PCa patients compared with controls or compared with PCa without distant metastases. DMRs were identified based on absolute differences in methylation levels (β-values) calculated in 300-bp genomic windows between the compared cohorts, including only windows with > 2 NRPKMs in at least one sample. Correction for multiple testing was applied using the Benjamini-Hochberg method.

DMR analysis	PLASMA # significant windows	URINE # significant windows
advanced PCa (M1) vs. all controls	# samples: 45 (9 vs. 36)	# samples: 44 (9 vs. 35)
all DMRs	712 / 980,927 windows	48 / 992,774 windows
hypermethylated DMRs [logFC > 1]	615 [490]	29 [22]
hypomethylated DMRs [logFC < -1]	97 [22]	19 [11]
advanced PCa (M1) vs. localized PCa (all) + advanced PCa (N1 M0)	# samples: 73 (9 vs. 64)	# samples: 67 (9 vs. 58)
all DMRs	890 / 1,115,932 windows	64 / 1,112,691 windows
hypermethylated DMRs [logFC > 1]	835 [681]	44 [26]
hypomethylated DMRs [logFC < -1]	55 [7]	20 [9]

Abbreviations: DMR = differentially methylated regions, logFC = log fold change, NRPKM = normalized reads per kilobase million mapped reads

Plasma cfDNA

In plasma cfDNA, 712 DMRs were identified between metastatic PCa patients and controls (Table 19; Figure 33A), and 890 DMRs were detected in the comparison between metastatic PCa patients and PCa patients without distant metastases (Table 19; Figure 33B). The majority of DMRs were hypermethylated in plasma cfDNA from metastatic PCa patients compared to the other two groups (Table 19, Figure 33A and B). Two thirds of all DMRs were located in either distal intergenic or intron regions, whereas ~20% were associated with promoter regions (Figure 33C and D). Methylated CpGs were predominantly detected inside CpG islands or open sea regions within the described genomic regions (Figure A19).

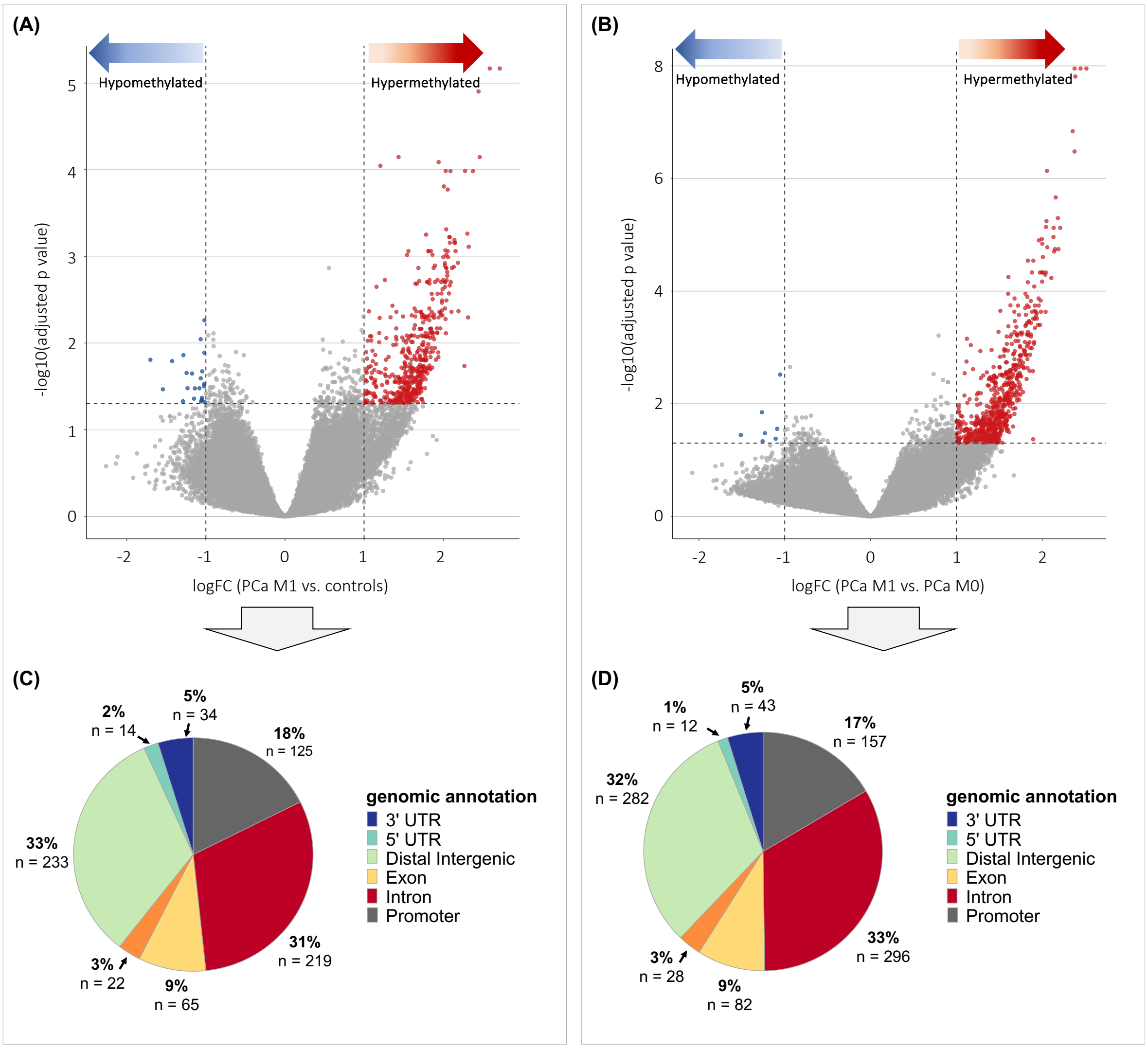
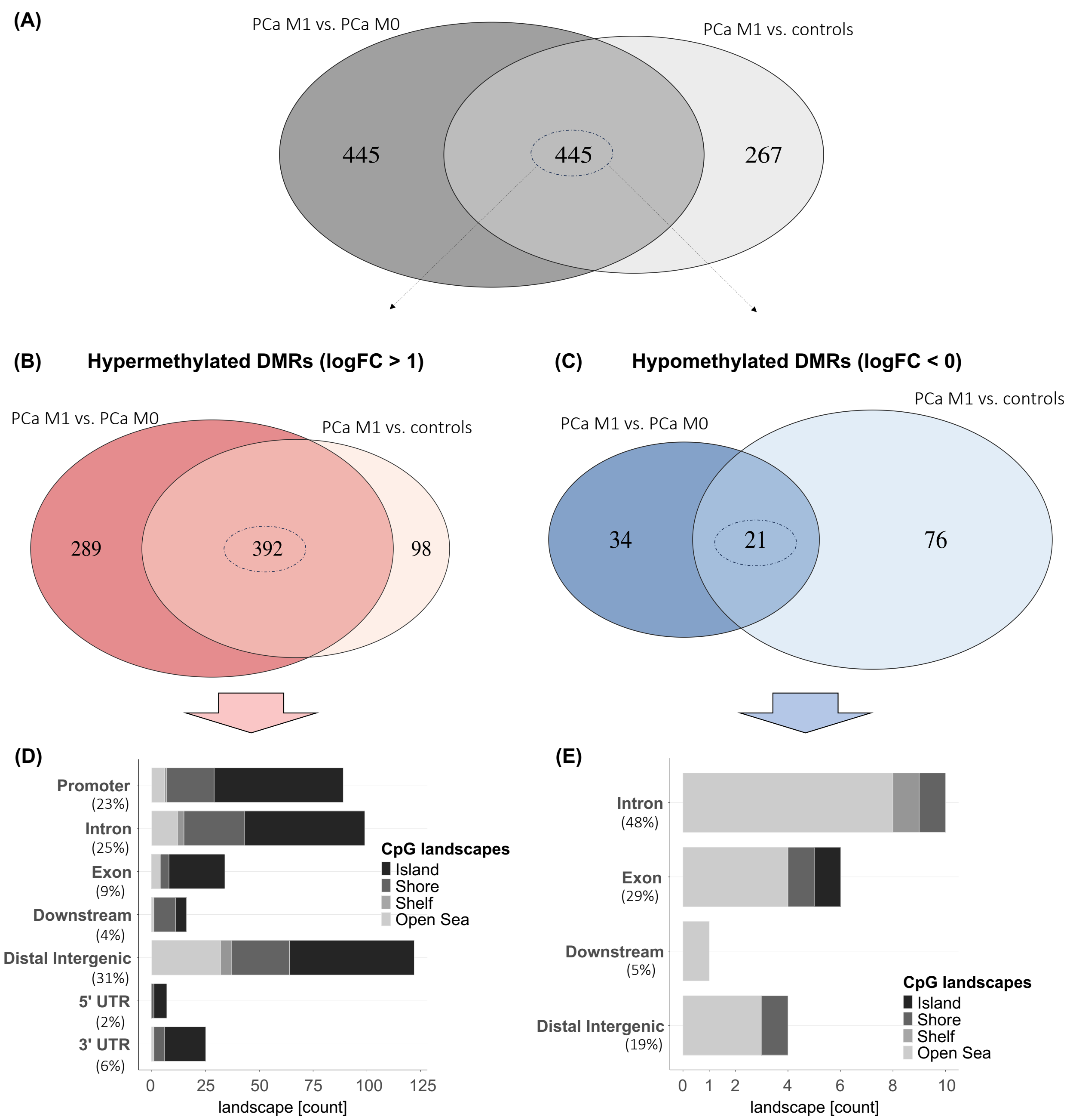


Figure 33: DMR analysis in plasma cfDNA from metastatic PCa patients compared with controls or PCa patients without distant metastases.
Overview of the results from the DMR analysis in plasma cfDNA comparing **(A)** metastatic PCa patients with controls (PCa M1 vs. controls) and **(B)** metastatic PCa patients with PCa patients without distant metastases (PCa M1 vs. PCa M0). **(A–B)** Volcano plots showing the distribution of logFC and adjusted p values across **(A)** 980,927 and **(B)** 1,115,932 genomic regions, respectively. Correction for multiple testing was performed with Benjamini-Hochberg method, and significance was defined as adjusted $p < 0.05$. Vertical dashed lines indicate logFC = ±1 and horizontal dashed lines indicate an adjusted p value of 0.05. **(C–D)** Pie charts illustrating the genomic annotation of significant DMRs, showing their distribution across genomic features (3' or 5'UTR, distal intergenic region, downstream region, exon, intron, or promoter region) for **(C)** 712 DMRs identified in metastatic PCa patients versus controls, and **(D)** of 890 DMRs identified in metastatic PCa patients versus PCa patients without distant metastases. Abbreviations: logFC = log fold change, UTR = untranslated region.

Overlapping the DMRs identified in the two comparisons, metastatic PCa patients (PCa M1) versus controls or metastatic PCa patients versus PCa patients without distant metastases (PCa M0), revealed a total of 445 regions that were differentially methylated in plasma cfDNA from metastatic PCa patients compared with both reference groups (Figure 34A). These regions accounted for 50% of all DMRs identified in the comparison PCa M1 versus controls and 62.5% of all DMRs identified in the comparison PCa M1 versus PCa M0. Among these shared DMRs, 424 regions were hypermethylated (logFC > 0) in metastatic PCa, including 392 regions with a logFC > 1 (Figure 34B). In addition, 21 DMRs were hypomethylated (logFC < 0) in metastatic PCa (Figure 34C), three of which showed a logFC < -1 and were located in two exons and one intron. Approximately half of the 392 hypermethylated regions with logFC > 1 were located in distal intergenic regions or introns, and 90 DMRs (23%) were identified in promoter regions (Figure 34D). Across these genomic contexts, hypermethylated DMRs were distributed among CpG islands, shores, shelves, and open sea regions, with a predominance of CpG islands, followed by CpG shores. Approximately half of the hypomethylated regions were located in introns, followed by exons and distal intergenic regions, and were mainly associated with CpG open sea regions (Figure 34E).

Hierarchical clustering of plasma cfDNA samples from all PCa patients and controls, based on β-values in the 392 hypermethylated DMRs (logFC > 1) identified in metastatic PCa patients compared with both controls and PCa patients without distant metastases, revealed distinct methylation patterns in plasma samples from metastatic PCa patients relative to all other samples (Figure 34F). Plasma samples from metastatic PCa patients did not cluster into a single group, but instead formed two separate clusters located on opposite sides of the heatmap. These clusters also included additional plasma samples from advanced PCa patients with lymph node metastases (N1M0).

Overall, β-values were low in both tumor and control samples. However, two plasma samples from metastatic PCa patients harbored increased methylation levels and deviated from all other samples (Figure 34F). These two samples corresponded to ctDNA-positive tumor samples with the highest TFx in the CNV analyses performed both without and with size selection of plasma cfDNA, thereby complementing the findings of the methylation analysis.


(A)
PCa M1 vs. PCa M0
PCa M1 vs. controls
445
445
267
(B)
Hypermethylated DMRs (logFC > 1)
PCa M1 vs. PCa M0
PCa M1 vs. controls
289
392
98
(C)
Hypomethylated DMRs (logFC < 0)
PCa M1 vs. PCa M0
PCa M1 vs. controls
34
21
76
(D)
Promoter (23%)
Intron (25%)
Exon (9%)
Downstream (4%)
Distal Intergenic (31%)
5' UTR (2%)
3' UTR (6%)
0 25 50 75 100 125
landscape [count]
CpG landscapes
Island
Shore
Shelf
Open Sea
(E)
Intron (48%)
Exon (29%)
Downstream (5%)
Distal Intergenic (19%)
0 1 2 4 6 8 10
landscape [count]
CpG landscapes
Island
Shore
Shelf
Open Sea

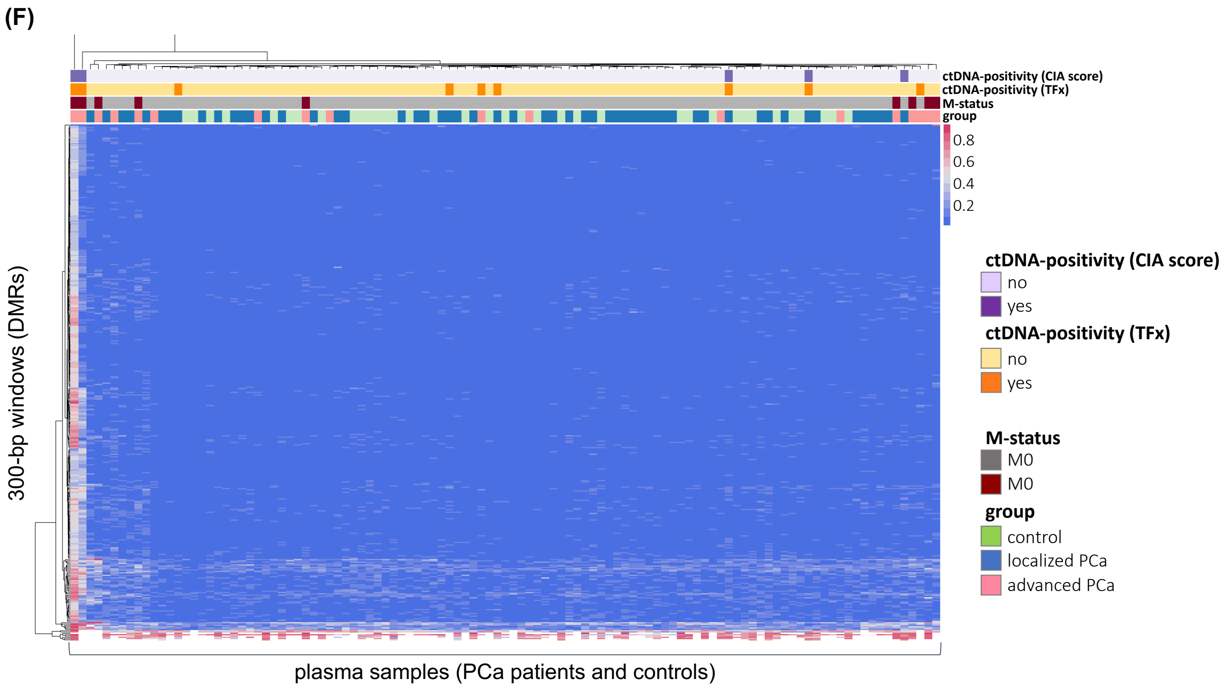


Figure 34: Shared DMRs in plasma cfDNA from metastatic PCa patients compared with PCa patients without distant metastases and controls.
This figure summarizes DMRs identified in both comparisons, metastatic PCa patients versus controls (PCa M1 vs. controls) and metastatic PCa patients versus PCa patients without distant metastases (PCa M1 vs. PCa M0), and highlights regions shared between the two analyses. **(A)** Venn diagram showing all identified and overlapping significant DMRs in plasma cfDNA from the two comparisons PCa M1 vs. controls and PCa M1 vs. PCa M0. **(B–C)** Venn diagrams showing all identified and overlapping significant **(B)** hypermethylated DMRs (logFC > 1) and **(C)** hypomethylated DMRs (logFC < 0) in plasma cfDNA from the same comparisons. **(D–E)** Genomic annotation and CpG context distribution of the overlapping **(D)** hypermethylated DMRs and **(E)** hypomethylated DMRs are shown below the Venn diagrams, including distributions across genomic features (3' or 5'UTR, distal intergenic region, downstream region, exon, intron, or promoter region) and CpG-associated landscapes (CpG island, shore, shelf, or open sea). **(F)** Heatmap of β-values across 392 overlapping hypermethylated DMRs (logFC > 1) identified in plasma cfDNA from metastatic PCa patients compared with both controls and PCa patients without distant metastases. The heatmap includes all plasma samples from PCa patients and controls. Additional annotations indicate tumor status, presence of distant metastases (M0 or M1), and ctDNA positivity (yes/no) based on genomic analyses (TFx and CIA score). Adapted from Riediger *et al.* [410]

Urinary cfDNA

In urinary cfDNA, 48 DMRs were identified between metastatic PCa patients and controls (Table 19; Figure 35A), and 64 DMRs were detected in the comparison between metastatic PCa patients and PCa patients without distant metastases (Table 19; Figure 35B). Similar to plasma cfDNA, the majority of DMRs in urinary cfDNA from metastatic PCa patients were hypermethylated compared with the other two groups (Table 19, Figure 35).

In both comparisons, the largest proportion of all DMRs (42–45%) was located in introns, followed by distal intergenic regions and exons, each representing approximately one fifth of all regions (Figure 35C and D). Promoter regions accounted for less than 10% of all DMRs. The distribution of methylated CpGs across CpG-associated landscapes within these genomic regions was more heterogeneous, with a slight predominance of CpG islands and open sea regions (Figure A19).

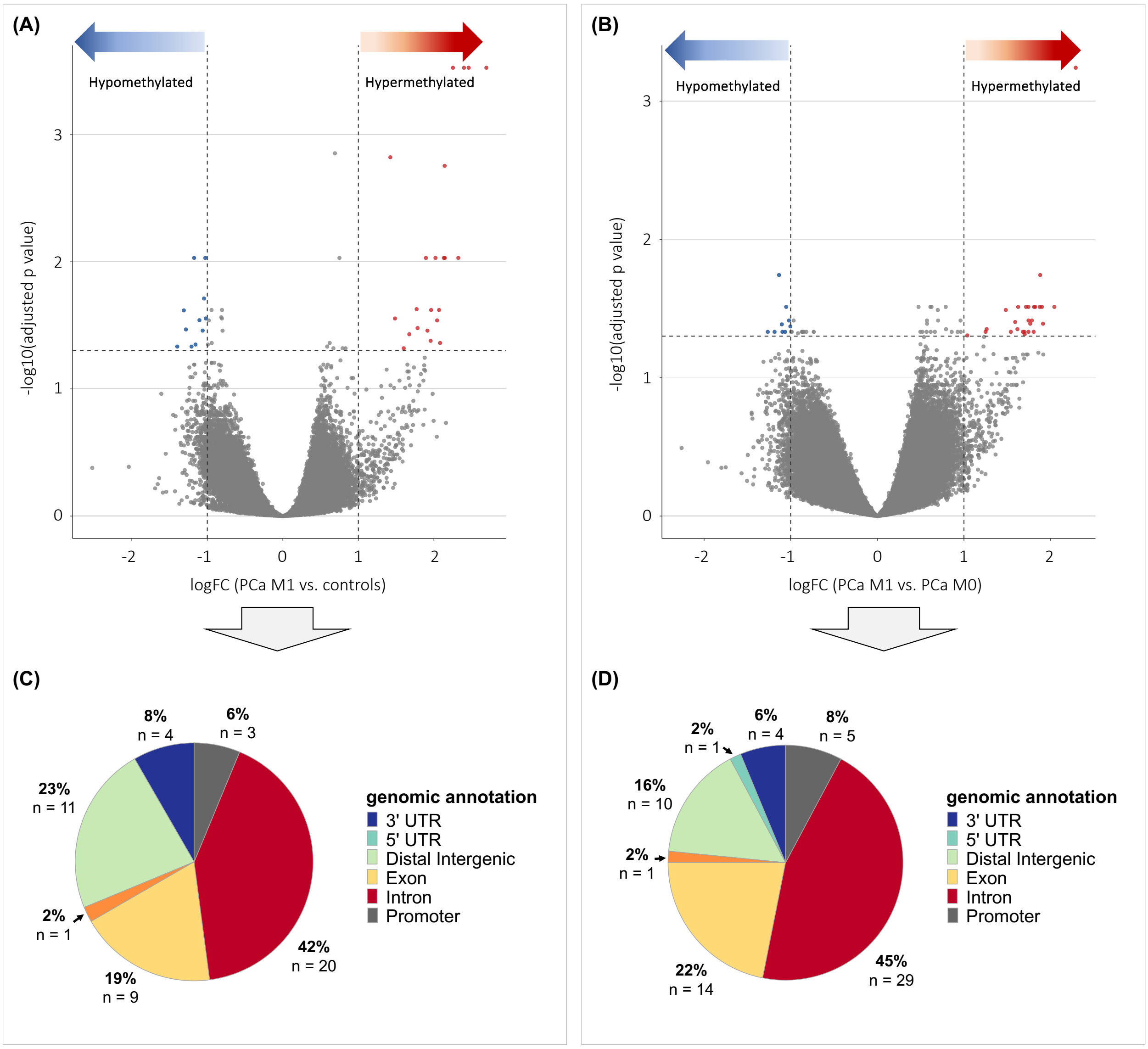
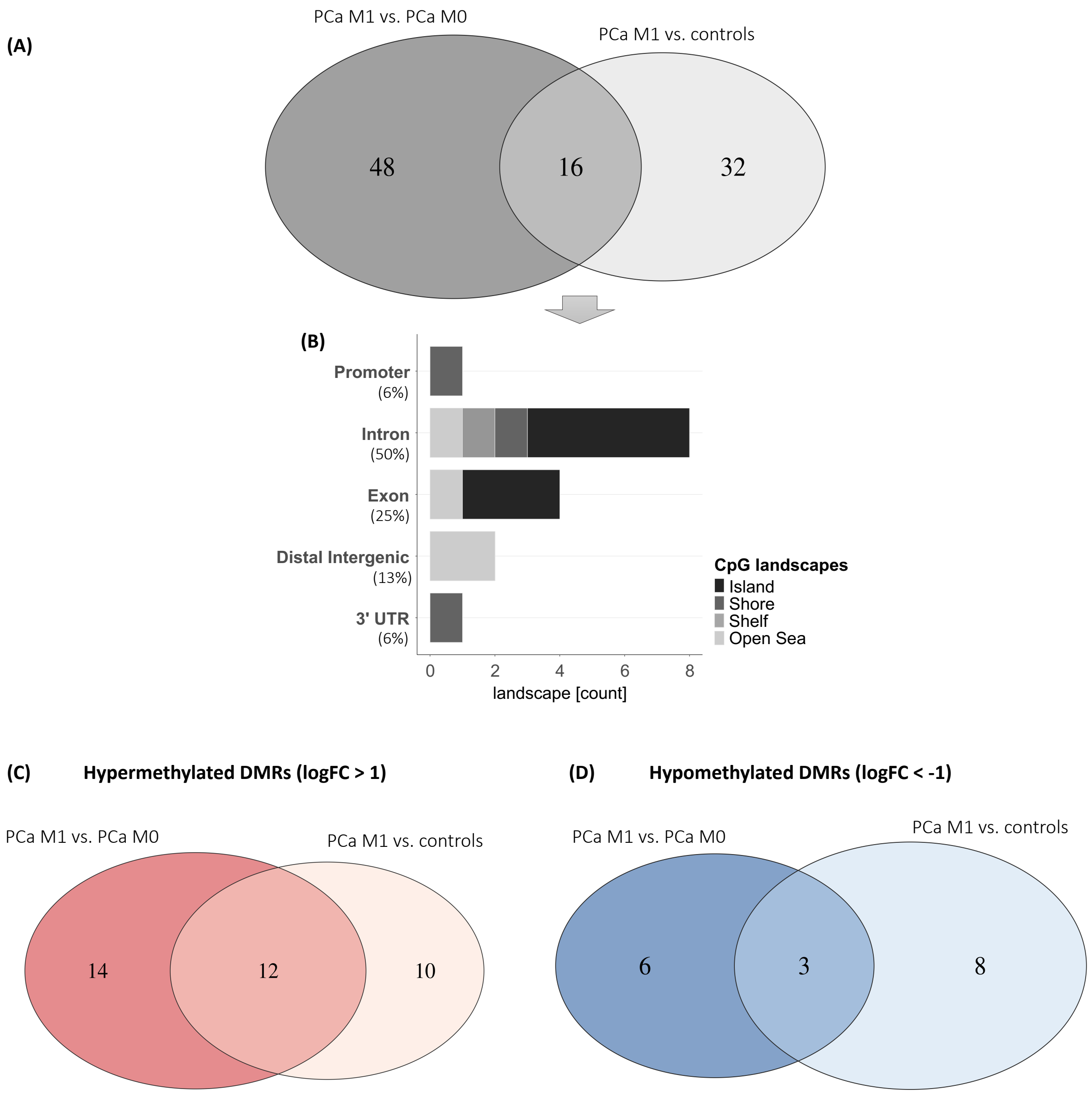


Figure 35: DMR analysis in urinary cfDNA from metastatic PCa patients compared with controls or PCa patients without distant metastases.
Overview of the results from the DMR analysis in urinary cfDNA comparing **(A)** metastatic PCa patients with controls and **(B)** metastatic PCa patients with PCa patients without distant metastases. **(A–B)** Volcano plots showing the distribution of logFC and adjusted p values across **(A)** 992,774 and **(B)** 1,112,691 genomic regions, respectively. Correction for multiple testing was performed with Benjamini-Hochberg method, and significance was defined as adjusted $p < 0.05$. Vertical dashed lines indicate logFC = ±1 and horizontal dashed lines indicate an adjusted p value of 0.05. **(C–D)** Pie charts illustrating the genomic annotation of significant DMRs, showing their distribution across genomic features (3' or 5'UTR, distal intergenic region, downstream region, exon, intron, or promoter region) for **(C)** 48 DMRs identified in metastatic PCa patients versus controls, and **(D)** of 64 DMRs identified in metastatic PCa patients versus PCa patients without distant metastases.

Overlapping the DMRs identified in the two comparisons, metastatic PCa patients (PCa M1) versus controls or metastatic PCa patients versus PCa patients without distant metastases (PCa M0), revealed a total of 16 regions that were differentially methylated in urinary cfDNA from metastatic PCa patients compared with both reference groups (Figure 36A). Of these regions, eight DMRs were located in introns, four in exons, and one hypermethylated DMR was located in a promoter region (Figure 36B). Overall, the 16 shared DMRs comprised 13 hypermethylated regions, of which 12 showed a logFC > 1 (Figure 36C), and three hypomethylated regions, all with a logFC < -1 (Figure 36D). The three hypomethylated regions were located in one intron, one exon and one 3'UTR.

Hierarchical clustering of urinary cfDNA samples from all PCa patients and controls, based on β-values in the 16 hyper- and hypomethylated DMRs identified in metastatic PCa patients compared with both controls and PCa patients without distant metastases, revealed that urine samples from metastatic PCa patients separated from all other samples, analogous to the findings in plasma cfDNA (Figure 36E).

Urine samples from metastatic PCa patients did not cluster into a single group, but instead formed two separate clusters located on opposite sides of the heatmap. The cluster on the left side included a single urine sample from a metastatic PCa patient that demonstrated a prominent methylation pattern and was also ctDNA-positive in the CNV analysis. This sample clustered together with two additional ctDNA-positive urine samples in the CNV analysis, originating from one localized PCa patient and one advanced PCa patient with lymph node metastases (N1M0). A third urine sample from an advanced PCa patient (N1M0) clustered nearby, but the sample was ctDNA-negative in the CNV analysis. On the right side of the heatmap, eight of nine samples from metastatic PCa patients formed a cluster that also included one urine sample from an advanced PCa patient (N1M0) with detectable ctDNA in the CNV analysis. In addition, two urine samples from a localized PCa patient and a cancer-free control were included in this cluster.


(A)
PCa M1 vs. PCa M0
PCa M1 vs. controls
48
16
32
(B)
Promoter (6%)
Intron (50%)
Exon (25%)
Distal Intergenic (13%)
3' UTR (6%)
0
2
4
6
8
landscape [count]
CpG landscapes
Island
Shore
Shelf
Open Sea
(C)
Hypermethylated DMRs (logFC > 1)
PCa M1 vs. PCa M0
PCa M1 vs. controls
14
12
10
(D)
Hypomethylated DMRs (logFC < -1)
PCa M1 vs. PCa M0
PCa M1 vs. controls
6
3
8

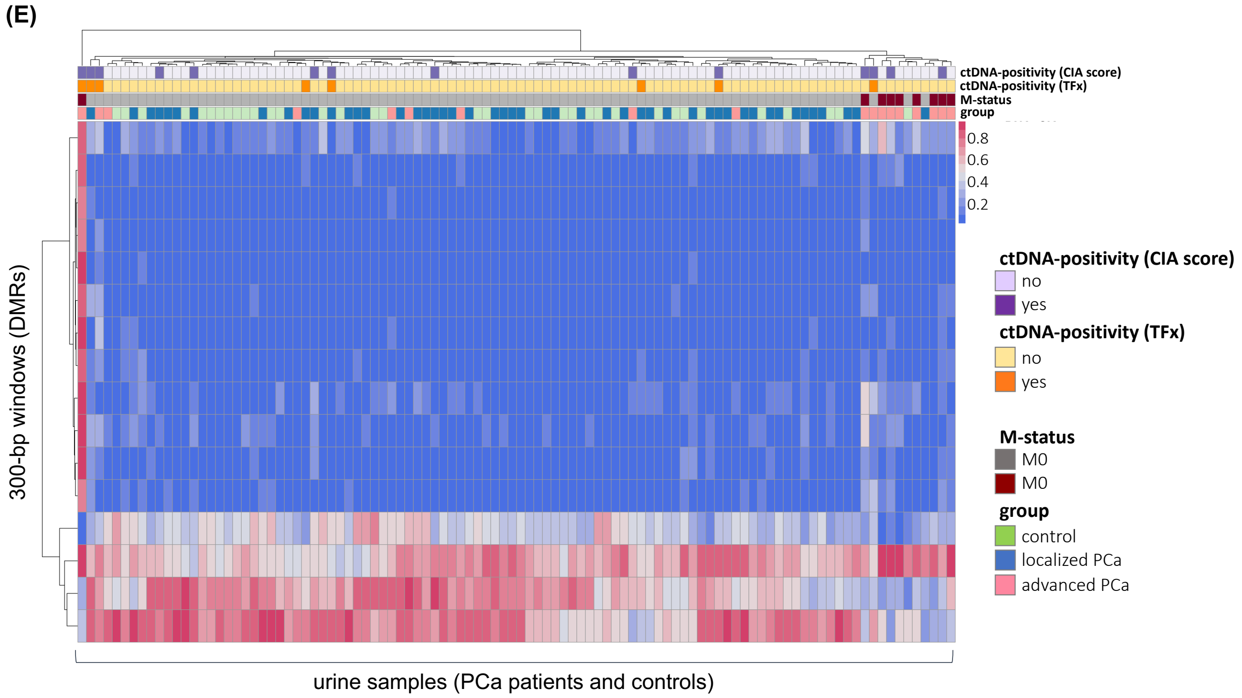


Figure 36: Shared DMRs in urinary cfDNA from metastatic PCa patients compared with PCa patients without distant metastases and controls.
This figure summarizes DMRs identified in both comparisons, metastatic PCa patients versus controls (PCa M1 vs. controls) and metastatic PCa patients versus PCa patients without distant metastases (PCa M1 vs. PCa M0), and highlights regions shared between the two analyses. **(A)** Venn diagram showing all identified and overlapping significant DMRs in urinary cfDNA from the two comparisons PCa M1 vs. controls and PCa M1 vs. PCa M0. **(B)** Genomic annotation and CpG context distribution of the overlapping DMRs, including distributions across genomic features (3’ or 5’UTR, distal intergenic region, downstream region, exon, intron, or promoter region) and CpG-associated landscapes (CpG island, shore, shelf, or open sea). **(C–D)** Venn diagrams showing all identified and overlapping significant **(C)** hypermethylated DMRs (logFC > 1) and **(D)** hypomethylated DMRs (logFC < -1) in urinary cfDNA from the same comparisons. **(E)** Heatmap of β-values across 16 overlapping hyper- and hypomethylated DMRs identified in urinary cfDNA from metastatic PCa patients compared with both controls and PCa patients without distant metastases. The heatmap includes all urine samples from PCa patients and controls. Additional annotations indicate tumor status, presence of distant metastases (M0 or M1), and ctDNA positivity (yes/no) based on genomic analyses (TFx or CIA score). Adapted from Riediger *et al.*[410]

Comparing the results of the DMR analyses in plasma and urinary cfDNA, eight shared DMRs were identified in the comparison between metastatic PCa patients and controls (Supplementary Table S15), and seven shared DMRs were identified in the comparison between metastatic PCa patients and PCa patients without distant metastases across both liquid biopsy sources (Table A16). All shared DMRs were hypermethylated in metastatic PCa patients compared with the respective reference groups.

Three of these hypermethylated regions were identical in plasma and urinary cfDNA. They were located in two introns and one exon on chromosomes 3, 20, and 7, respectively, and were associated with the genes *SOX2 overlapping transcript (SOX2-OT)*, *zinc finger and BTB domain containing 46 (ZBTB46), and protein tyrosine phosphatase receptor type N2 (PTPRN2*).

3.7.1.2 DMRs identified in gDNA from PCa tissue were informative of tumor-associated methylation changes

I also performed differential analysis between eight fresh-frozen PCa tissue and their matched buffy coat samples to identify tumor-specific methylation patterns in PCa tissue. To this end, I assessed DMRs in genome-wide windows of 300 bp in size, after filtering out low-coverage regions and retaining only windows with more than 2 NRPKMs in at least one sample.

In total, 32,146 significant DMRs were identified, corresponding to 3.89% of all analyzed windows (adjusted p value < 0.01). These comprised 18,021 hypermethylated DMRs with logFC > 0 and 14,125 hypomethylated DMRs with logFC < 0, indicating a moderately higher number of hypermethylated than hypomethylated DMRs (1.28-fold). Applying a more stringent logFC threshold based on differences in β-values between the two cohorts, 6,015 hypermethylated DMRs with logFC > 2 and 1,364 hypomethylated DMRs with logFC < -2 remained, representing a 4.4-fold higher number of hypermethylated compared with hypomethylated regions (Figure 37). The largest proportion of both hyper- and hypomethylated DMRs was located in introns, followed by distal intergenic regions. Hypomethylated DMRs were predominantly associated with CpG open sea regions, and hypermethylated DMRs were mainly located in CpG islands or open sea regions (Figure A20). Promoter regions accounted for 13% of hypermethylated and 6% of hypomethylated DMRs, respectively (Figure 37B and C).

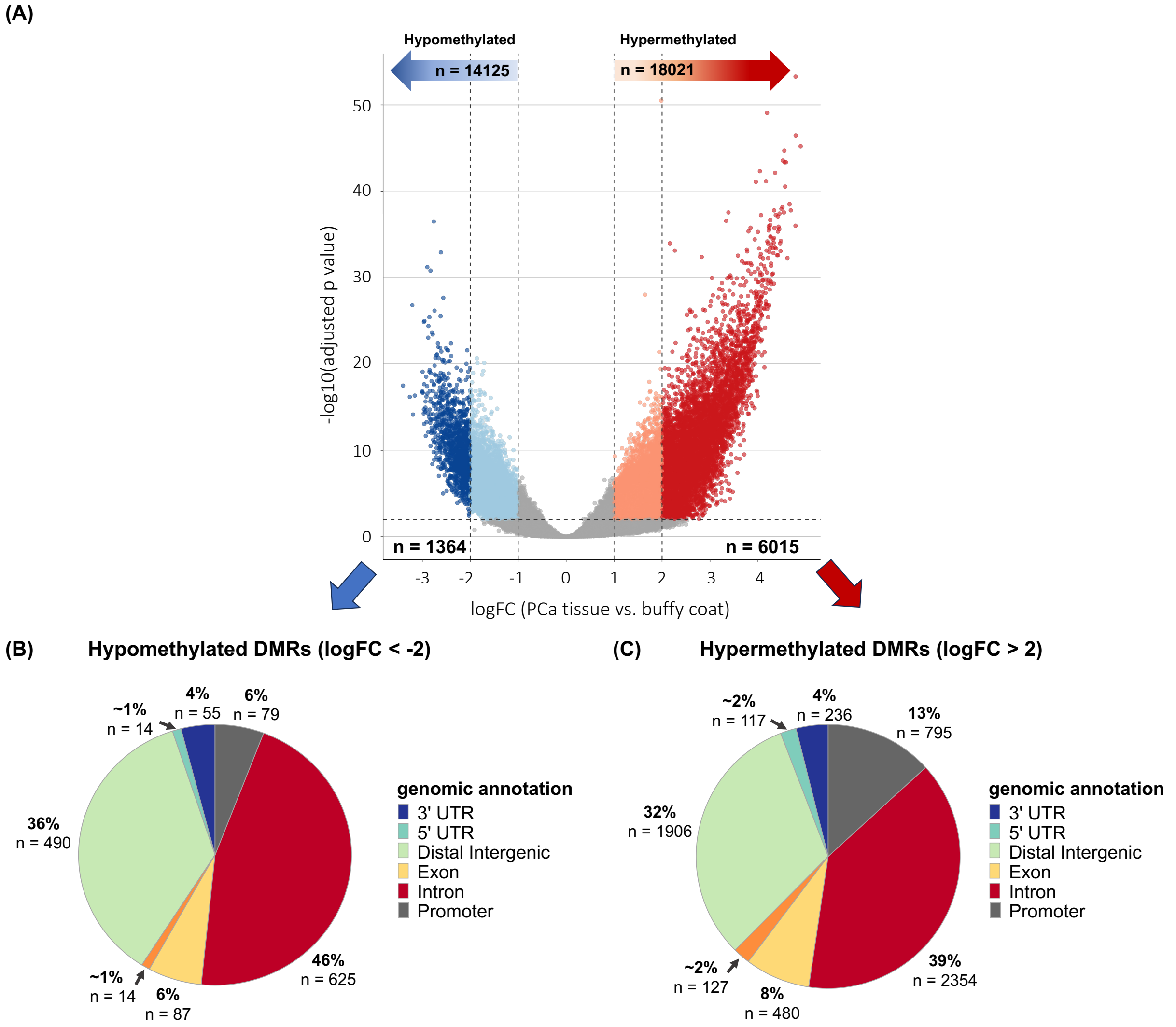


Figure 37: DMR analysis in gDNA from PCa tissue samples compared with matched buffy coat samples. **(A)** Overview of the results from the DMR analysis in gDNA comparing PCa tissue with matched buffy coat samples. The Volcano plot shows the distribution of logFC and adjusted p values across 810,174 genomic regions. Correction for multiple testing was performed with Benjamini-Hochberg method, and significance was defined as adjusted p < 0.01. Vertical dashed lines indicate logFC = ±1 and ±2, respectively. The horizontal dashed line indicates an adjusted p value of 0.01. **(B–C)** Pie charts illustrating the genomic annotation of significant DMRs, showing their distribution across genomic features (3’ or 5’UTR, distal intergenic region, downstream region, exon, intron, or promoter region) for **(B)** 1,364 hypomethylated DMRs and **(C)** 6,015 hypermethylated DMRs identified in PCa tissue samples versus buffy coat samples. Adapted from Riediger *et al.* [410]

I validated the results from the DMR analysis in PCa tissue samples with an external PCa tissue methylation dataset published by Börno and colleagues[403]. The authors performed DMR analysis based on MEDIP-seq data from 51 PCa tissue samples and 53 normal prostate tissue samples. All tissue samples originated from newly diagnosed, untreated PCa patients who underwent prostatectomy. PCa tissue samples contained more than 70% tumor cell content. For normal prostate tissue samples, only sections containing exclusively normal tissue material with an epithelial cell content of 20–40% were included. The authors performed DMR analysis using RPM values in genome-wide 500-bp windows, and identified 146,714 DMRs between PCa tissue samples and normal prostate tissue samples, of which 85,406 were hypermethylated and 61,308 were hypomethylated in PCa tissue. The identified DMRs were ranked according to their significance level, determined using the Mann-Whitney U test, with p values ranging from 7.90e-19 to 1.07e-17. The authors reported the "TOP 100" hypermethylated DMRs with the lowest p values[403], comprising 100 individual 500-bp windows and 92 merged windows of 1,000 bp in size.

I compared the 6,015 hypermethylated DMRs with logFC > 2 identified in PCa tissue samples in my DMR analysis to the TOP 100 hypermethylated DMRs reported by Börno and colleagues[403] (Figure 38A). This comparison revealed that 55 of the 100 unique 500-bp windows from the external methylation dataset overlapped with 67 unique 300-bp windows from my analysis. Permutation testing based on 5,000 random iterations confirmed that the overlap between the two region sets was significant and non-random (p = 2.00e-04). Eight of the 67 hypermethylated regions belonged to the highest 10% of regions ranked by logFC among all hypermethylated DMRs with logFC > 0 identified in my DMR analysis. By contrast, none of the identified hypomethylated regions in my DMRs analysis overlapped with the TOP 100 hypermethylated DMRs reported by Börno and colleagues. One third of the overlapping hypermethylated regions was located in introns, and approximately another third (29%) was associated with promoter regions (Figure 38B). The methylated CpGs were predominantly located in CpG islands.

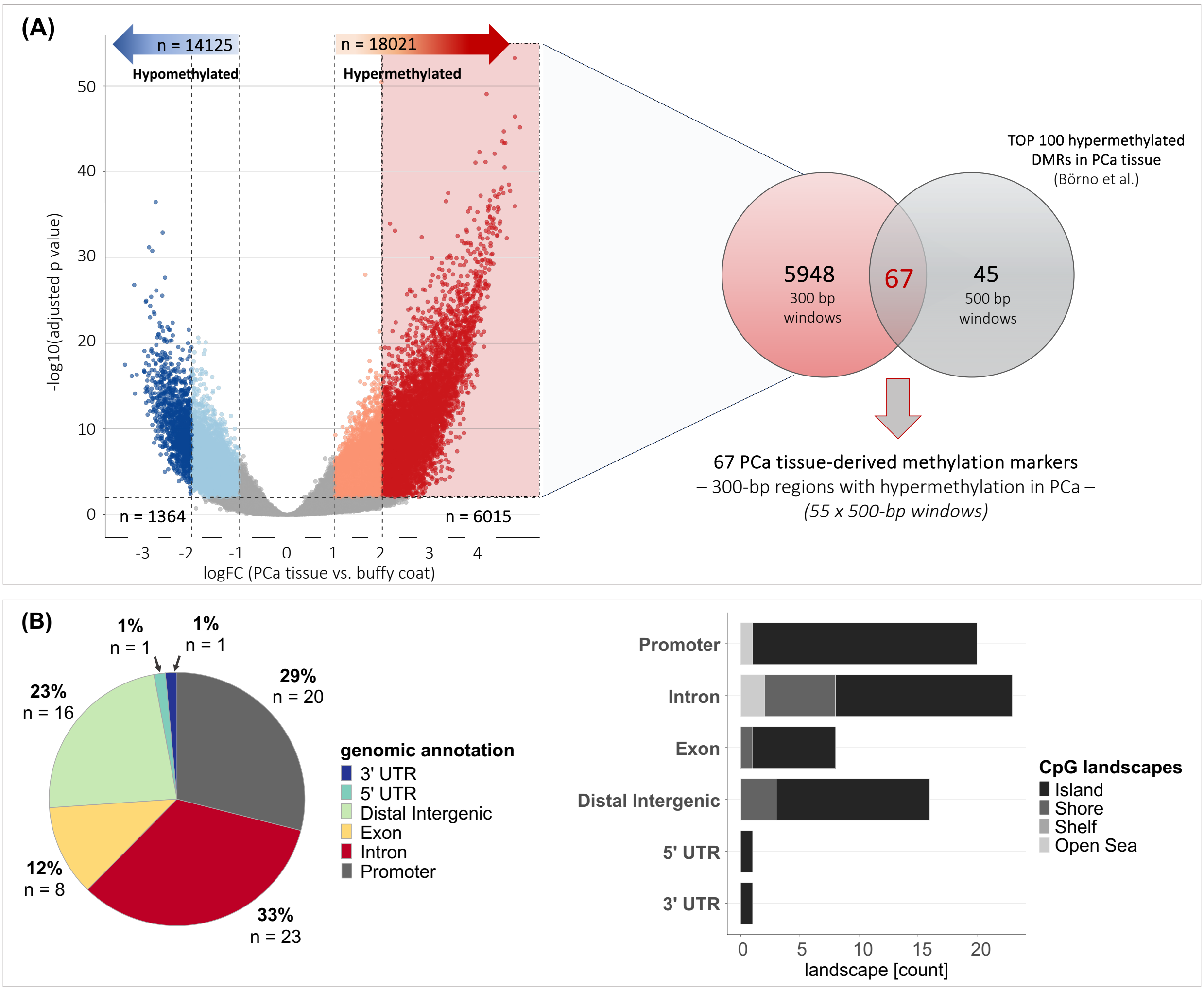
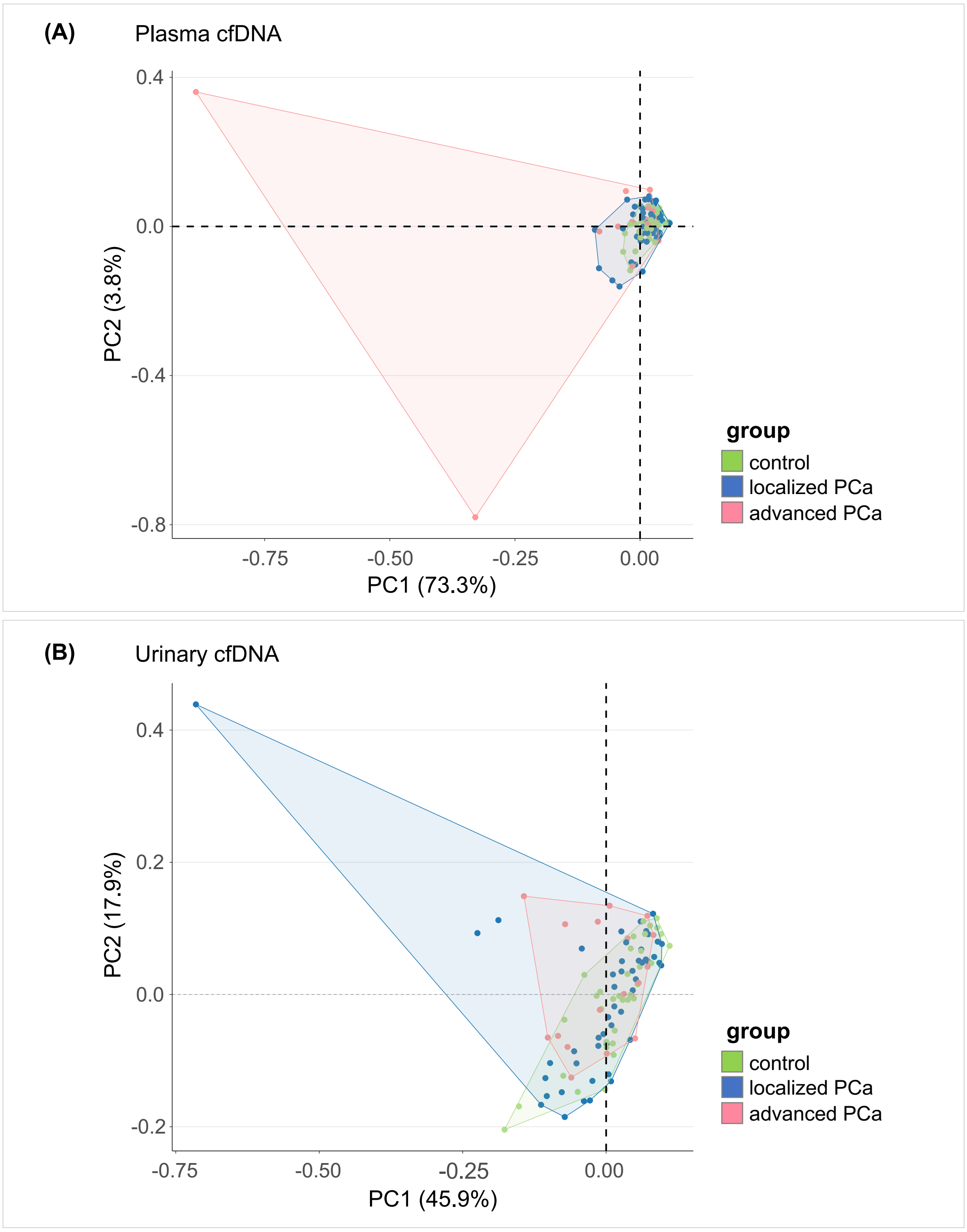


Figure 38: Identification of PCa tissue-derived methylation marker regions.
(A) DMR analysis between PCa tissue samples and matched buffy coat samples identified significant hyper- and hypomethylated regions in PCa tissue. A total of 6,015 hypermethylated regions (300-bp windows; logFC > 2) identified in PCa tissue were intersected with the 100 top-ranked hypermethylated DMRs (500-bp windows) obtained from an external PCa tissue dataset (Börno et. al. [403]), resulting in 67 overlapping regions (300-bp windows). These PCa tissue-derived methylation marker regions were subsequently applied to cfMeDIP-seq data from liquid biopsy samples to assess methylation levels in plasma and urinary cfDNA. **(B)** Genomic annotation and CpG context distribution of the 67 PCa tissue-derived methylation marker regions, showing their distributions across genomic features (3' or 5'UTR, distal intergenic region, downstream region, exon, intron, or promoter region) and CpG-associated landscapes (CpG island, shore, shelf, or open sea). Adapted from Riediger *et al.* [410]

I applied the 67 shared hypermethylated DMRs identified in PCa tissue in my DMR analysis and validated using the external methylation dataset reported by Börno and colleagues[403] to the cfMeDIP-seq data from liquid biopsy samples. Subsequently, I assessed the methylation status of tumor and control samples within these regions.

Unsupervised clustering by PCA of cancer-free controls, localized PCa patients, and advanced PCa patients revealed that the three cohorts did not cluster separately based on β-values in the 67 selected 300-bp windows, neither in plasma cfDNA nor in urinary cfDNA (Figure 39A and B). However, strong outlier samples were observed in both liquid biopsy sources.

Hierarchical clustering of plasma and urine samples from all PCa patients and controls based on β-values within the 67 methylation marker regions showed that β-values were low in both tumor and control samples (Figure 39A). Consistent with the PCA results, as small number of plasma and urine samples from advanced PCa patients exhibited distinctly increased β-values. These samples were also ctDNA-positive based on the genomic analyses. One plasma and one urine sample with the highest TFx values and CIA scores clustered closely together on the right side of the heatmap and displayed methylation patterns that differed from all other samples. These two samples were located adjacent to a cluster of urine samples with slightly increased β-values in a limited number of windows, alongside two additional clusters that predominantly contained either plasma or urine samples. On the left side of the heatmap, another mixed cluster of plasma and urine samples presented moderately increased β-values across several windows. These samples mainly originated from advanced PCa patients, with a small number of samples from localized PCa patients. Approximately half of these samples showed ctDNA-positive results in the genomic analyses, either based on the estimated TFx from the CNV analysis or based on the assessment of the CIA score.



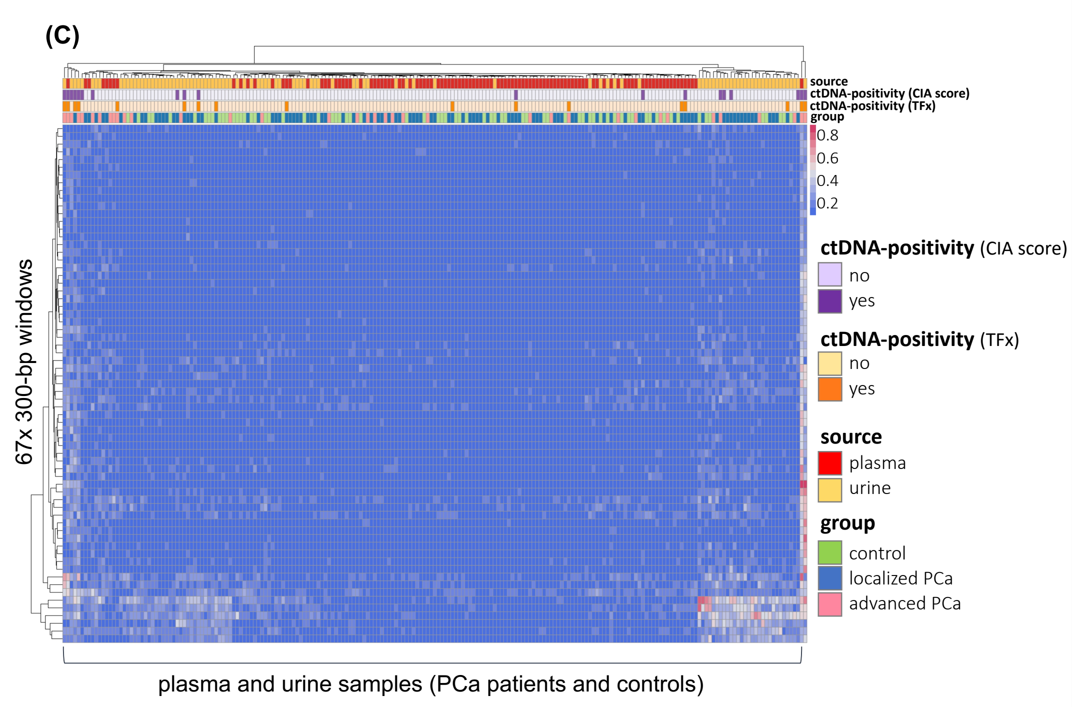
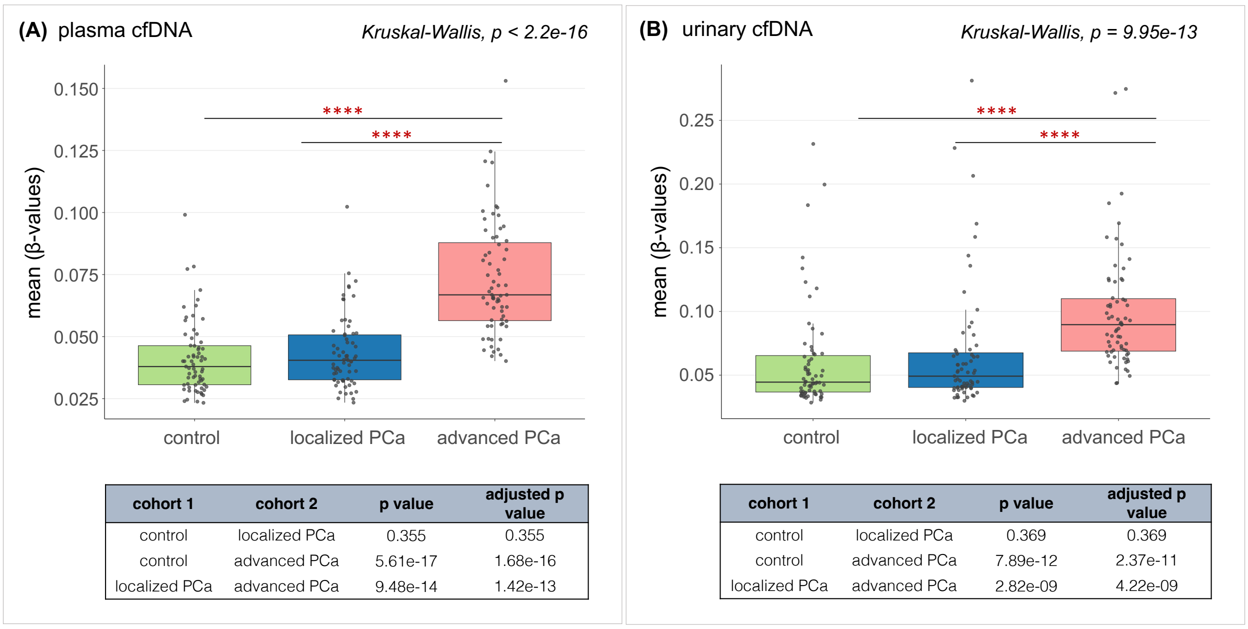


Figure 39: Unsupervised clustering based on 67 PCa tissue-derived methylation marker regions in plasma and urinary cfDNA.

(A–B) Unsupervised clustering based on PCA of β-values across 67 PCa-tissue derived methylation marker regions for **(A)** all plasma samples and **(B)** all urine samples from PCa patients and controls. The x- and y-axes of the PCA plots represent the first and second principal component, explaining the largest and second largest proportions (%) of the variance, respectively. **(C)** Heatmap of β-values across 67 PCa tissue-derived methylation marker regions, including all plasma and urine samples from PCa patients and controls. Additional annotations indicate biofluid type, tumor status, and ctDNA positivity (yes/no) based on genomic analyses (TFx and CIA score). Abbreviations: PC1/2 = principal component 1/2, PCA = principal component analysis

I compared the distributions of β-values in each of the 67 PCa tissue-derived marker regions among controls, localized PCa patients, and advanced PCa patients. For each region, β-values were calculated as the mean across all plasma or urine samples within each cohort, respectively. Advanced PCa patients harboured significantly higher mean β-values across the 67 regions compared with localized PCa patients or controls in both plasma and urine (Figure 40; Figure A21; Table A17).



cohort 1	cohort 2	p value	adjusted p value
control	localized PCa	0.355	0.355
control	advanced PCa	5.61e-17	1.68e-16
localized PCa	advanced PCa	9.48e-14	1.42e-13

cohort 1	cohort 2	p value	adjusted p value
control	localized PCa	0.369	0.369
control	advanced PCa	7.89e-12	2.37e-11
localized PCa	advanced PCa	2.82e-09	4.22e-09

Figure 40: Distribution of methylation levels across 67 PCa-tissue derived methylation marker regions in plasma and urinary cfDNA.
Distribution of mean β-values across 67 PCa-tissue derived methylation marker regions in **(A)** plasma and **(B)** urinary cfDNA from localized PCa patients, advanced PCa patients, and controls. Each data point represents the mean *β*-value across all plasma or urine samples, respectively, within each cohort for a single marker region. The three cohorts were compared using Kruskal-Wallis testing, followed by Dunn's post hoc test. Correction for multiple testing was performed with Benjamini-Hochberg method. Only significant differences are shown in the figures, and a summary of statistical testing results is provided in the tables below the figures. Significance codes: ****1e-04, ***0.001, **0.01, *0.05

In addition, I applied the 67 PCa tissue-derived DNA methylation marker regions to an external liquid biopsy-based methylation dataset from PCa patients. Chen and colleagues performed cfMeDIP-seq on 133 plasma samples[354], including 30 samples from untreated patients with localized PCa and 103 samples from advanced patients with mCRPC who had already received systemic therapy. Plasma samples from mCRPC patients originated from three different cohorts and were collected either at baseline, during treatment or at timepoint of progression. I assessed absolute methylation levels within the 67 methylation marker regions in plasma cfDNA from the external cohort of localized PCa patients and mCRPC patients.

Both PCA and hierarchical clustering revealed a distinct separation between plasma samples from localized PCa patients and mCRPC patients based on β-values in the 67 genomic regions (Figure 41; Figure A22). I equally performed PCA using β-values from 67 randomly selected genomic regions with a CpG density > 6, comparable to the CpG density in the 67 methylation marker regions (Figure A22). In this analysis, plasma samples from localized PCa patients and mCRPC patients also formed two clusters, but the separation was less pronounced compared with the separation observed using the 67 methylation marker regions.

I also performed hierarchical clustering of all plasma samples from the external cohort published by Chen and colleagues[354], combined with either plasma or urine samples from my cohort, based on β-values in the selected 67 methylation marker regions (Figure 41). In both liquid biopsy sources, most samples from cancer-free controls and primary localized or advanced PCa patients in my cohort clustered together with plasma samples from localized PCa patients in the external cohort.

The majority of plasma samples from my cohort and plasma samples from localized PCa patients in the external cohort harbored low β-values, whereas mCRPC patients from the external cohort revealed high β-values in most of the 67 genomic regions. Two plasma samples from advanced PCa patients in my cohort clustered together with plasma samples from mCRPC patients (Figure 41A). These two plasma samples had previously been identified as strong outliers in my methylation dataset and equally showed detectable ctDNA based on increased TFx values in the CNV analysis, indicating the presence of a pronounced tumor signal.

Urine samples from my cohort also harbored low β-values within the 67 methylation marker regions, except for some regions in which β-values were increased in a subset of urine samples (Figure 41B). Two distinct clusters comprising exclusively urine samples were identified. One cluster was located adjacent to the group of external plasma samples from mCRPC patients and shared a few methylation patterns with this group. The other cluster was located on the left side of the heatmap and included urine samples with increased β-values in a limited number of genomic regions. One urine sample from an advanced PCa patient in my cohort clustered together with plasma samples from mCRPC patients. This urine sample had previously been identified as a strong outlier in my methylation dataset and also was ctDNA-positive in the CNV analysis. However, this urine sample did not correspond to either of the two previously described outlying plasma samples from my cohort. Eight additional urine samples from two localized PCa patients and six advanced PCa patients clustered with the group of mCRPC patients, although they were positioned at the outermost left edge of the cluster (Figure 41B).

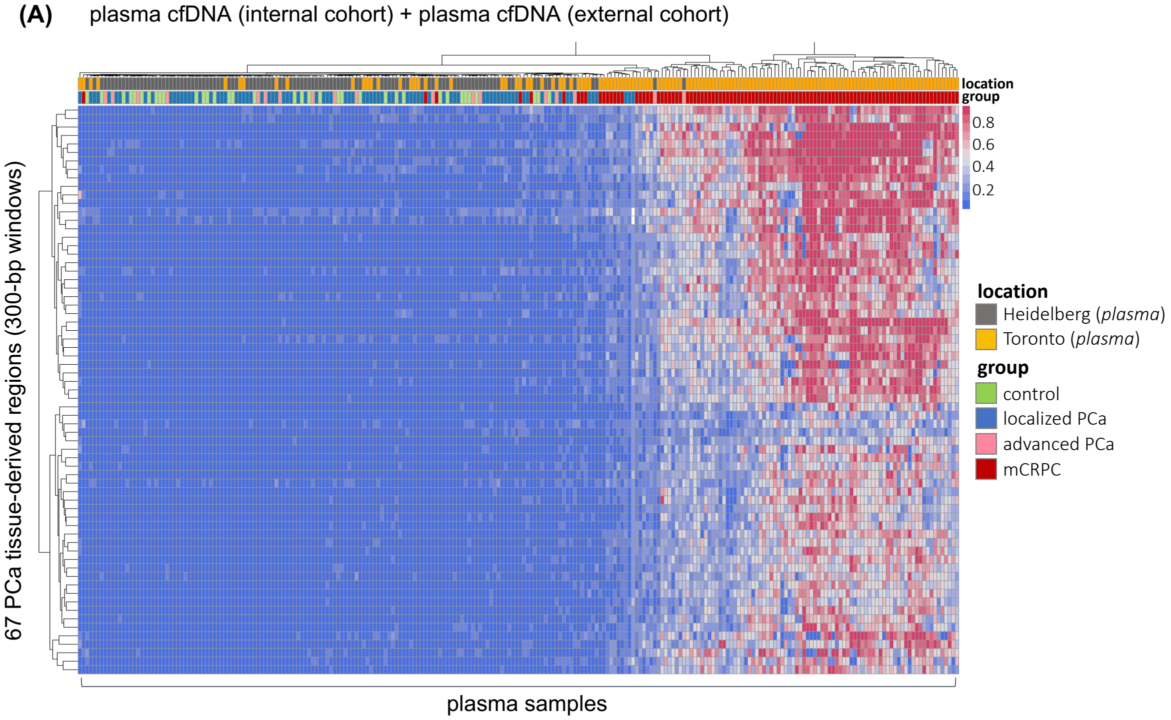


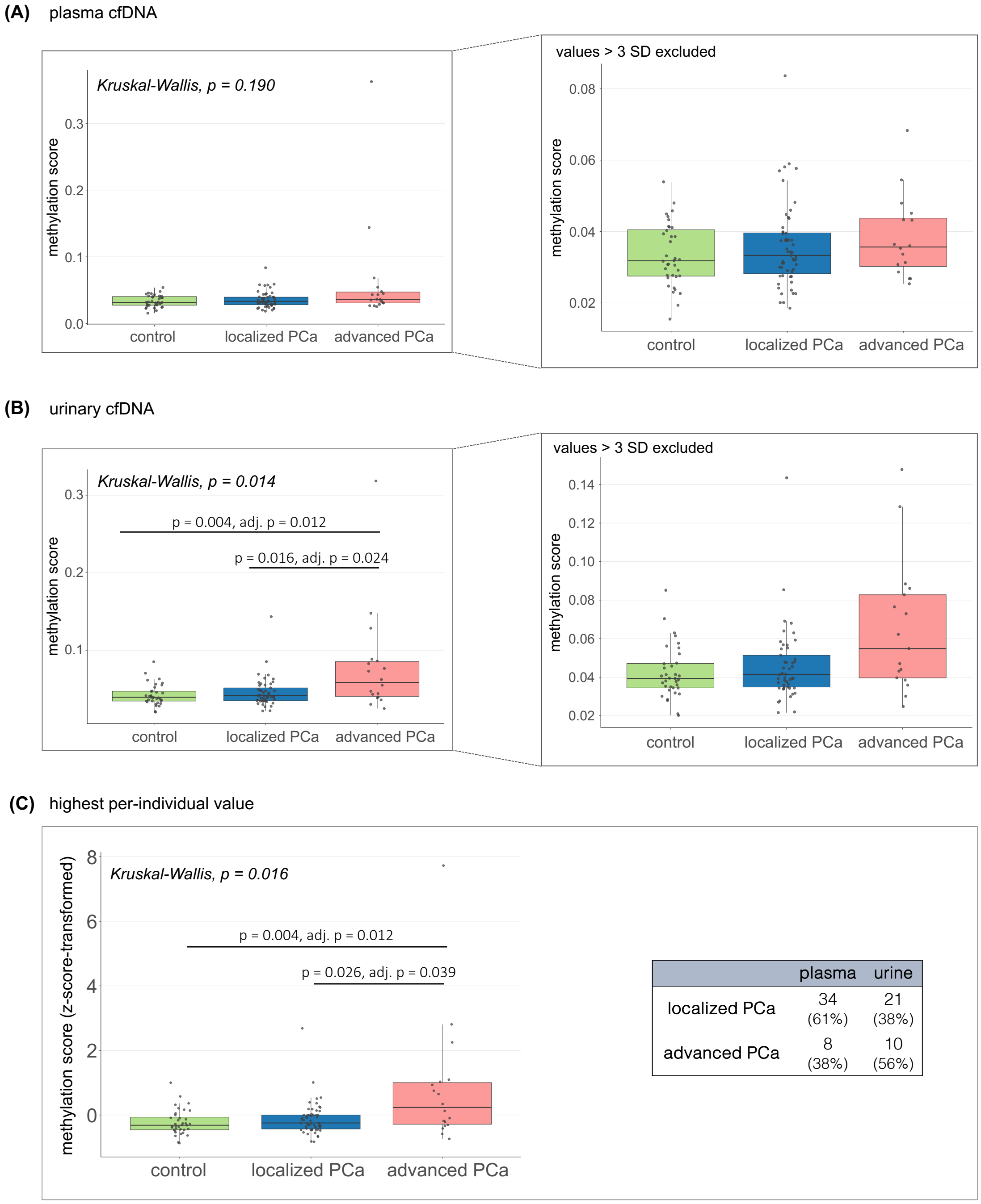
Figure 41: Hierarchical clustering based on 67 PCa tissue-derived methylation marker regions across liquid biopsy samples from the internal cohort and an external PCa cohort.
Heatmap of β-values across 67 PCa tissue-derived methylation marker regions, including plasma samples from patients with localized PCa or mCRPC from an external PCa cohort (Toronto, Chen et al. [354]) together with **(A)** all plasma and **(B)** all urine samples from PCa patients and controls from the internal Heidelberg cohort. Adapted from Riediger *et al.* [410]

3.7.2 PCa tissue-derived methylation markers supported ctDNA detection in plasma and urine

The previous sections described the identification of DMRs through pairwise comparisons between groups, either between PCa patients and controls or between PCa subgroups. However, these analyses did not allow for the assessment of the methylation status at the level of individual plasma or urine samples. To address this question, I calculated a methylation score as a single summary parameter for each sample, reflecting the overall methylation status across the 67 PCa tissue-derived methylation marker regions. The methylation score was defined as the median of β-values across these 67 genomic regions for each sample.

An increasing trend in methylation scores from controls to localized PCa patients and to advanced PCa patients was observed in both plasma and urinary cfDNA (Figure 42; Table A17). Comparative analysis revealed no significant differences in plasma cfDNA (Figure 42A). In contrast, advanced PCa patients showed significantly higher methylation scores than localized PCa patients and controls in urinary cfDNA (Figure 42B). Overall, methylation scores were higher in urine than in plasma samples.

In order to capture the strongest molecular signal per individual, the higher of the z-score-transformed methylation scores from the two biofluids was determined. These highest per-individual values showed a modest increase from controls to localized PCa patients, and a pronounced increase in advanced PCa patients, who exhibited significantly higher scores compared with both localized PCa patients ($p = 0.026$, adjusted $p = 0.039$) and controls ($p = 0.004$, adjusted $p = 0.012$; Figure 42C). Consistent with the genomic analyses, the highest methylation scores in localized PCa patients were more frequently observed in plasma cfDNA (61%, n = 34), whereas in more than a half of advanced PCa patients the highest scores were detected in urinary cfDNA (56%, n = 10; Figure 42C).



	plasma	urine
localized PCa	34 (61%)	21 (38%)
advanced PCa	8 (38%)	10 (56%)

Figure 42: Overview of synoptic methylation scores derived from PCa-tissue derived methylation marker regions in plasma and urinary cfDNA.
(A–B) Distribution of methylation scores, calculated as the median of methylation levels (β-values) in 67 PCa tissue-derived methylation marker regions, across **(A)** plasma and **(B)** urine samples from localized PCa patients, advanced PCa patients, and controls. Additional boxes on the right display zoomed views on the distributions, with methylation scores exceeding 3 SD above the median excluded for visualization. **(C)** Distribution of highest per-individual methylation score values across localized PCa patients, advanced PCa patients, and controls. For each individual, the higher of the z-score-transformed methylation score values from matched plasma and urine samples was used to capture the strongest molecular signal across biofluids. The table adjacent to the plot indicates the proportion of PCa patients in whom the highest signal was detected in plasma or urine, respectively. **(A–C)** Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one individual. The three cohorts were compared using Kruskal-Wallis testing, followed by Dunn's post hoc test. Correction for multiple testing was performed using the Benjamini-Hochberg method. Only significant differences are shown. Abbreviations: SD = standard deviation. Adapted from Riediger *et al.* [410]

I also applied the methylation score to investigate the presence of ctDNA in plasma and urine samples from PCa patients, defining the 95th-percentile of methylation scores from control samples as the ctDNA detection threshold (plasma: 0.046, urine: 0.065).

In plasma cfDNA, 12 tumor samples harbored methylation scores above the defined ctDNA detection threshold, including seven samples from localized PCa patients (5x localized intermediate-risk PCa; 2x localized high-risk PCa) and five samples from advanced PCa patients (1x N1M0, 4x M1; Table A18). Two plasma samples from metastatic PCa patients harbored distinctly higher methylation scores than all other samples. These corresponded to the two previously mentioned outlying samples, which also demonstrated pronounced tumor signals in the genomic analyses.

In urinary cfDNA, 12 tumor samples also exceeded the defined ctDNA detection threshold, including four samples from localized PCa patients (2x localized intermediate-risk PCa; 2 x localized high-risk PCa) and eight samples from advanced PCa patients (4x N1M0, 4x M1). Five patients showed increased methylation scores in matched plasma and urine samples. In all five cases, methylation scores were higher in urine than in the corresponding plasma samples (Table 20; Table A18).

In summary, 24 plasma and urine samples exhibited increased methylation signals across the 67 PCa tissue-derived hypermethylated marker regions, including 10 matched plasma and urine samples from the same patients. Using this approach, tumor signals were detected in 19 patients with localized or advanced PCa. CtDNA detection rates based on methylation scores exceeding the respective detection thresholds for plasma and urinary cfDNA were moderate to low (Table 20), but were overall higher than those obtained from the cfDNA fragmentation and genomic instability analyses. In localized PCa patients, a positive tumor signal was identified in 13% of plasma samples and 8% of urine samples (Table 20). In contrast, detection rates were more than 2-fold higher in plasma samples (28%) and 5.5-fold higher in urine samples from advanced PCa patients (44%; Table 20). Complementary analysis of plasma and urine samples further improved detection rates for both localized PCa patients (16%) and advanced PCa patients (56%).

Table 20: CtDNA detection in plasma and urinary cfDNA based on the methylation score.
Number of localized and advanced PCa patients with detectable ctDNA (methylation score above 95th-percentile of values from controls) in plasma or urine separately, for matched plasma and urine samples (ctDNA detected in both biofluids from the same patient), and for complementary ctDNA detection in either plasma or urine. 68 PCa patients with matched plasma and urine samples; 4 localized PCa patients with missing urine samples.

tumor stage	analyzed liquid biopsy sample	# ctDNA-positive patients
localized PCa	plasma, n = 55	7 (13%)
	urine, n = 50	4 (8%)
	matched plasma + urine, n = 50	2 (4%)
	plasma or urine, n = 55	9 (16%)
advanced PCa	plasma, n = 18	5 (28%)
	urine, n = 18	8 (44%)
	matched plasma + urine, n = 18	3 (17%)
	plasma or urine, n = 18	10 (56%)

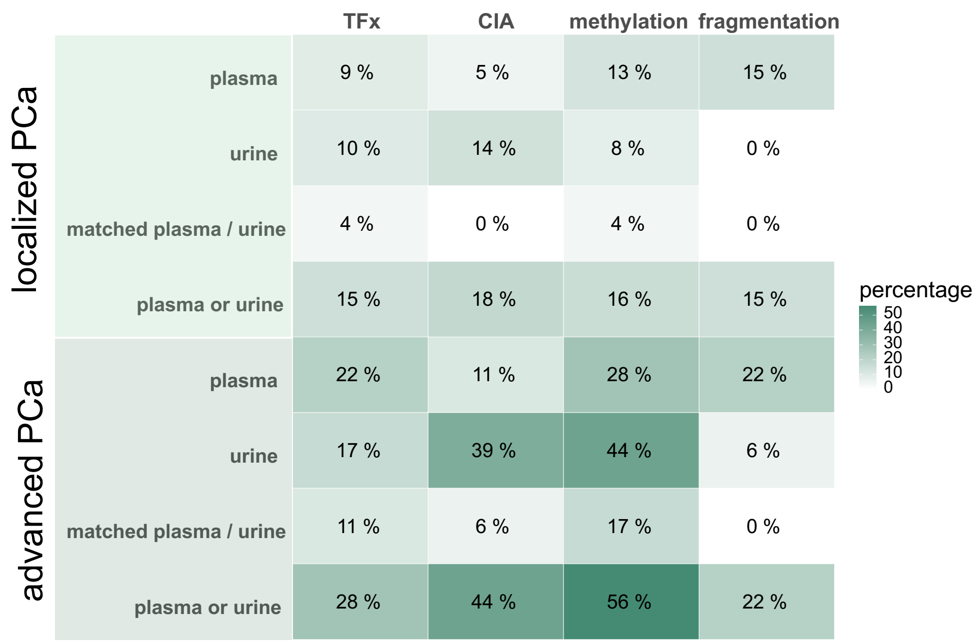
3.8 Integration of genomic and epigenomic analyses in plasma and urine improved ctDNA detection

I performed genomic and epigenomic profiling of plasma and urinary cfDNA, and evaluated the corresponding features for their applicability to detect ctDNA in plasma and urine samples from PCa patients. Chromosomal and genomic instability analyses were represented by the estimated TFx derived from the CNV analysis and by the determination of the CIA score. Epigenomic analyses comprised the evaluation of methylation scores in both liquid biopsy sources, as well as the assessment of two different fragmentation features for plasma (10-bp oscillation score) and urinary cfDNA (P163–160 bp). I selected these two fragmentation features due to their pronounced differences between tumor and control samples, with the 10-bp oscillation score in plasma cfDNA additionally supported by previous reports describing it as a relevant fragmentation feature for tumor characterization[217,279].

The individual analysis types revealed low to moderate ctDNA detection rates in both localized and advanced PCa patients, based on increased values for the genomic and epigenomic plasma and urinary cfDNA features exceeding their respective ctDNA detection thresholds (as detailed in Table 15, Table 17, Table 18 and Table 20, and summarized in Figure 43).

In localized PCa patients, the highest detection rates were observed based on the plasma cfDNA fragmentation feature (15%), followed by the CIA score in urinary cfDNA (14%), and the methylation score in plasma cfDNA (13%). All other plasma and urinary cfDNA parameters revealed ctDNA detection rates of 10% or lower, and no ctDNA-positive samples from localized PCa patients were detected based on the urinary cfDNA fragmentation feature. Complementary analysis of plasma and urine samples increased ctDNA detection rates to 18% for the CIA score, 16% for the methylation score, and 15% for the estimated TFx derived from the CNV analysis.

In advanced PCa patients, ctDNA detection rates were consistently higher than in localized PCa patients. The highest detection rates were achieved using the methylation score (44%) and the CIA score (39%) in urinary cfDNA, followed by the methylation score in plasma cfDNA (28%), and both the estimated TFx and the cfDNA fragmentation feature in plasma cfDNA (22% each). Complementary analysis of plasma and urine samples further increased ctDNA detection rates in advanced PCa patients, reaching 56% for the methylation score, 44% for the CIA score, and 28% for the estimated TFx. Even the lowest ctDNA detection rate observed in advanced PCa patients exceeded the highest ctDNA detection rate achieved in localized PCa patients.



		TFx	CIA	methylation	fragmentation
localized PCa	plasma	9 %	5 %	13 %	15 %
	urine	10 %	14 %	8 %	0 %
	matched plasma / urine	4 %	0 %	4 %	0 %
	plasma or urine	15 %	18 %	16 %	15 %
advanced PCa	plasma	22 %	11 %	28 %	22 %
	urine	17 %	39 %	44 %	6 %
	matched plasma / urine	11 %	6 %	17 %	0 %
	plasma or urine	28 %	44 %	56 %	22 %

Figure 43: CtDNA positivity in plasma and urine from PCa patients based on genomic and epigenomic cfDNA analyses.
CtDNA detection rates, defined as the proportion of ctDNA-positive plasma and urine samples from localized PCa and advanced PCa patients, based on four complementary (epi)genomic cfDNA analyses: 1) estimated TFx based on the CNV analysis with *ichorCNA* (plasma: analysis with *in silico* size selection for 90–150bp fragments, urine: analysis without size selection), 2) CIA score, 3) methylation score, 4) cfDNA fragmentation (10 bp-oscillation score in plasma cfDNA, P163–169 bp in urinary cfDNA). CtDNA detection rates (percentages, %) are shown separately for plasma and urine, for matched plasma and urine samples (ctDNA detected in both biofluids from the same patient), and for the complementary analysis, considering ctDNA detection in either plasma or urine. Matched plasma and urine samples were available for 68 PCa patients, while urine samples were missing for four localized PCa patients. Adapted from Riediger *et al.*[410]

Complementary information was provided by the two liquid biopsy sources, as well as by the four different genomic (TFx, CIA score) and epigenomic (methylation score, cfDNA fragmentation) analysis types.

In localized PCa patients, 11 of 55 plasma samples (20%) showed increased values above the ctDNA detection threshold in one of the four analysis types, and six samples (11%) were positive in two analysis types. None of the plasma samples from localized PCa patients was ctDNA-positive in three or all four analyses (Table A19). Among urine samples from localized PCa patients, seven of 50 samples (14%) exceeded the ctDNA detection threshold in one of the four analysis types, while one sample was ctDNA-positive in three of the four analyses. All remaining urine samples were ctDNA-negative across all analysis types.

In advanced PCa patients, two of 18 plasma samples (each 6%) were ctDNA-positive in three and in all four analysis types, respectively, while four (22%) and two (11%) plasma samples were ctDNA-positive in one or two of the four analyses, respectively (Table A19). Among urine samples from advanced PCa patients, three of 18 samples (17%) exceeded the ctDNA detection threshold in one analysis type, and another three samples (17%) were ctDNA-positive in three of the four analyses. All remaining samples were ctDNA-negative across all analysis types.

When combining all four genomic and epigenomic analyses across both liquid biopsy sources for each patient, 33 of 73 PCa patients (45%) showed at least one positive result in any analysis type (Figure 44). Using this combined approach, ctDNA detection rates of 41.8% were achieved for localized PCa patients and further increased to 55.5% for advanced PCa patients.

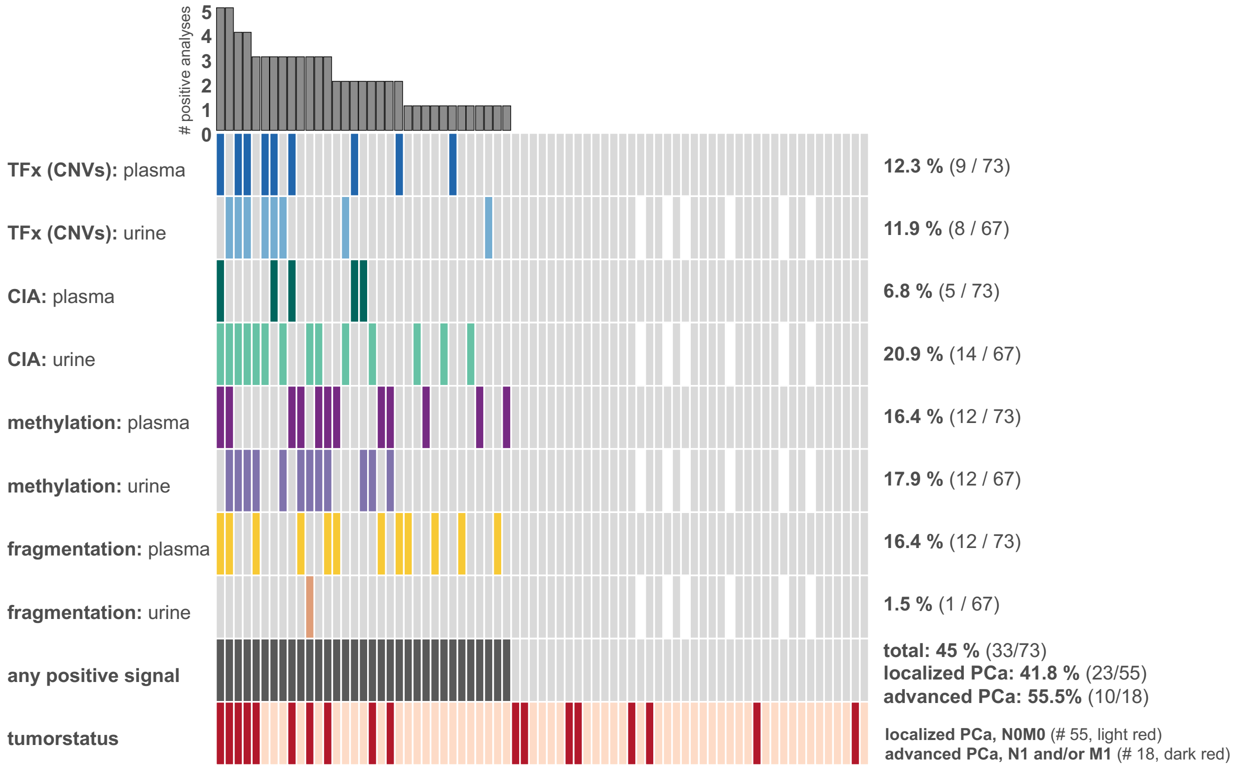


Figure 44: Improved ctDNA detection in localized and advanced PCa patients through multimodal liquid biopsy analyses.
Oncoprint summarizing results from genomic (estimated TFx from the CNV analysis and CIA score) and epigenomic (cfDNA fragmentation features and methylation score) analyses performed on plasma and urinary cfDNA from localized and advanced PCa patients. CtDNA-positive results in the respective analyses are indicated by colored tiles, whereas white tiles display unavailable urine samples (n = 4). The bar plot at the top displays the total number of positive analyses per patient. The second-to-last row indicates the number of PCa patients with at least one positive signal across all analyses ("any positive signal"). Abbreviations: # = number of. Adapted from Riediger *et al.*[410]

3.9 Multimodal models integrating PSA levels and cfDNA features improved PCa risk stratification

The combined genomic and epigenomic cfDNA analyses in plasma and urine identified ctDNA in 23 of 55 localized PCa patients and 10 of 18 advanced PCa patients. The ctDNA-positive cohort of advanced PCa patients included four of nine patients (44%) with lymph node-only disease (N1M0) and six of nine patients (67%) with distant metastases (M1 ± N1; Table A20). Among the advanced PCa patients with only lymph node metastases, two of the four ctDNA-positive patients were lymph node-positive based on both pathological assessment of the lymph nodes after surgery and PET/CT imaging. In contrast, one ctDNA-positive and three of five ctDNA-negative patients were only identified as lymph node-positive following surgery, while the PET/CT imaging showed no suspicious sign of malignant lymph nodes. Among advanced PCa patients with distant metastases, five of the six ctDNA-positive patients harbored both lymph node and distant metastases. Two of three ctDNA-negative patients were lymph node-negative and distant metastases were located in the bones. The ctDNA-positive metastatic PCa patients had a median PSA level of 24.6 ng/ml, whereas the three ctDNA-negative metastatic PCa patients had a lower median PSA level of 7.3 ng/ml. For the lymph node-only patients, the median PSA levels were reversed, and ctDNA-negative patients harbored a higher median PSA level (22.1 ng/ml) compared with ctDNA-positive patients (5.9 ng/ml). With regard to primary tumor characteristics, such as the tumor size (T stage) and the Gleason scoring, 80% of all ctDNA-positive advanced PCa cases (N1M0 or M1) were locally advanced (T3 or T4) and only two metastatic PCa patients harbored locally-restricted (both T2a) tumors (Table A20). Furthermore, the ctDNA-positive patients harbored predominantly tumors with Gleason score >= 8 (Gleason 8, n = 3; Gleason 9, n = 3; Gleason 10, n = 1) and only three cases were scored as Gleason 7a (n = 1) or 7b (n = 2). In contrast, ctDNA-positive localized PCa patients harbored predominantly tumors with Gleason 7b or lower pattern (Gleason 7a, n = 15; Gleason 7b, n = 1; Gleason 6, n = 1), whereas only four tumors were scored as Gleason 8 (n = 1) or Gleason 9 (n = 3; Table A20). The three tumors with Gleason 9 pattern, and three additional tumors were locally-advanced (T3 + T4), whereas the majority of ctDNA-positive localized PCa patients (74%) harbored locally-restricted tumors (T1+T2). However, when stratifying PCa patients based on Gleason scoring (Gleason 6a/7a vs. Gleason >= 7b) or based on T stage (locally-restricted vs. locally-advanced), no significant differences were identified in cfDNA metrics (per-patient highest value for TFx, CIA score, methylation score) and no significant associations were revealed to binary ctDNA positivity.

Comparing the molecular results to the available data from clinical PSA testing, the median PSA level of all ctDNA-positive localized and advanced PCa patients was 8.1 ng/ml, ranging from 2.7–249 ng/ml. Therefore, ctDNA was also detectable in cases with low or intermediate PSA levels <10 ng/ml (Figure 45). The highest PSA levels were observed in ctDNA-positive advanced PCa patients with distant metastases. Stratifying PCa patients by PSA level ranges (below/above 10 ng/ml), ctDNA was present in 43% of localized PCa patients and 67% of advanced PCa patients with low to intermediate PSA levels < 10 ng/ml, as well as in 38% of localized PCa patients and 50% of advanced PCa patients with PSA levels >10 ng/ml (Figure 45). Higher PSA levels were associated with an increasing number of positive genomic and epigenomic analyses, especially in advanced PCa, but PSA levels did not significantly correlate with cfDNA metrics.

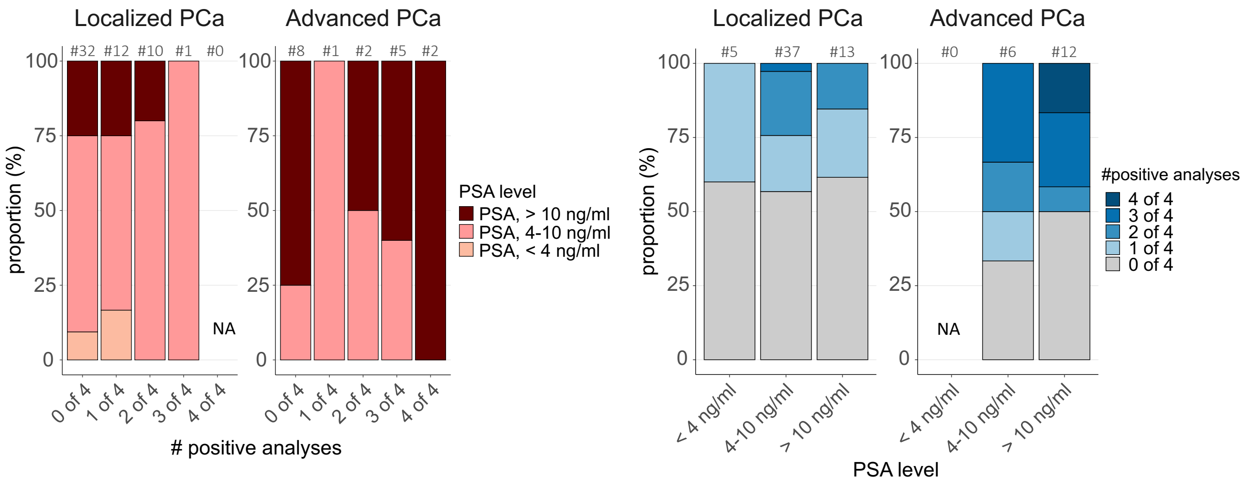


Figure 45: Association between ctDNA positivity and PSA levels in localized PCa and advanced PCa patients. Patients are grouped according to the number of ctDNA-positive analyses (0–4 out of four total analyses) and stratified by serum PSA levels (< 4 ng/mL, 4–10 ng/mL, and > 10 ng/mL). Total patient numbers are indicated above each bar. CtDNA positivity was determined based on four complementary (epi)genomic features assessed in plasma and urinary cfDNA: 1) estimated TFx, 2) CIA score, 3) methylation score, 4) cfDNA fragmentation (10 bp-oscillation score in plasma cfDNA, P163–169 bp in urinary cfDNA). An analysis was considered positive, if ctDNA was detected either in plasma or urine from the respective patient. Adapted from Riediger *et al.*[410]

I constructed GLMs combining clinical parameters with either binary ctDNA status or with quantitative cfDNA features to evaluate whether the integration of cfDNA features improved model discrimination for PCa risk stratification across different potential clinical scenarios during the diagnostic procedure. Two GLMs combining either PSA levels, clinical T stage and Gleason score or only PSA levels and T stage achieved an AUC of 0.96 (95% CI 0.92–1.0) and 0.93 (95% CI 0.87–0.99) for distinguishing localized PCa from advanced PCa, respectively. Both GLMs were not improved by the addition of the binary ctDNA status (ctDNA positivity: yes/no), resulting in an AUC of 0.96 (95% CI 0.92–1.0, DeLong, p = 0.480; Figure A23) and 0.94 (95% CI 0.89–1.00; DeLong, p = 0.303; Figure A23). Among PCa patients with intermediate to low PSA levels < 10 ng/ml, inclusion of ctDNA positivity slightly, but non-significantly improved model performance (AUC = 0.98, 95% CI 0.93–1.0) compared to clinical parameters (PSA level, T stage) alone (AUC = 0.94, 95% CI 0.87–1.0; DeLong, p = 0.305; Figure A23). Next, I simulated an early PCa screening scenario in which PSA testing was used to guide initial risk stratification and inform decision-making for subsequent diagnostic procedures. This assessment was focused on men with low to intermediate PSA levels < 10 ng/ml. A GLM predicting localized PCa versus advanced PCa was established, including PSA alone or PSA levels combined with ctDNA positivity. The model based on PSA alone achieved an AUC of 0.54 (95% CI 0.32–0.76), whereas the inclusion of ctDNA positivity increased the AUC to 0.64 (95% CI 0.34–0.93), indicating a non-significant trend toward improved discrimination (DeLong, p = 0.422; Figure A23).

Based on previous findings indicating distinct differences in genomic and epigenomic cfDNA features across disease stages, I extended the multivariable analysis to include quantitative cfDNA features (highest per-patient values for TFx, CIA score and methylation score, as well as cfDNA fragmentation features (plasma 10-bp oscillation score, urinary P163-169 bp)) rather than the binary information of ctDNA positivity. Although the majority of cfDNA features were elevated in advanced PCa patients compared with localized PCa patients, as well as in PCa patients versus controls, multivariable logistic regression models yielded lower predicted probabilities for the advanced PCa group. The multimodal GLM including PSA levels and cfDNA features significantly increased the AUC to 0.91 (95% CI 0.81–1.00) compared to PSA alone (AUC of 0.57, 95% CI 0.35–0.79, DeLong, p = 0.003; Figure 46).

Threshold-based classification performance improved accordingly, with F1 scores increasing from 0.34 in the PSA-only model to 0.67 in the multimodal model based on PSA levels and cfDNA features. The PR-AUC was higher in the PSA-only model (0.86) compared to the multimodal model (0.70). Bootstrap resampling for the multimodal GLM yielded a mean AUC of 0.96 (95% CI 0.86–1.00) and the likelihood ratio test also confirmed a significant improvement in model fit when cfDNA features were included ($\chi^2 = 13.27$, df = 7, p = 0.021). An additional analysis comparing PCa patients with PSA levels < 10 ng/ml to PSA-elevated controls with PSA levels in the same intermediate range (2–10 ng/ml) revealed a similar pattern: the PSA-only model achieved an AUC of 0.51 (95% CI 0.34–0.68), whereas the multimodal model significantly improved discrimination to an AUC of 0.82 (95% CI 0.70–0.94; DeLong, p = 0.004; Figure 46). Similarly, F1 scores increased from 0.35 in the PSA-only model to 0.88 in the multimodal model, while the PR-AUC was higher in the PSA-only model (0.21) compared with the multimodal model (0.14). Predicted probabilities were lower for PCa patients compared to controls. Bootstrap resampling for the multimodal GLM yielded a mean AUC of 0.852 (95% CI 0.74–0.97) and the likelihood ratio test confirmed a significant improvement in model fit when cfDNA features were included ($\chi^2 = 12.72$, df = 7, p = 0.026).

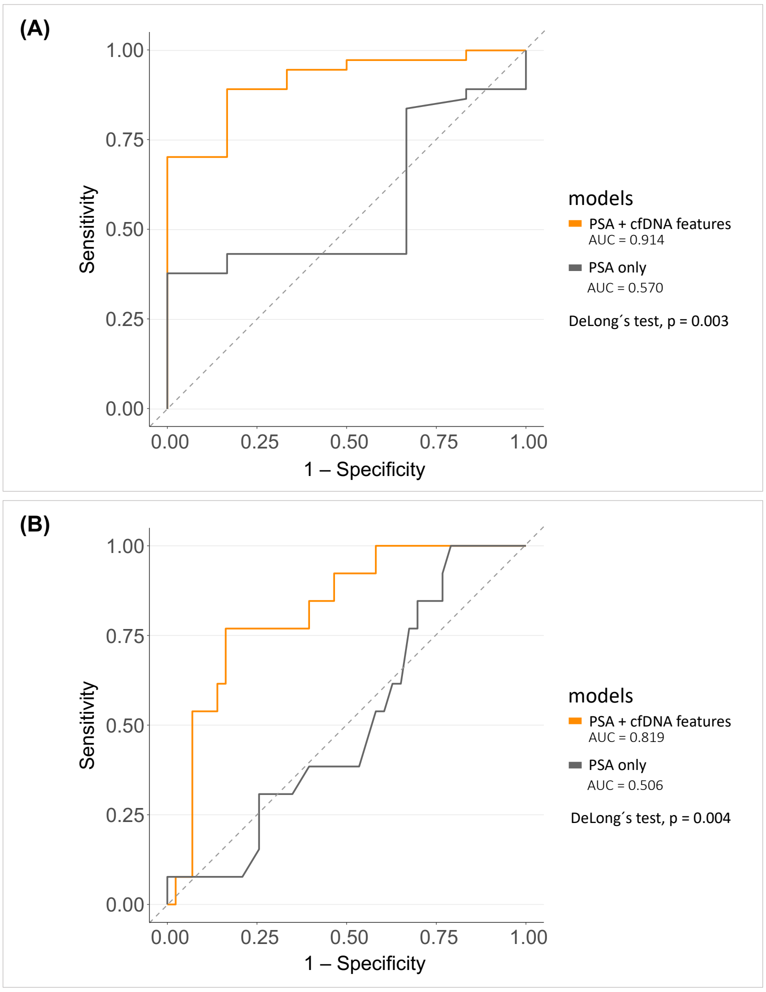


Figure 46: ROC analysis of multimodal GLMs combining PSA levels with cfDNA features.
ROC curves for GLMs based on PSA levels alone (PSA-only model) and PSA levels combined with five genomic and epigenomic cfDNA features (PSA + cfDNA features model). The cfDNA features include the highest per-patient values of TFx, CIA score, and methylation score, as well as two cfDNA fragmentation features derived from plasma and urine (10-bp oscillation score in plasma cfDNA and the proportion of fragments within the 163–169 bp range in urinary cfDNA). **(A)** Models distinguishing localized PCa from advanced PCa among patients with PSA levels < 10 ng/ml. **(B)** Models distinguishing PCa patients with PSA levels < 10 ng/ml from PSA-elevated controls (2–10 ng/ml). **(A–B)** The diagonal reference line indicates random classification performance. Abbreviations: AUC = area under the curve, GLM = generalized linear model, ROC = receiver operating characteristic. Adapted from Riediger *et al.*[410]

4 Discussion

PCa is a heterogeneous disease with regard to its molecular landscape, clinical disease courses and prognosis[117,118,161]. Current diagnostic procedures for clinical decision-making and risk stratification, including PSA testing, clinical staging, and histopathological assessment, incompletely capture tumor heterogeneity and biological aggressiveness[118]. Consequently, there is a clinical need for complementary biomarkers and diagnostic strategies that incorporate molecular characteristics to improve individualized risk stratification and support patient-oriented treatment decisions. Liquid biopsy analyses provide a minimally invasive approach to assess tumor-derived molecular information and the potential to capture spatial and temporal tumor heterogeneity[233-236]**.** However, cfDNA analyses are challenging in PCa, due to limited ctDNA shedding and often low genomic burden, especially in localized disease[250,263,353,412]. Epigenomic changes occur early in tumorigenesis with further accumulation during tumor progression, suggesting their relevance for early tumor characterization[254,288]. Tumor-specific methylation markers have previously been characterized in PCa and successfully profiled in both plasma and urinary cfDNA[183,191].

Based on these considerations, the aim of this PhD project was to establish a multimodal liquid biopsy framework that integrates genomic and epigenomic cfDNA analyses in matched plasma and urine samples from patients with newly diagnosed PCa. The study cohort comprised 73 PCa patients at initial diagnosis with predominantly localized disease and few advanced cases with lymph node- or distant metastases, as well as 36 cancer-free controls. By combining multiple molecular modalities and two liquid biopsy sources, this approach aimed to overcome current limitations of cfDNA analyses in PCa and to support improved ctDNA detection and molecular tumor characterization of heterogeneous PCa, particularly in localized disease and in the context of risk stratification. For this purpose, lcWGS and cfMeDIP-seq were applied on plasma and urinary cfDNA to assess chromosomal instability, CNVs, cfDNA fragmentation patterns, and DNA methylation profiles. Complementary genomic and epigenomic characteristics of plasma and urinary cfDNA were identified, and differences across disease stages were evaluated. Molecular cfDNA profiling results were interpreted in the context of PCa tissue data generated within this study, further supported by validation against an independent external PCa tissue dataset and established molecular characteristics of primary PCa reported in the literature. Based on these analyses, representative genomic and epigenomic cfDNA features were applied for ctDNA detection, and ctDNA positivity results were subsequently assessed in association to clinical characteristics of the PCa patients.

Several key findings emerged from this proof-of-concept study, introducing a novel multimodal approach in newly diagnosed PCa. Genomic and epigenomic cfDNA features showed different patterns in PCa patients and controls, with increasing tumor-associated signals observed in advanced disease. The complementary assessment of multiple genomic and epigenomic cfDNA features in matched plasma and urine samples improved ctDNA detectability compared with the analyses of single parameters and single biofluids, underscoring the value of multimodal strategies. Overall, ctDNA was detected in 45% of newly diagnosed PCa patients, including 42% of localized PCa patients and 56% of advanced PC patients. Of clinical relevance, ctDNA was detectable in a substantial proportion of patients with low to intermediate PSA levels < 10 ng/ml, highlighting the potential of multimodal cfDNA analyses to complement PSA testing and to refine risk stratification prior to prostate biopsy.

4.1 Clinical characteristics of PCa patients and controls

This study combined genomic and epigenomic cfDNA analyses of plasma and urine samples from 73 PCa patients and 36 cancer-free controls. The majority of PCa patients harbored localized disease, reflecting the epidemiological distribution of disease stages at initial diagnosis, as PCa is more frequently diagnosed at localized stage[9]. The control cohort included men with increased PSA levels and/or further PCa risk factors, but negative prostate tissue biopsy, as well as men undergoing examinations or surgery for benign urological conditions. Comprehensive clinical data were available for all participants. PCa patients were significantly older and exhibited higher PSA levels than controls, which was expected given the known association between age, PSA levels, and PCa risk[1,56,57]. PSA levels were further increased in advanced PCa patients compared with localized PCa patients, consistent with the established role of PSA as a prostate-specific biomarker that is widely used in PCa screening, and elevated or dynamically increasing levels have been associated with increased PCa risk[54,413-416]. The potential influence of age on cfDNA analyses has been controversially discussed in the research community[253]. Moss and colleagues reported significantly higher cfDNA concentrations in older individuals, but identified similar distributions of the cfDNA composition between younger or older age, based on methylation deconvolution of healthy plasma cfDNA[197]. Furthermore, it has been shown that higher age in cancer patients compared to controls did not lead to significant differences in the assessment of (urinary) cfDNA fragment distributions[205], suggesting that age alone is unlikely to confound cfDNA-based molecular analyses.

To explore potential PCa-associated risk factors and to obtain additional lifestyle- and health-related information on the study population, questionnaire-based data were collected for the majority of PCa patients and controls. The questionnaire assessed factors which have previously been proposed to impact PCa risk, including diet, alcohol consumption, smoking behavior, and physical activity[23-25], as well as infertility, sterilization, recent urinary tract infections or urinary retention[26-30]. Family history of PCa was equally assessed in the questionnaire, since it was known to significantly increase PCa risk[12,13]. Overall, the questionnaire-based analysis did not reveal significant differences between PCa patients and controls, limiting its interpretability with regard to the identification of potential risk factors within this cohort. Smoking status differed significantly between the groups, but (heavier) smokers were more frequently identified in the control cohort in tumor patients, contrary to expectations. In addition, participants seemed to have difficulties in accurately estimating quantitative lifestyle parameters, such as physical activity, alcohol or meat consumption, and smoking frequency per week, which likely reduced data reliability. Moderate cohort sizes further limited statistical power, particularly for detecting subtle, significant effects. Although several factors have been proposed to impact PCa risk, previous large-scale studies equally reported no significant differences between PCa patients and controls for several lifestyle-related factors, including body weight, smoking status, alcohol consumption, and physical activity. For example, Travis *et al.* analyzed a cohort of 643 PCa patients and 643 matched controls and did not identify significant associations for these factors[417].

4.2 Impact of preanalytical variables and NGS library preparation on cfDNA analyses

The preanalytical workflow for collection and subsequent processing of liquid biopsy samples, as well as the experimental setup prior to NGS were known to be essential to achieve good quality of the sequencing data for the subsequent downstream analyses. Blood and urine samples were collected and further processed applying a standardized workflow based on international recommendations from experts in the liquid biopsy field[418] to avoid biases due to preanalytical factors. Standardization and harmonization of preanalytical aspects has been a highly discussed topic in the research field for many years[216,419,420]. While extensive long-term experiences with blood sample handling have been gained over many years[418], comprehensive recommendations for optimal processing of urine samples remained limited at the time of initiation of this project. I established my workflow for urine supernatant and cell pellet based on the consensus of different previous studies[421-424]. The timepoint and condition of urine sampling has been considered as one crucial aspect. Several publications showed that the timepoint and way of sampling (e.g., early in the morning versus during the day, first-voided urine versus spontaneous urine versus midstream urine) might influence further analyses[421]. Urinary cfDNA concentrations may vary on a daily bases or depending on the time of the day, and cfDNA derived from morning urine tends to show lower cfDNA concentration, but higher levels of viral DNA[421,425,426]. However, these preanalytical aspects have been controversially discussed. One study assessed differences in urinary cfDNA fragmentation with respect to sampling timepoint and the immediate versus delayed addition of EDTA, using different time intervals prior to further processing[282]. Based on these analyses, no remarkable differences were observed, neither in the cfDNA yield, nor in the fragment size distribution. The authors concluded that urinary cfDNA might be affected by different physiological and pathological conditions[282], and a limitation of their study was the lack of reported confounding factors in association to the urine collection, such as time of day, hydration status and time since last void[282]. In the present study, urine was collected between 7 am–12 pm within the clinical routine in the Department of Urology at Heidelberg University Hospital, and men were asked to collect midstream urine. Alternative procedures for urine sampling could not be implemented in the local setup. The timeframe between sample collection and further processing was limited to a maximum of 6 hours, which has been shown to be appropriate for obtaining good cfDNA quality[427] and could be readily implemented in the local workflow. Hydration status and time since last urination were not documented in our project, but it was required that men were sober. One option to account for differences in the glomerular filtration and, thus, potentially improve the analysis of urinary cfDNA, might be the normalization of cfDNA data with renal retention parameters, such as urine creatinine, as already proposed by others[280,428]. On the other hand, the aim of the study was to identify robust tumor biomarkers. This setting could not account for all possible influencing factors, and the identified biomarkers should therefore be capable of reliably distinguishing tumor signals from other physiological or pathological states.

The preanalytical aspects also involved the cfDNA extraction from plasma and urine, for which various different procedures and commercial kits with diverging advantage and disadvantages were available[422,424,429]. Important considerations for cfDNA extraction involved the achievable range of fragment lengths and the yield of extracted cfDNA. Depending on the cfDNA extraction method used, a fragment length bias may be introduced which has been demonstrated to result in a limited yield of shorter fragments[422-424,429]. This aspect was crucial for the extraction of urinary cfDNA, which is known to harbor higher proportions of short fragments[280]. I selected the QIAamp MinElute ccfDNA kit, since this and other cfDNA extraction kits from the manufacturer QIAGEN were broadly used in the field[418,429,430], and it had also been well established in the our laboratory[237,301]. In addition, it enabled cfDNA extraction from volumes of up to 10 ml, as available for the urine samples. QC analyses after cfDNA extraction using this kit demonstrated good performance for both plasma and urine samples. However, fragment length biases, particularly for urine samples, or variability in cfDNA yield could not be fully excluded. A more detailed investigation of these aspects would have required comprehensive comparisons with alternative cfDNA extraction kits, which was considered disproportionately time-consuming and laborious, and was therefore not pursued. However, alternative methods for cfDNA extraction and subsequent NGS library preparation may be considered in future studies, as they could further enhance the analysis of short cfDNA fragments[431].

After sample preprocessing and cfDNA extraction, library preparation for lcWGS and cfMeDIP-seq was performed as an essential experimental step prior NGS. The methylation analysis of plasma and urinary cfDNA represented one of the major objectives of my PhD thesis, and an appropriate solution was identified to integrate the library preparation for both lcWGS and cfMeDIP-seq into a single workflow. I selected the cfMeDIP-seq protocol, published by Shen and colleagues[333], for methylation profiling, because it enabled a cost-effective, genome-wide assessment of methylated regions with a resolution of 100–200 bp based on low DNA input amounts[333]. This protocol had already been established in our laboratory prior to the initiation of my project, and I further optimized it to meet the requirements of the specific plasma and urinary cfDNA analyses performed for this thesis. Adaptions to the published protocol[333] involved the use of a different library preparation kit (KAPA Hyper Prep Kit), which had already been well established in our laboratory, as well as the incorporation of UMIs to improve the detection and removal of PCR and sequencing errors during the bioinformatic processing of the sequencing data. Prior to the methylation enrichment, 20% of the input sample, rather than 10% as described in the original protocol from Shen *et al.* [333], were separated and further processed as non-enriched sample in the same manner as the corresponding methylated-enriched fraction. Library preparation was primarily performed using 7 ng of cfDNA input. Following separation of 20% of the initial input material, at least 1 ng of cfDNA was available for the non-enriched libraries, while the remaining 80% corresponded to approximately 5 ng of cfDNA input for the 5mC-enriched libraries. Both amounts had previously proven sufficient to yield high-quality libraries for the respective analysis types in earlier sequencing projects conducted by our research group[237,238,301,432]. Non-enriched samples served as reference for laboratory QC of the methylation enrichment following library preparation. In addition, they were subjected to lcWGS, enabling background correction of the corresponding cfMeDIP-seq data during the bioinformatic analysis, as well as serving as input data for CNV and fragmentation analysis. Particular emphasis was placed on temperature-sensitive steps within the cfMeDIP-seq protocol, which required careful execution to ensure appropriate efficiency of the methylation enrichment. Several protocol optimizations, such as execution of pipetting steps in a cold room (< 4°C), contributed to the successful implementation of the methylation enrichment. The methylation QC based on the qPCR quantification of 5mC- and 5C-spike-ins confirmed the appropriate enrichment of methylated DNA in cfMeDIP-seq libraries. The recovery of 5mC- and 5C-spike-ins, as well as the specificity of the methylation enrichment were calculated based on the qPCR data. I considered quality thresholds proposed by Shen and colleagues[333] (5mC-recovery > 20%, 5C-recovery < 1%, specificity ≥ 99%) as reference, but not as fixed threshold or binding exclusion criterion. This decision was based on the observation that the calculated 5mC- and 5C-recoveries were influenced by the efficiency of the qPCR quantification. Accordingly, PCR efficiencies were continuously monitored by quantification of serial dilutions of 5mC- and 5C-spike-ins to ensure consistent results.

In addition to the laboratory QC, the efficacy of the methylation enrichment was assessed using QC metrics derived from the cfMeDIP-seq data. In this context, a relative methylation enrichment score was calculated for each sample. Shen and colleagues proposed a target threshold for the relative methylation enrichment score greater than 3. However, the authors of the cfMeDIP-seq protocol did not define the rationale for this proposed threshold, although reported values in their data ranged between 3 and 4[312,345]. In the present study, this threshold was not achievable for the majority of cfMeDIP-seq samples, despite the optimization of crucial steps within the protocol and the confirmation of an appropriate methylation enrichment based on the laboratory methylation QC. Preliminary own analyses indicated that the insert size length (i.e., cfDNA fragment length) might influenced the calculation of the methylation enrichment score, with shorter fragments tending to yield higher scores. This effect might partially explain deviations from the proposed threshold and variability in the methylation enrichment scores among different DNA sources, given the distinct differences between plasma and urinary cfDNA fragmentation profiles, as well as differences compared with sheared gDNA derived from PCa tissue and buffy coat samples. Consistent with these observations, several sequencing-based methylation QC metrics differed significantly between libraries derived from sheared gDNA and cfDNA. These differences had to be interpreted with caution, as the number of PCa tissue and buffy coat samples (n = 8 each) was substantially smaller than the number of plasma and urine samples (plasma, n = 109; urine, n = 107), likely affecting the statistical testing. Furthermore, gDNA samples were mechanically sheared prior to library preparation, whereas cfDNA is intrinsically fragmented and characterized by shorter fragment length distributions. These differences in fragment length were expected to affect sequencing-based QC metrics, including the saturation analyses, which explicitly depended on the abundance of short reads. This likely explained the observed differences between gDNA- and cfDNA-based libraries, as well as between matched plasma and urine samples, as urinary cfDNA exhibited a higher proportion of short fragments. Matched lcWGS data served as an essential reference for the interpretation of the methylation QC results. The methylation enrichment scores obtained from lcWGS data were distinctly decreased compared to the corresponding cfMeDIP-seq data, indicating that methylated DNA fragments were effectively enriched in the cfMeDIP-seq libraries. This was further supported by the CpG coverage metrics, which showed a sufficient proportion of covered CpGs and a low proportion of reads without CpG pattern in (cf)MeDIP-seq libraries, while matched lcWGS libraries showed markedly higher proportions of reads without CpG pattern and lower proportions of CpGs covered by more than five reads. These consistent differences underscored the specificity and efficacy of the methylation enrichment despite variability in individual QC metrics. Taken together, these results supported an overall successful methylation enrichment, and no samples were excluded based on methylation QC criteria. However, a reduced efficacy of the methylation enrichment could not be entirely ruled out and may represent a potential explanation for comparatively low signals observed in subsequent methylation analyses.

4.3 Plasma and urinary cfDNA fragmentation characteristics in PCa patients and controls

I assessed plasma and urinary cfDNA fragmentation based on the overall distribution of insert sizes derived from paired-end lcWGS and cfMeDIP-seq data. This approach enabled the identification of shared patterns and source-specific and/or tumor-specific differences in global fragmentation profiles and allowed the extraction of quantitative features from plasma and urinary cfDNA fragmentation distributions based on lcWGS data for comparative analysis between tumor and control samples.

The fragmentation analysis of plasma and urinary cfDNA revealed distinct characteristics in fragment length distributions between the two liquid biopsy sources, as well as differences between PCa patients and cancer-free controls. The characteristics of plasma and urinary cfDNA fragmentation have been previously described in different tumor entities[282,328,433], and result from differences in underlying biological processes, such as cfDNA release mechanisms and nucleosomal DNA packaging[272]. Plasma cfDNA fragment profiles harbor a characteristic peak at 167 bp, representing the DNA wrapped around the nucleosome plus linker DNA[246]. In some cases, a second peak at ~334 bp or a third peak at ~501 bp, displaying multiples of the first peak, were also observed in the present study. CfDNA fragments shorter than 150 bp exhibited an oscillation pattern with up to 11 local maxima and minima at an approximately 10-bp periodicity. In contrast, urine samples showed an elongated peak extending to approximately 400 bp, rather than a single prominent peak, along with a high proportion of short cfDNA fragments (< 100–150 bp). The 10 bp-oscillation pattern was also observed in short urinary cfDNA fragments, extending to fragment lengths of approximately 300 bp, and was more pronounced than in plasma cfDNA. Small cfDNA fragments shorter than 200 bp length are known to originate from apoptotic cells, whereas larger fragments are associated with other types of cell death e.g., necrosis[245,249,273]. The composition of urinary cfDNA is more heterogeneous, and its origin remains less well understood than those of plasma cfDNA. Glomerular filtration of plasma cfDNA with selective passage of short fragments (< 100–250 bp) [272,278,279] has been discussed as one potential mechanism. Subsequently, cfDNA fragments are exposed to enzymatical and mechanical degradation during their passage through the urinary tract[280]. In contrast, longer urinary cfDNA fragments (> 1 kb) may originate from urinary tract cells[277]. In addition, ctDNA derived from urological tumor cells, such as PCa cells, may be directly released into the urine, bypassing renal filtration and with reduced exposure to degradation. Consequently, urinary cfDNA presents as a heterogeneous mixture of transrenal and directly released, non-tumorous cfDNA and ctDNA.

The plasma and urinary cfDNA fragmentation analysis included the global assessment of cfDNA fragment length distributions and the evaluation of specific fragmentation features, i.e., proportions of defined fragment length ranges or pairwise ratios between these proportions. The analyses were inspired by previously published studies, demonstrating the utility of cfDNA fragmentation patterns for tumor characterization[213,275,288].

Initial differences between plasma samples from PCa patients and cancer-free controls were observed during the assessment of the global cfDNA fragmentation profiles. Comparative analysis using Kolmogorov-Smirnov testing revealed significant differences between the median cumulative frequency distributions of tumor and control samples. These significant differences were observed when assessing the entire fragment length range of 30–700 bp, as well as for shorter fragments (30–250 bp) and longer fragments (250–700 bp), respectively. Interestingly, the significant differences did not remain when investigating the subranges of fragment lengths associated with the first cfDNA peak (30–250 bp) or the second peak (250–700 bp), respectively. In these subrange analyses, frequencies of fragment lengths were calculated using the fragment length ranges associated with the first or second peak as the reference, rather than using the entire fragment length range (30–700 bp size). A comprehensive assessment of this phenomenon revealed that the frequencies of fragment length ranges associated with the two peaks were not independent, but rather mutually influenced each other. The proportion of longer fragments (250–700 bp) relative to all fragments (30–700 bp) was higher in controls compared to tumor patients. Conversely, the proportion of small fragments (30–250 bp) relative to all fragments was lower in controls compared to tumor patients. When assessing the first or the second peak as a subrange, tumor samples harbored a distinctly higher first peak and a slightly higher second peak. The analysis of the entire fragment length range (30–700 bp), combined with separate analyses of both subranges, revealed that the control samples showed a more evenly distributed pattern across the entire range, apart from the two (or three) visible peaks, and exhibited increasing frequencies when the subranges were analyzed, however without exceeding those observed in the tumor samples. The tumor samples, especially in the case of advanced PCa, had higher concentrations around both peaks, and these tendencies were more pronounced when the subranges were analyzed. This effect was stronger for the second peak, which was notably higher in advanced PCa samples than in controls in the subrange analysis, whereas controls showed a higher second peak when frequencies were calculated relative to the entire fragment length range. The results from Kolmogorov-Smirnov testing, based on the cumulative frequency distributions of tumor and control samples, reflected these findings. When comparing the cumulative frequencies across the entire fragment length range (30–700 bp), the cumulative distributions captured the spread and the combined effect of both peaks, revealing significant differences due to distinct overall distribution patterns in tumor samples versus control samples. Isolating the subranges reduced these effects, and the distributions became more similar, resulting in non-significant differences. This comprehensive assessment was performed to identify the most appropriate fragment length range to serve as a reference for plasma cfDNA fragmentation analysis. It was therefore concluded that the overall distribution patterns across the entire plasma cfDNA fragment length range were more informative for identifying differences between tumor and control samples compared to the isolated subranges. In other studies, either the full range of fragment lengths spanning the mono-nucleosomal and di-nucleosomal peak (or longer fragments) was considered for the analyses[323,434], or no concrete reference fragment length range was reported[205,213,274]. A few publications explicitly mentioned that they considered only fragment length ranges below 250–300 bp in their fragmentation analyses[435,436], since the characteristic mono-nucleosomal peak in plasma was known to contain the majority of tumor-derived DNA[275,323,437]. However, ctDNA has also been detected in the di-nucleosomal peak at longer fragment sizes. Ganesamoorthy and colleagues demonstrated that shared variants in plasma cfDNA and tumor tissue, representing tumor-derived fragments, were enriched at fragment lengths associated with the di-nucleosomal peak (300–350 bp), using variant analysis of matched plasma cfDNA, germline DNA, and tumor tissue-derived DNA samples from four breast cancer patients and two patients with benign tumors[434]. Markus and colleagues performed a comprehensive assessment of ctDNA enrichment in fragment size bins of 10 bp, ranging from 0 to 400 bp, and identified the highest ctDNA enrichment and an increased extent of CNVs in fragments with lengths of 126–135 bp and 240–324 bp, respectively[438]. In addition, fragment length profiles of five patients with high-grade serous ovarian cancer showed an increased peak height and a shift toward shorter fragment sizes for the di-nucleosomal peak compared to healthy controls[438]. It has been shown that ctDNA fragments were slightly shifted toward shorter sizes compared to non-tumorous cfDNA[205,274,276,288,323], provided that either the average fragment lengths or short fragment length ranges (< 250–300 bp) were considered. The differences in fragment lengths between cancer patients and controls have also been observed in PCa[267,439], although studies evaluating cfDNA fragmentation patterns in PCa remain limited[281]. Souza *et al.* reported that plasma cfDNA fragment sizes were longer in healthy controls (range: 174–7,387 bp) compared to PCa patients with T1/T2 prostate tumors (range: 157–4,912 bp), and were further decreased in patients with T3/T4 tumors (range: 29–53 bp)[267]. Chen *et al.* evaluated total cfDNA concentrations and fragment sizes in plasma samples from a cohort of 112 localized PCa patients, 122 mCRPC patients, and 34 healthy controls using the 2100 Bioanalyzer gel electrophoresis system, a similar technology to the Fragment Analyzer instrument that I used for QC during the library preparation[439]. They compared plasma cfDNA fragmentation between localized PCa patients and controls, and reported that localized PCa patients harbored significantly shorter cfDNA fragments (median average size: 173 bp, range: 135–280 bp) than healthy controls (median average size: 177.5 bp, range: 142–265 bp). I also observed a similar tendency toward shorter fragments in plasma samples from tumor patients compared to controls in the present study. The average fragment length was lower in PCa patients (198 bp) compared to controls (205 bp), and advanced PCa patients harbored the shortest average fragment length (196bp). However, statistical testing revealed that these differences were not statistically significant between the two cohorts.

Mouliere and colleagues performed comprehensive analyses of various plasma cfDNA fragmentation features across patients with multiple cancer entities and cancer-free controls, and demonstrated that the proportion of the fragment length range of 20–150 bp (P20–150 bp) had the highest AUC of all features for tumor characterization[274]. In addition, the proportions of the fragment length ranges of 160–180 bp, 180–220 bp and 250–320 bp, as well as the oscillation score obtained from the 30–150 bp fragment length range, when combined with results from the CNV analysis, represented the strongest features in a logistic regression model[274]. Mouliere and colleagues also assessed three pairwise ratios of proportions of fragment length ranges (P20–150 bp/P160–180 bp, P100–150 bp/P163–169 bp, P20–150 bp/P180–220 bp), but they did not include them in further classification analyses, because of strong correlations observed between these ratios and other fragmentation features[274]. In the present study, I observed distinct differences in plasma cfDNA for the two fragment length ranges of 30–60 bp and 150–300 bp between all PCa patients and controls, but these differences did not remain significant after correction for multiple testing. The proportion of fragments with a 30–60 bp length showed a decreasing trend from controls to localized PCa patients, and further to advanced PCa patients. In another study by Mouliere and colleagues, this fragment length range was not reported to be significantly lower in plasma cfDNA from controls compared to tumor patients, but a reversed significant difference was identified in urinary cfDNA (tumor patients > controls) [205]. I did not identify such significant differences in the urinary cfDNA of the present cohort, but instead observed a higher proportion of fragments within the 30–60 bp range in urinary cfDNA from controls compared to PCa patients. As already described, plasma cfDNA fragments were more concentrated at the first peak (~167 bp) in PCa patients compared to controls. Accordingly, PCa patients harbored higher proportions of fragment length ranges associated with this peak (163–169 bp or 160–180 bp) in plasma cfDNA, though differences were not statistically significant. The ratios of short fragments (30–100 bp or 30–150 bp) compared to these two fragment length ranges highlighted distinct differences between tumor patients and controls, primarily driven by the difference between advanced PCa patients and controls, and the ratio P30–100 bp/P163–169 bp remained statistically significant after correction for multiple testing. These findings were consistent with literature data, since the characteristic first peak in plasma cfDNA is known to contain the majority of tumor-derived DNA[275,323,437].

Within the plasma cfDNA fragmentation analysis, I also assessed the 10-bp oscillation pattern, reflecting local maxima and minima within the 30–150 bp fragment length range, and calculated a composite oscillation score for each sample based on the differences in the amplitudes of these local maxima and minima. Mouliere and colleagues demonstrated in two studies that the oscillation score differed significantly between tumor and control samples and was applicable for tumor characterization[213,274]. It is assumed that this 10-bp oscillation pattern originates from the varying accessibility of DNA wrapped around the nucleosome and, thus, from periodic enzymatic cleavage of DNA[246,433]. Consistent with previous findings, I also observed a significant difference in plasma cfDNA oscillation scores between tumor and control samples in the present cohort, with PCa patients showing a reduced amplitude of the 10-bp oscillation compared to controls. Mouliere and colleagues observed the same trend in cerebrospinal fluid of glioma patients with detectable CNVs compared to glioma patients without detectable CNVs[213]. Based on the significant results observed in my data and the previously described findings in the literature, I selected the 10-bp oscillation score in the fragment length range 30–150 bp as informative plasma cfDNA fragmentation feature for tumor characterization and for further comparison with other genomic and epigenomic cfDNA analyses. Based on the assessment of the 10-bp oscillation score, ctDNA was detected in 15% of localized PCa patients and 22% of advanced PCa patients. Thereby, this feature revealed the highest ctDNA detection rate in plasma from localized PCa among all genomic or epigenomic features. In advanced PCa patients, the ctDNA detection rate was comparable to that based on the estimated TFx from the CNV analysis, exceeded that based on the CIA score, and was slightly lower than the detection rate achieved with the methylation score.

In addition to the observed 10-bp oscillation pattern in 30–150 bp fragments, my data revealed a consecutive oscillation pattern within the 150–300 bp fragment length range, but this pattern was more pronounced in urinary cfDNA, with up to 15 additional oscillations and an average 10-bp periodicity. In plasma cfDNA, only four oscillations were observed on average, with a variable range of local maxima and minima per sample and a highly variable periodicity. Chandrananda and colleagues similarly described three additional peaks at 151, 173 and 177 bp with decreased periodicity (~5 bp) in these longer plasma cfDNA fragments[433]. However, a comprehensive depiction of the oscillation pattern in urinary cfDNA within the 150–300 bp fragment length range has, to my knowledge, not been presented before.

Urinary cfDNA fragmentation profiles from both PCa patients and controls revealed distinctly higher intersample variability compared to plasma cfDNA, and common characteristics or clear differences between the groups were difficult to define. Urinary cfDNA fragmentation profiles were compared between PCa patients and controls using Kolmogorov-Smirnov testing based on the cumulative frequency distributions of fragment lengths. However, comparative analysis revealed no significant differences in any of the assessed fragment length ranges. This contrasts with results presented by Mouliere and colleagues, who reported significant differences in the median cumulative frequency distributions of urinary cfDNA fragment lengths between high-grade glioma, patients with other CNS pathologies, and healthy controls [205]. It has to be considered that ctDNA derived from urine samples of glioma patients could originate exclusively from a transrenal source, with potentially lower heterogeneity of urinary cfDNA fragmentation compared to urological tumor entities, for which an additional direct release of ctDNA into urine may occur. Urinary cfDNA profiles in the present cohort harbored an elongated peak spanning a broad fragment length range from 30 bp to approximately 400 bp. For this reason, the separate assessment of subranges associated with individual peaks, similar to the plasma cfDNA analysis, was not considered reasonable. However, group-wise assessment of urinary cfDNA fragmentation profiles revealed that several advanced PCa cases harbored an additional peak at 160–170 bp, similar to their matched plasma cfDNA profiles. These fragments could originate from cfDNA or ctDNA directly released into urine and could potentially be considered a tumor-derived signal. The described characteristic was predominantly observed in individual urinary cfDNA fragmentation profiles from (advanced) PCa patients, but was also present in four control samples. These four cancer-free controls did not show any shared clinical characteristics or laboratory abnormalities in blood and urine analyses which could explain the origin of the prominent peak. In addition, technical causes related to library preparation or NGS could be excluded, since the four samples were not prepared in the same batch and were partially not processed together in the same sequencing run. Two of the four men showed an increased PSA level and underwent prostate tissue biopsy, revealing no evidence of malignancy. In both cases, benign prostate hyperplasia was diagnosed, including areas with chronic and acute inflammation. These pathological processes may have caused cfDNA shedding from prostate cells. However, additional cancer-free controls with similar histopathological features were identified and did not demonstrate the prominent peak in their urinary cfDNA samples. Of note, the urine samples from both men were also outliers in the CNV analysis, which was likewise performed using the lcWGS data. One of them additionally harbored the plasma sample with the highest peak at ~167 bp in the plasma cfDNA fragmentation analysis of all controls, while the plasma sample of the other men showed the second-highest methylation score among all controls. Interestingly, this man was diagnosed with a localized intermediate-risk PCa one year later, and it can be discussed whether early molecular signs were already present in the current sample. However, no liquid biopsy or PCa tissue samples from the time point of PCa diagnosis were available for molecular analyses and comparison with existing data from cfDNA profiling. The other two men had an appointment at the urological department for the resection of a testicular hydrocele and a renal cyst (Bosniak II/F), respectively, but did not undergo PCa screening (PSA levels < 2 ng/ml). The renal cyst could potentially be associated with a direct shedding of cfDNA into urine. However, further degradation of the cfDNA during passage through the urinary tract would have been expected to more likely result in shorter fragments (< 100 bp) rather than in a prominent peak at ~167 bp fragment length. Similarly, the presence of a testicular hydrocele could not explain the occurrence of the prominent peak in urinary cfDNA. None of the four men harbored any other pathological features, such as stones or urinary statis, which could induce mechanical irritation and thus local cfDNA shedding from urothelial cells. Prostate MRI data from the affected PCa patients and controls were evaluated to further investigate the hypothesis that the prominent peak at 160–180 bp fragment length in urinary cfDNA originated from cfDNA, or even ctDNA, directly released into urine. In this context, it was assessed whether any suspicious lesions or other abnormalities were visible in close proximity to the urethra, the seminal vesicles, or the bladder, potentially enabling direct access to urine. The majority of PCa patients and cancer-free controls either harbored no suspicious findings close to the urethra or bladder, or had no annotated imaging data available that specifically addressed the presented scientific question. However, a few representative examples supported the formulated hypothesis. Bladder invasion by the prostate tumor was observed in one advanced PCa patient, while two other PCa patients had prostate tumors with extracapsular extension, and an additional patient harbored a tumor with seminal vesicle invasion.

The urine sample from the advanced PCa patient harboring a bladder invading prostate tumor also exhibited an increased CIA score, while both the estimated TFx derived from the CNV analysis and the methylation score were not increased. However, the corresponding plasma sample from the same patient demonstrated distinct genomic alterations in the CNV analysis and was ctDNA-positive in all four (epi)genomic analyses. In two additional advanced PCa patients, prostate tumors also showed bladder invasion, and their urine samples likewise harbored a peak at 160–180 bp, although less prominent than in the previously mentioned cases. In these two cases, ctDNA was also detectable in urine based on the assessment of both the CIA score and the methylation score, and one of them also demonstrated apparent genomic alterations in the CNV analysis. Overall, these PCa patients represented suitable examples supporting the assumption that close proximity of prostate tumor cells to the urinary tract may result in a direct release of ctDNA into urine, manifesting as a prominent tumor signal in urinary cfDNA. However, considering the different observed scenarios, a final conclusion in support of this hypothesis could not be drawn and will require further investigation. In this context, comprehensive tissue sequencing data from well-annotated lesions by prostate MRI would be highly beneficial for direct comparison with the findings obtained from the liquid biopsy-based analyses.

Analogous to the distinct peak observed in a subset of urinary cfDNA samples, several urinary cfDNA fragmentation features related to these mono-nucleosomal fragment lengths revealed distinct differences between (advanced) PCa patients and cancer-free controls, including the proportion of fragments within the 163–169 bp range, which was increased in tumor samples. Consequently, I selected the proportion of fragments with a length of 163–169 bp (P163–169 bp) as tumor-informative urinary cfDNA fragmentation feature for tumor characterization and for further comparison with other genomic and epigenomic cfDNA features. In the assessment of the feature P163–169 bp for ctDNA detection in tumor samples, only one advanced PCa patient with lymph-node metastases harbored an increased value above the defined detection threshold based on the 95th-percentile of values from controls. Two control samples demonstrated a prominent peak at mono-nucleosomal fragment lengths and, consequently, revealed the highest P163–169 bp values of all urine samples. This aspect was previously discussed in this section, and the reason for the outlying values could not be explained. Eleven PCa samples harbored elevated P163–169 bp values that were lower than these outlying measurements, but higher than all remaining control samples, comprising six localized PCa patients and five advanced PCa patients. The increased proportion of urinary cfDNA fragments with a length of 163–169 bp in tumor samples has previously been described in a few studies on cfDNA fragmentation profiling of urine samples, as well as in cfDNA derived from seminal fluid of PCa patients[270,440,441]. Cheng and colleagues assessed urinary cfDNA in bladder cancer, which is likewise a urological tumor entity with direct contact between tumor cells and urine[270]. In this study, a distinct increase in fragments with an approximate length of 150–180 bp was exclusively observed in urinary cfDNA fragmentation profiles from muscle-invasive bladder cancer patients compared to those from non-muscle-invasive cancer patients or cancer-free controls[270]. The authors reported a higher proportion of urinary cfDNA fragments with lengths > 105 bp in tumor patients with advanced, muscle-invasive bladder cancer, but not in non-muscle-invasive bladder cancer patients, compared to cancer-free controls. Further analyses suggested that these fragments (> 105 bp) contained tumor-derived cfDNA. The authors concluded that non-tumor cfDNA, originating from either glomerular filtration of cfDNA in the blood or from healthy urinary tract cells, was degraded during the passage through the urinary tract and was represented in the short cfDNA fragments, whereas direct shedding of ctDNA from bladder tumor cells into urine without further degradation contributed to longer fragments[270]. Another study by Russo and colleagues assessed urinary cfDNA fragmentation from urothelial bladder cancer patients using the Bioanalyzer gel electrophoresis system[440]. They reported a first peak at approximately 178 bp, and a second or even third peak, as well as longer DNA (1–10 kb) in most of the urine samples, with variable peak heights between samples. The authors concluded that ctDNA was present in both shorter and longer fragments, since the ratio between 178 bp and longer fragment lengths was not associated with the mutant allele frequency (MAF) of the *telomerase reverse transcriptase (TERT)* 228G>A mutation in urinary cfDNA, determined by digital droplet PCR[440]. Additionally, the MAF was not significantly different when comparing fractions enriched for shorter or longer fragments, respectively[440]. Mouliere and colleagues performed cfDNA fragmentation analyses in plasma, urine and cerebrospinal fluid from glioma patients, patients with other diseases of the central nervous system (CNS), and cancer-free controls[205]. The presented cfDNA fragmentation profiles of urine samples from all cohorts harbored a high proportion of short fragments with a distinct oscillation pattern, but no obvious peak at 160–180 bp. A discrete peak was observed in a few individual cfDNA fragmentation profiles from controls, patients with other CNS diseases and patients with high-grade glioma, as well as in the median fragmentation profile of all controls, but was shifted to fragments lengths of approximately 140–150 bp. The authors assessed the proportion of different fragments lengths ranges in 30-bp steps between 30 bp and 300 bp based on lcWGS data from urinary cfDNA. They did not detect a significant difference between tumor and non-tumorous samples for the proportion of 151–180 bp fragments, but did observe distinct differences for shorter fragment lengths (< 90 bp). The proportion of the fragment length range 30–60 bp was significantly increased in low-grade and high-grade glioma compared to controls or patients with other CNS diseases, and this feature also revealed the highest classification performance in a binary classification, distinguishing glioma patients from controls (AUC = 0.885). In contrast to my data, the authors reported significant differences in the median cumulative frequency distributions of urinary cfDNA fragmentation between high-grade glioma and controls (Kolmogorov-Smirnov test, D statistic = 0.476, $p < 0.001$) or patients with other CNS diseases (Kolmogorov-Smirnov test, D statistic = 0.287, $P < 0.001$), respectively[205]. Furthermore, they reported that the median fragment length decreased from controls (137 bp) to patients with other CNS diseases (108 bp), and further to glioma patients (101 bp), with a significantly lower median fragment length in glioma patients compared to the other two groups. In a separate analysis based on hybrid-capture panel sequencing of plasma and urinary cfDNA derived from eight glioma patients, they assessed fragment lengths in sequencing reads with and without identified mutations and concluded that tumor-derived fragments were shorter than non-mutant fragments in both plasma and urine (160 bp vs. 169 bp in plasma cfDNA, and 101 bp vs. 133 bp in urinary cfDNA, respectively[205]). However, the authors also remarked that the reported median urinary fragment sizes differed from those reported in other studies previously described in this section[270,282]. Cheng and colleagues reported higher proportions of longer fragments (> 105 bp) in urinary cfDNA from bladder cancer patients compared to controls[270]. Markus and colleagues identified the modal size of urinary cfDNA to be 80–81 bp in the majority of 30 cancer-free controls, or 111–112 bp in a few additional control samples. Similarly, a cohort of 22 cancer patients, including non-metastatic pediatric solid cancers and pancreatic cancers at different stages, revealed a modal urinary cfDNA fragment length of 80–81 bp or 111–112 bp, and in one case 123 bp[282]. The data from these two studies were consistent with my findings in the present study, showing that the modal size was located at 111 bp for both PCa patients and controls, but increased to 148 bp when only advanced PCa patients were considered. However, the median fragment length was higher in controls (169 bp) compared to localized and advanced PCa patients (167 bp and 166 bp, respectively), in line with the results published by Mouliere and colleagues[205].

Despite the distinct characteristics observed in urinary cfDNA fragmentation between tumor and control samples in my analysis, none of the urinary cfDNA fragmentation features revealed significant differences between the two groups. The comparison of the results reported in different studies and my data in the present study highlighted the variability of the findings on urinary cfDNA fragmentation, which was also observed for plasma cfDNA, albeit to a lesser extent. The urinary cfDNA analyses might have been influenced by physiological and pathological factors specific to the individual patient or cancer-free control (e.g., hydration, pH, urea concentration or comorbidities including chronic kidney failure, diabetes, urinary tract infection), as well as by pre-analytical aspects concerning sample collection and processing, or the use of cfDNA extraction methods with varying selectivity for different fragment lengths[422-424,429]. All of these aspects could affect the composition and, consequently, the fragmentation of urinary cfDNA, as previously discussed by others[205,270,282]. These circumstances might have also contributed to the observed variability in my cohort of PCa patients and cancer-free controls. Some controls harbored increased CRP levels as sign of inflammation, potentially altering the cfDNA composition toward increased proportions of cfDNA originating from blood cells[197,311,442]. Furthermore, a subset of the control cohort harbored benign urological conditions, leading to changes in the urothelial tract and, thus, potentially influencing the composition of the urinary cfDNA. For example, ureteric stones might cause mechanical irritation and subsequent shedding of healthy cfDNA from epithelial cells[280,443]. Similarly, infectious or inflammatory processes might influence urinary cfDNA composition[444,445] and increase non-specific background signals in the analyses, thereby impacting the identification of differences relative to PCa patients. The majority of PCa patients harbored localized disease and/or small tumor sizes, and only low amounts of ctDNA were expected in addition to the high non-tumorous background. Low ctDNA abundance and high background noise likely affected the analyses of both plasma and urinary cfDNA, and may explain the absence of expected significant differences between tumor and control samples or the discrepancies with results reported in the literature.

Finally, I performed cfDNA fragmentation analysis based on cfMeDIP-seq data and identified distinct differences in the fragmentation patterns compared to matched lcWGS data in both plasma and urinary cfDNA. The main differences concerned a relatively higher proportion of longer fragments (~300–400 bp), leading to a higher second peak and a reduced first peak in plasma cfDNA, and resulting in a lower proportion of smaller fragments and a shift toward longer fragment lengths in urinary cfDNA compared with the corresponding lcWGS data. The presence of a prominent second peak in cfMeDIP-seq data, compared to the fragmentation profile of extracted, non-enriched plasma cfDNA, was also observed in the data reported by Shen and colleagues[333], but the authors did not provide further details on this observation. The altered cfDNA fragmentation profile of cfMeDIP-seq data compared to lcWGS data might be due to the immunoprecipitation procedure, which may favor specific fragments lengths. In addition, it may also have an underlying biological explanation, such as methylated fragments potentially exhibiting distinct size distributions. Differences in cfDNA fragmentation have previously been shown to be associated with CpG methylation[446,447], where CpG methylation appeared to influence fragment end positions and was associated with higher cfDNA abundance compared to regions with unmethylated CpGs. Additionally, genome-wide hypomethylation has been linked to overall shorter cfDNA fragment lengths in cancer patients[447]. Consequently, it would have been of interest in this cohort to investigate differences between the lengths of methylated cfDNA fragments compared to non-enriched fragments derived from lcWGS data, and to assess whether these characteristics harbored potential for discriminating PCa patients and controls. However, no final conclusion could be drawn based on my data, due to the impact of the antibody-based enrichment, which likely involved both biological effects and technical biases. For this reason, cfDNA fragmentation analyses based on cfMeDIP-seq data were not further extended.

4.4 Complementary genomic information from CNV and chromosomal instability profiling

I performed genomic profiling to investigate genome-wide CNVs and chromosomal instability in plasma and urinary cfDNA from PCa patients and controls. CNVs were assessed in 1000-kb genomic windows using lcWGS data and the *ichorCNA* algorithm, which enabled the inference of ctDNA presence and estimation of the corresponding TFx based on genome-wide copy number gains and losses. To enable comparison across data types, I complemented CNV profiles derived from lcWGS by CNV assessment based on cfMeDIP-seq data. In addition, I performed CNV profiling on primary PCa tissue samples using lcWGS data to identify PCa-associated genomic alterations and to facilitate comparison with recurrent CNVs detected in liquid biopsy samples.

Beyond CNV profiling, I evaluated chromosomal instability as a complementary genomic feature in plasma and urinary cfDNA. This analysis enabled calculation of a synoptic CIA score, providing an additional genome-wide measure of genomic instability alongside TFx estimates. Together, these genomic features allowed a more comprehensive assessment of plasma and urinary cfDNA. In this context, the complementary analysis of genomic instability enabled the specific determination of ctDNA within the total cfDNA and could hereby reflect the heterogeneous tumor burden[254].

4.4.1 Identification of genome-wide CNVs in liquid biopsies and PCa tissue

The CNV analysis revealed challenges for ctDNA detection in plasma and urinary cfDNA from both localized and advanced PCa patients. Only two advanced PCa patients with distant metastases showed prominent alterations and a high estimated TFx in plasma cfDNA, whereas all other plasma samples harbored only a few CNVs and low TFx values, or showed no tumor-derived signals at all. Three additional ctDNA-positive PCa patients with low TFx values were identified based on the CNV analysis without size selection. However, they turned ctDNA-negative after *in silico* size selection for 90–150 bp fragments, suggesting either a false-positive result in the CNV analysis without size selection or a loss of tumor-informative reads through size selection and subsequent downsampling of sequencing reads. It has previously been shown that *in silico* size selection for 90–150 bp fragments could improve the sensitivity of the CNV analysis based on lcWGS data from plasma cfDNA[204,274]. In my cohort, the same approach also increased the number of ctDNA-positive tumor samples, and identified seven additional ctDNA-positive plasma samples compared to the CNV analysis without size selection, except for the three samples that became ctDNA-negative after size selection. The two plasma samples with distinct CNV profiles and elevated TFx values in the CNV analysis performed without size selection showed further increases in the TFx values following size selection, similar to the median TFx of all PCa patients, which also increased after size selection. The median TFx of controls equally increased following size selection, which might indicate that the size selection procedure amplified background signals in the data. However, the increase in the median TFx was more pronounced in PCa patients compared to controls, supporting the expected enrichment of ctDNA in 90–150 bp fragments. The median TFx of tumor samples also had to be interpreted with caution, since it was below the ctDNA detection threshold, and therefore noise and inaccuracy may have contributed to the observed findings. The proportion of cfDNA fragments within the 90–150 bp range was not significantly increased in plasma samples from tumor patients compared to controls, consistent with findings reported in another PCa study [439].

The authors who investigated an improved ctDNA detection in tumor samples through *in silico* size selection in plasma cfDNA hypothesized that size selection might also improve the sensitivity of ctDNA detection in urinary cfDNA[205]. However, a comprehensive assessment of appropriate fragment length ranges for *in silico* size selection of urinary cfDNA fragments has, to my knowledge, not been reported before. Consequently, I first performed CNV analyses based on lcWGS data from urinary cfDNA without size selection and subsequently repeated these analyses following iterative *in silico* size selections across 15 different fragment length ranges. For each CNV analysis following fragment length-specific *in silico* size selection, several aspects were evaluated to assess a potential enrichment of ctDNA in tumor samples and a corresponding improvement over the CNV analysis without size selection. These evaluations were based on changes in median TFx values in PCa patients and controls, as well as changes in the number of ctDNA-positive tumor samples. The size selection for an optimal fragment length range was expected to significantly increase median TFx values in PCa patients without substantially affecting TFx values in controls or, at least, to result in a higher TFx increase in tumor samples than in controls, thereby increasing the number of ctDNA-positive tumor samples. Across the evaluated fragment length ranges, *in silico* size selection led to changes in estimated TFx values in both tumor and control samples. However, these effects were heterogeneous and dependent on the selected fragment length range. While the size selection for certain fragment length ranges led to significant differences in TFx values in tumor samples, these changes were not consistent across analyses and were frequently accompanied by similar or even stronger TFx increases in control samples. Furthermore, size selection did not consistently increase TFx values in tumor samples relative to the CNV analysis without size selection. Taken together, the comprehensive evaluation of *in silico* size selections for different fragment length ranges prior to CNV analyses did not identify a fragment length range that reliably improved CNV-based ctDNA detection in urinary cfDNA of PCa patients. One reason might be the low number of urine samples with detectable ctDNA based on both CNV analysis without size selection and CNV analyses following size selections for different fragment length ranges, limiting the interpretability with regard to the ctDNA enrichment. Many samples harbored TFx values slightly below the ctDNA detection threshold, a range in which high background noise was expected and systematic statistical testing could not be robustly applied. Another possible explanation involved the heterogeneity of urinary cfDNA composition, which had already been discussed in detail in the context of cfDNA fragmentation. The presence of ctDNA was either expected in shorter or longer urinary cfDNA fragments, respectively, or even in both, depending on the origin and the underlying release mechanism of the ctDNA in each sample. In addition, physiological or pathological aspects that may influence the background noise in the CNV analysis had to be considered. To conclude, no single fragment length range for *in silico* size selection prior to CNV analysis that was applicable to all urine samples could be identified. In future studies, sample-specific or group-specific fragment length ranges could be selected based on common characteristics, as determined by prior cfDNA fragmentation analysis. However, further research on the composition of urinary cfDNA in urological tumor entities, as well as on additional influencing factors, is needed before such approaches can be applied.

Despite these limitations, the CNV analysis in urinary cfDNA, even without size selection, showed a higher ctDNA detection rate compared to the CNV analysis in plasma cfDNA without size selection, and showed a similar performance compared to the plasma CNV analysis with size selection. Eight ctDNA-positive PCa patients were identified based on the CNV analysis in urinary cfDNA, including five patients with localized intermediate-risk PCa, two patients with lymph node metastases and one patient with distant metastases. Seven of these eight urine samples harbored TFx values greater than 10% and demonstrated prominent chromosomal alterations in the corresponding CNV profiles. Furthermore, two patients with localized PCa and two patients with advanced PCa also harbored ctDNA-positive plasma samples. However, detectable alterations in matched plasma and urine samples were concordant only in one advanced PCa patient with lymph node metastases, indicating complementary results between plasma and urinary cfDNA in all other ctDNA-positive patients. This could be explained by the different origins of plasma and urinary cfDNA in relation to the location of the prostate tumor, potentially favoring ctDNA shedding into one or the other liquid biopsy source. Consequently, the complementary analysis of plasma and urinary cfDNA could enhance ctDNA detection and uncover tumor heterogeneity. However, the sensitivity of the CNV algorithm and the presence of background noise within the analysis could also play a pivotal role, potentially resulting in both false positive and false-negative signals.

The challenges associated with ctDNA detection in liquid biopsies from newly diagnosed PCa patients were consistent with findings reported in previous studies. CNVs have been detected in plasma and urinary cfDNA across different tumor entities, with improved detection in advanced disease[247,250,270,289]. However, genomic cfDNA analyses in PCa remain challenging. Heitzer and colleagues presented data from genome-wide profiling of plasma cfDNA in metastatic, hormone-sensitive PCa (CSPC) and CRPC using a lcWGS approach[305], which is comparable to the genomic analyses presented in this study. They detected multiple CNVs, including known alterations from PCa tissue analysis, such as *TMPRSS2-ERG* deletion on chromosome 2, and *AR* amplification in CRPC, but not in CSPC[305]. Kohli and colleagues reported low ctDNA fractions in 15% of their study cohort, which comprised metastatic PCa patients expected to harbor higher ctDNA abundance, representing a limiting factor for the sensitive detection of key alterations[307]. PCa is known to belong to the tumor entities with rarely detectable ctDNA[250]. In localized PCa, it is even more challenging to characterize recurrent genomic features, due to the very low ctDNA abundance, and the fact that well-characterized alterations often arise during tumor progression. Further approaches were attempted to address the challenge of ctDNA detection in localized PCa disease[263,310]. Hennigan *et al.* performed ultra-low-pass WGS for CNV assessment, as well as highly sensitive, personalized targeted panel sequencing for somatic mutation detection on plasma cfDNA of PCa patients[263]. In the use of both techniques, ctDNA could only be detected in metastatic PCa, but not in plasma samples from localized PCa patients, although high-risk patients with subsequent disease recurrence were included in the cohort. Lau *et al.* applied a sensitive, personalized analysis to plasma samples of eight patients with localized, intermediate-risk or high-risk PCa, using targeted panel sequencing informed by individual WGS results derived from PCa tissue, but detected SNVs or structural variants in only two of the eight patients[310]. The application of tagged amplicon deep sequencing in a second cohort of patients with localized PCa (n=189) also identified *TP53* alterations in ctDNA in only 12% of cases[310].

Previous studies have demonstrated that CNVs can be inferred from methylation-enriched sequencing data derived from cfMeDIP-seq or MBD-seq, respectively, based on plasma cfDNA from PCa patients[354,355]. In the present study, CNV analysis using the *ichorCNA* algorithm revealed inconsistent results between cfMeDIP-seq and lcWGS data in both plasma and urine, with only moderate correlations observed between estimated TFx values derived from the two sequencing approaches (Spearman's $\rho = 0.32$ in plasma and 0.62 in urine), in contrast to the strong concordance reported by Lleshi *et al.* (Spearman's $\rho = 0.98$) [355]. Given these inconsistencies, CNV results and subsequent analyses were restricted to lcWGS data.

Several limitations and difficulties were encountered when performing CNV analyses on lcWGS data using the bioinformatic algorithm (R package) *ichorCNA*[286], affecting both the sensitivity and specificity of the applied approach. CNVs were detectable in only a few plasma and urine samples, which could be caused by both biological reasons, such as the predominantly early tumor stage of the majority of PCa patients with associated low ctDNA abundance[263], as well as by technical limitations, including the limited sensitivity of the applied algorithm. Various parameter settings, compared to the default settings, were tested to improve the sensitivity of the *ichorCNA* algorithm, but certain settings resulted in heterogeneous results, thereby limiting the interpretation of the findings. Additionally, misleading annotation and color coding of gains and losses in CNV profiles were encountered, which were related to the underlying segmentation of the genome for CNV determination. These aspects would have required a manual review of all estimated TFx values and corresponding CNV profiles for each sample, as proposed by the authors of the *ichorCNA* algorithm[286]. This approach was not pursued, since it would have introduced a subjective component into the analysis and would also not be feasible in future studies with large sample sizes.

An additional important consideration with respect to the sensitivity and specificity of the algorithm concerned the definition of an appropriate ctDNA detection threshold. The authors of the *ichorCNA* algorithm reported a ctDNA detection limit of 0.03 based on their lcWGS data (0.1 x coverage) derived from plasma cfDNA[286]. I defined an individual ctDNA detection threshold for the present study based on the 95th-percentile of the TFx values obtained from plasma or urinary cfDNA in the control cohort, consistent with all other genomic and epigenomic analyses in the present study to balance appropriate specificity and sensitivity of the algorithm. This threshold was selected to account for differences in my own laboratory and bioinformatic workflow compared to the original implementation studies of the *ichorCNA* algorithm. In addition, this approach was chosen to ensure comparability between CNV analyses of plasma or urinary cfDNA, as well as between CNV analyses performed with and without *in silico* size selection, since no predefined detection limits were available for these different approaches. The defined ctDNA detection thresholds in plasma cfDNA for both CNV analyses performed with and without size selection were distinctly lower than the threshold of 0.03 defined by Adalsteinsson *et al.* [286]. Applying a ctDNA threshold of 0.03 to plasma cfDNA in this study would have resulted in the loss of three ctDNA-positive tumor samples in both CNV analyses with and without size selection. The threshold for the CNV analysis with size selection was 1.25-fold higher (0.023) than that of the CNV analysis without size selection (0.018). This difference may be explained by additional signal noise introduced through the reduction in the number of available sequencing reads following *in silico* size selection.

Contrastingly, the ctDNA threshold (0.041) in urinary cfDNA was higher than 0.03, originally defined based on plasma cfDNA. Twelve additional urine samples would have been classified as ctDNA-positive in the CNV analysis without size selection based on a lowered threshold of 0.03. Overall, urine samples appeared to harbor higher estimated TFx values than plasma samples, which could be have both biological reasons due to increased ctDNA release into urine, and technical reasons due to a higher background noise in the CNV analysis. For this reason, the default threshold of 0.03 would not have been an appropriate choice for the CNV analysis of urinary cfDNA.

The CNV analysis of both plasma and urinary cfDNA revealed outlying control samples, which impacted the maximum TFx value as well as the definition of the ctDNA detection threshold. Outlying control samples with the highest TFx values in both CNV analyses corresponded to samples previously described as outliers in the cfDNA fragmentation analysis. The plasma control sample with the highest TFx in the CNV analysis with size selection for 90–150 bp fragments showed a deviating fragmentation profile, as both the first and second cfDNA peaks were shifted toward longer fragment sizes, potentially explaining the outlying result observed in the CNV analysis with size selection for shorter fragment lengths. For urinary cfDNA, the control samples with the highest TFx values harbored a visible peak at approximately 160–180 bp in their fragmentation profiles and have previously been discussed in detail in this context. One of these controls also harbored a corresponding plasma sample with the highest first peak of all controls in the fragmentation profile, but did not show detectable alterations in the CNV analysis. The second cancer-free control was diagnosed with PCa one year after sample collection. The corresponding plasma and urine samples showed both the second-highest TFx values among all control samples in the CNV analyses without size selection, however, they did not reveal concordant alterations in their CNV profiles. Following size selection in plasma cfDNA, the TFx value was decreased, although ctDNA enrichment would have been expected through size selection, if the detectable alterations in the plasma control sample had represented an early sign of malignancy. Consequently, false-positive results in the CNV analysis of control samples could not be excluded, which may have been associated with the same as yet unknown factors underlying the deviating findings observed in the cfDNA fragmentation analysis. An additional influential finding was the systematic reduction of log ratios in chromosome 19 in the CNV profiles from both tumor and control samples, which was interpreted by the *ichorCNA* algorithm as a deletion of this chromosome and consequently influenced the estimation of the TFx. This phenomenon was also observed in the two plasma control samples with highest TFx values in the CNV analyses performed without and with size selection, respectively. It had previously been reported by the authors of the *ichorCNA* algorithm that chromosome 19 showed a systematic decrease in their assessment of tumor and control samples and therefore likely influences the CNV estimation[286]. I decided not to exclude chromosome 19 from the CNV analysis, since the decrease in chromosome 19 was also present in control samples, which served as a reference for tumor samples and defined the ctDNA detection threshold.

Despite the limitations of the applied algorithm, it could ultimately be concluded that the complementary CNV analyses of plasma and urinary cfDNA enhanced ctDNA detection in both advanced and localized PCa patients. While the individual analyses of plasma or urine samples identified ctDNA in 9%–10% of localized PCa patients, as well as in 17–22% of advanced PCa patients, the combined analysis increased the detection rates to 19% in localized PCa and 39% in advanced PCa patients. Only a minority of these PCa patients harbored detectable ctDNA in both matched plasma and urine, but benefited from the complementary analysis of both liquid biopsy sources. When comparing the results from the CNV analyses with the established clinical biomarker PSA, 9 of 13 PCa patients with detectable ctDNA harbored a PSA level in the intermediate range of 4–10 ng/ml[63], highlighting the potential of cfDNA analyses as an additive diagnostic tool. All PCa patients with a ctDNA-positive urine sample showed a PSA level below 10 ng/ml, except for one PCa patient with distant metastases and a PSA level of 11 ng/ml.

I further performed CNV analysis in eight fresh-frozen PCa tissue samples, which allowed a direct comparison with the findings obtained from liquid biopsy samples. Recurrent genomic alterations detected in plasma and urinary cfDNA were concordant with the results from the tissue-based CNV analyses, as well as with well-established genomic events in primary PCa in the literature. The CNV analysis in PCa tissue samples revealed three distinct categories of samples with many, few or no detectable genomic alterations. Four PCa tissue samples showed numerous CNVs, all of which consisted exclusively of deletions. These samples uniformly exhibited a Gleason 7a pattern, but varying tumor sizes (2x pT2c, 1x pT3a, 1x pT3b). One of these cases also showed lymph node involvement (pN1 with 1/21 positive lymph nodes). One additional PCa tissue sample harbored both amplified and deleted regions and represented an aggressive tumor phenotype, characterized by a

Gleason 9 pattern, advanced local tumor stage (pT3b), and lymph node involvement (pN1 with 2/22 positive lymph nodes). In contrast, three additional PCa tissue samples did not show any detectable CNVs, despite sufficient tumor cell content (40%, 50% and 90%, respectively) and advanced local tumor stage (2x pT3a and 1x pT3b). These findings were concordant with previous studies reporting the identification of characteristic clusters based on the analysis of somatic CNVs in primary PCa[161,162]. The Cancer Genome Atlas Research Network identified three distinct groups in the molecular analysis of 333 primary prostate cancers, including one group with high genomic instability characterized by extensive CNVs, a second group lacking detectable CNVs, and a major third group with intermediate CNV levels, comprising approximately one half of all analyzed PCa tissue samples[161]. Taylor and colleagues similarly reported the heterogeneous occurrence of CNVs in primary PCa[162]. They identified six distinct clusters, of which four were characterized by low CNV burden, ranging from few alterations, predominantly involving deletions, to nearly unaltered genomes, and two additional clusters with high CNV burden, characterized by genome-wide amplifications and deletions, or by predominantly gains in chromosome 7 or 8q[162]. These two clusters included both localized tumors and the majority of metastatic PCa cases. With regard to the pathological Gleason score, PCa with Gleason scores 6 and 7 were present across all clusters, whereas Gleason 8 tumors were present in five of six clusters, and Gleason 9 tumors were only present in two clusters (cluster 5 and the adjacent cluster 4), which were characterized by genome-wide amplifications and deletions or a reduced but still elevated CNV burden, respectively[162].

The eight PCa tissue samples from my cohort could be appropriately assigned to the three previously reported groups, comprising samples with unaltered genome, few deletions, or genome-wide amplifications and deletions, the latter corresponding to the PCa tissue sample with a Gleason 9 pattern, respectively. The genomic alterations observed in the five PCa tissue samples with detectable CNVs were concordant with the most frequently reported alterations in primary PCa. Large-scale genomic rearrangements, as well as chromosomal arm-level alterations, including gains on chromosomes 7 and 8q, and losses on chromosomes 8p, 10q, 13q, 16q, and 18, are known to be common in primary PCa[161,448]. Alterations on chromosome 8 have been reported in 30–50% of cases for 8p loss, and in 20–40% for 8q gain[162]. In addition, one of the most well-known alterations associated with PCa is the *TMPRSS2-ERG* fusion, which is detectable in approximately 40–50% of primary PCa cases[162]. In my CNV analysis of PCa tissue samples, deletions were identified on chromosomes 8p, 13q and 18, as well as on 5p, 6p and 12p in at least 40–50% of the five PCa tissue samples with detectable CNVs. Amplifications in the single PCa tissue sample harboring gains were predominantly observed on chromosomes 7 and 8q, and additionally on chromosomes 3, 5, 11, 12 and 13. However, the genomic PCa tissue analysis based on lcWGS data was not suitable for the detection of the *TMPRSS2-ERG* fusion due to the low sequencing coverage.

The most frequently altered regions were also observed in plasma or urinary cfDNA from localized and advanced PCa patients. Deletions on chromosomes 5q, 6q, 8p, 13q, and 18 occurred in more than 40% of both PCa tissue and liquid biopsy samples. Deletions on chromosome 16q were detected in more than 50% of liquid biopsy samples, but only in 20% of PCa tissue samples. Additional amplified regions on chromosomes 3q, 5p, 7, and 8q were also detected in at least 40% of liquid biopsy samples and were shared with the single PCa tissue sample harboring amplifications. However, the corresponding liquid biopsy samples from the eight PCa patients with available PCa tissue samples did not harbor detectable ctDNA, limiting the direct comparison of detectable CNVs between PCa tissue and liquid biopsy sources. With regard to specific genes of interest targeted by altered regions in PCa tissue and liquid biopsy samples, *phosphodiesterase 4D* (*PDE4D;* 5q11.2-q12.1), *PIK3R1* (5q13.1) and *repulsive guidance molecule BMP co-receptor b* (*RGMB*; 5q15) were identified in both PCa tissue and liquid biopsy samples, whereas *docking protein 5* (*DOK5*; 20q13.2) was detected exclusively in liquid biopsies, and other well-established PCa-associated genes (e.g., *PTEN*, *MYC*, *RB1, AR, TP53*) [161,162,286] were not involved in the detected CNV regions.

4.4.2 Genome-wide assessment of chromosomal instability in plasma and urinary cfDNA

I applied an additional genomic analysis on plasma and urinary cfDNA to assess genome-wide chromosomal instability and to define a single summary metric, the CIA score, for each sample, analogous to the estimated TFx determined with the *ichorCNA* algorithm based on detectable CNVs. With this approach, I aimed to obtain an alternative measure of genomic instability, potentially overcoming the previously described limitations of the CNV analysis using the *ichorCNA* algorithm, which were primarily related to the genome segmentation for CNV estimation and high background noise. The calculation of the CIA score was likewise based on mappability- and GC-content-corrected log ratios in genome-wide 1000-kb bins from tumor and control samples, but applied a different algorithm that did not rely on a subsequent segmentation. Instead, genome-wide bins with particularly low or high log ratios, and thus showing marked deviations from the remaining bins, were considered. Regions with only minor deviations from the copy-number-neutral baseline were not included in the calculation, thereby potentially reducing the non-informative background noise. The CIA score was calculated for plasma and urinary cfDNA from both PCa patients and controls, and the detection limit was again defined based on the 95th- percentile of CIA scores derived from the control cohort.

Five ctDNA-positive plasma samples from three localized PCa patients and two advanced PCa patients were identified based on an elevated CIA score, resulting in the same ctDNA detection rate as observed for the CNV analysis without size selection, which was also performed based on the same mappability- and GC-content-corrected log ratios in genome-wide 1-Mb windows. The calculation of the CIA score based on size-selected sequencing data (90–150 bp fragment) did not improve the analysis, but instead resulted in only four ctDNA-positive tumor samples. The two plasma samples with the highest TFx values in the CNV analyses both with and without size selection also harbored the highest CIA scores of all plasma samples, highlighting a high ctDNA burden in these two cases. The additional three plasma samples from localized PCa patients with increased CIA score did not overlap with the additional three ctDNA-positive plasma samples identified by the CNV analysis without size selection. Contrastingly, the three plasma samples with elevated CIA score harbored an estimated TFx of 0.0 in the CNV analysis without size selection. However, two of them became ctDNA-positive after size selection, with TFx values increasing to 0.036 and 0.059, respectively. This observation might suggest both an increased sensitivity of the CIA approach and ctDNA enrichment through size selection prior to the CNV analysis. However, it remained unclear why *in silico* size selection did not consistently lead to an improved CIA score determination.

In urinary cfDNA, 14 samples from seven localized PCa patients and seven advanced PCa patients harbored an increased CIA score above the ctDNA detection threshold, exceeding the detection rate observed in the CNV analysis in urinary cfDNA and potentially indicating increased sensitivity. Six of the eight ctDNA-positive urine samples with increased TFx values in the CNV analysis overlapped with ctDNA-positive urine samples identified based on the CIA score. Two additional urine samples with increased TFx, but a negative CIA score, were characterized by a single loss of chromosome 18 and no further detectable alterations. One of these samples showed a corresponding ctDNA-positive plasma sample, identified based on both the CIA score and the estimated TFx derived from the CNV analysis with size selection, but without evidence of the chromosome 18 deletion. Overall, urine samples showed a markedly higher correlation between corresponding TFx values and the CIA scores (Spearman's ρ = 0.76) compared to plasma (Spearman's ρ = 0.44), which may be explained by the limited number of plasma samples with increased TFx values and/or CIA scores, thereby limiting robust correlation analysis. A strong correlation was expected between the two analyses, since both assessed genomic instability and were based on the same underlying sequencing count data. However, the two different algorithms yielded divergent results and varying sensitivities, highlighting the complementary information provided by the two analyses.

The approach for the calculation of CIA scores was based on previously published studies implemented in plasma cfDNA, as well as in DNA derived from urinary cell pellets[400,401]. Liu and colleagues performed multiple annealing and looping-based amplification cycles-next-generation sequencing (MALBAC-NGS) on urinary cell pellet samples from bladder cancer patients, patients with non-malignant urological diseases, and healthy controls[401]. In this study, bladder cancer patients could be distinguished based on the CIA score, with an overall sensitivity of 83.1% and accuracy of 89.0%. The performance was confirmed in a validation cohort, achieving a higher sensitivity of 89.2% and specificity of 90%. Sensitivity was correspondingly higher in muscle-invasive and high-grade bladder cancer compared to non-muscle-invasive and low-grade bladder cancer[401]. Chen and colleagues calculated copy number instability (CNI) scores with a similar algorithm, but based on lcWGS data derived from plasma samples of patients with advanced gastric cancer, collected at baseline and during the course of different therapy regimens[400]. They identified increased CNI scores in 49% of plasma samples at baseline and observed a significant decrease in CNI scores after therapy initiation, with fluctuations during treatment[400]. CNI scores were also evaluated in relation to therapy response, and patients with partial response tended to have lower CNI scores compared to patients who developed resistance[400]. The presented studies used two different approaches for defining cut-off values for chromosomal instability scores. While Liu and colleagues defined the cut-off based on ROC curves analysis[401], Chen and colleagues used the CNI scores of healthy controls for threshold definition, similar to the approach I applied in the present study, setting the cut-off at the mean plus one standard deviation[400]. Beyond the described algorithms for chromosomal instability assessment, several other approaches have been reported that assess chromosomal instability using different conceptual and computational frameworks[449-454]. Chromosomal instability has been described as the characteristic of exhibiting high rates of genomic alterations, including gains and losses of whole chromosomes, as well as point mutations, small-scale genomic alterations, and larger chromosomal rearrangements[455]. Integrating these diverse aberration types into a single quantitative metric that expresses the full extent of chromosomal instability remains challenging, resulting in the development of multiple techniques capturing distinct aspects of chromosomal instability[456]. In this context, Oza and colleagues proposed an R package that calculates six different chromosomal instability scores based on both numerical and structural aberrations identified in array-based data, but not in NGS data[452]. The authors compared structural (number of altered segments, fraction of genome altered), numerical (CNVs and chromosomal break points) and overall (total aberration index, modified total aberration index) scores and observed variable correlations between the metrics, as well as differences in the performance when distinguishing cancer-free controls from tumor samples across different cancer entities[457,458]. The authors highlighted the complementarity of assessing structural and numerical aberrations or overall genomic instability, as each may reflect distinct biological processes[457,458]. These conclusions support my findings in the genomic instability analysis, which revealed both consistent and complementary results when applying two different algorithms to the same lcWGS dataset: (i) CNV analysis, capturing segmental chromosomal alterations, and (ii) the CIA score, representing a combined metric of genome-wide regions with the strongest deviation from the copy-number-neutral state.

4.5 Epigenomic profiling of tumor-associated methylation patterns in liquid biopsies and PCa tissue

The methylation analysis represented one of the key elements of my PhD project, since a high potential for the challenging task of ctDNA detection in localized PCa was expected due to the early occurrence of aberrant DNA methylation during carcinogenesis[189,195,196]. Regional and gene-specific hypermethylation has been observed in primary PCa[165,195], with global hypomethylation potentially co-occurring at early disease stages and both aberrations accumulating during tumor progression[191,200-203].

The methylation analysis was performed using cfMeDIP-seq, enabling a genome-wide assessment of methylated cfDNA fragments based on a straightforward, time- and cost-effective approach, applicable to different sources of cfDNA or gDNA[331-333]. The enrichment of methylated (cf)DNA fragments is associated with both advantage and disadvantage. On the one hand, enrichment may increase the detectability of otherwise low tumor-derived signals relative to the high background signal. On the other hand, this enrichment is also associated with several limitations and technical challenges, as described in the following. The cfMeDIP-seq approach relies on immunoprecipitation of methylated DNA, using a methylation-specific, monoclonal antibody, and the enrichment performance, such as the percentage of recovered methylated DNA fragments, is highly dependent on the efficacy and specificity of the antibody[332,333]. However, internal investigations within our laboratory have revealed that these parameters could vary between different antibody lots, consequently impacting the comparability between in-house analyses performed with different antibody lots, as well as comparisons with external studies, for which antibody lot numbers were often not reported. This limitation was also mentioned by Shen and colleagues, who proposed adjusting for potential batch effects either by including methylated control samples across batches or by applying bioinformatic normalization and batch-correction strategies[333]. In the present study, I conducted all (cf)MeDIP-seq experiments on liquid biopsy, PCa tissue and buffy coat samples using the same antibody lot, in order to minimize any batch effects within this cohort.

The methylation enrichment procedure resulted in relative rather than absolute methylation levels, requiring data transformation to obtain absolute (β-) values. This transformation potentially introduced additional bias or noise into the data. Another challenge was that the method achieved a resolution of only 100–300 bp[331,338], whereas most published methylation data in the literature derive from array-based approaches or WGBS with single-base resolution. The relative enrichment of methylated DNA fragments in the cfMeDIP-seq approach primarily allowed for the evaluation of hypermethylation, while the assessment of hypomethylation was highly limited[333,338]. A relatively lower coverage of a genomic region could reflect insufficient enrichment, potentially due to either hypomethylation, technical factors, or underrepresentation of the region in cfDNA itself[331]. The immunoprecipitation process could introduce bias toward specific genomic regions and fragment length ranges, since the affinity of the antibody may differ between long and short fragments, as well as between CpG-dense regions and regions containing few or single methylated cytosines[338]. The presence of CNVs could likewise influence the interpretation of relative methylation changes[332,333], because a methylated genomic region with an unaltered methylation status could appear to be (falsely) hyper- or hypomethylated, if the same region harbored an amplification or deletion, respectively. To avoid such misinterpretation, I prepared and sequenced a matched non-enriched sample for each cfMeDIP-seq sample, which served as control for the bioinformatic CNV correction and for evaluating methylation enrichment quality.

Despite these limitations, the advantages of the cfMeDIP-protocol outweighed its potential challenges, particularly with regard to the translational focus of my PhD project, which emphasized clinical relevance and future applicability, with a demand for time- and cost-effective, yet highly informative diagnostic approaches. Moreover, the potential of cfMeDIP-seq has previously been demonstrated for cfDNA methylation profiling across multiple tumor entities, including early-stage disease[237,312,344-346]. Additional studies, applying alternative methylation-enrichment techniques, such as MBD-seq, further support the potential of genome-wide, enrichment-based methylome profiling for tumor detection and characterization, even in cancers of unknown primary[342,343].

4.5.1 Characterization of DMRs identified in plasma and urinary cfDNA

The determination of DMRs between PCa patients and cancer-free controls, as well as between PCa subgroups with different risk profile, was used to identify genome-wide regions with altered methylation status in primary PCa, and specifically in aggressive disease forms. DMRs were identified between metastatic PCa patients (M1) and either cancer-free controls or PCa patients without distant metastases (N0M0 or N1M0), respectively, in both plasma and urine. Notably, distinctly lower numbers of DMRs were detected in the assessment of urinary cfDNA. Unexpectedly, significant DMRs could not be identified between all PCa patients and controls, nor between the different risk-stratified PCa subgroups, either in plasma or in urine.

The identified DMRs in plasma cfDNA comprised substantially more hypermethylated than hypomethylated regions in metastatic PCa patients compared to either controls or non-metastatic PCa patients, respectively. These results were concordant with previously reported findings[354,355]. Chen and colleagues performed DMR analyses across genome-wide 300-bp windows based on cfMeDIP-seq data obtained from metastatic CPRC patients compared to patients with localized PCa[354]. They identified 7.6-fold more hypermethylated than hypomethylated DMRs[354]. These hypermethylated DMRs correlated with detectable ctDNA in plasma and were enriched for regulatory regions, such as promoter regions, predominantly affecting tumor suppressor genes rather than oncogenes[354]. Both aspects were not observed in association with hypomethylated DMRs and the authors suggested that non-tumor sources may impact hypomethylated DMRs[354]. Technical aspects may also influence the methylation analyses, including the limited ability of the cfMeDIP-seq method to assess hypomethylation, as previously discussed[333,338]. Notably, the overall number of detectable DMRs between metastatic and localized PCa (19,048 hypermethylated and 2,493 hypomethylated DMRs) reported by Chen and colleagues[354] was substantially higher than the findings in the present study. However, in contrast to my cohort with primary metastatic PCa patients, the external cohort exclusively comprised patients with mCRPC, for whom higher ctDNA abundance and more pronounced tumor-derived signals were expected. In the study by Chen *et al.*, hierarchical clustering across all plasma samples based on normalized methylation signals within the identified DMRs revealed increased and heterogeneous signals for samples from mCRPC patients, whereas samples from localized PCa patients harbored homogenously low signals[354]. These distinct differences in methylation patterns enabled the profound detection of DMRs between the two biologically distinct cohorts. The findings showed high concordance with an additional PCa study, performing genome-wide assessment on cfDNA methylation using the enrichment-based technology MBD-seq[355]. DMR analysis was performed between plasma samples from primary metastatic PCa patients and cancer-free controls, and identified a large number of DMRs (n = 32,679), with the majority comprising hypermethylated regions (24,654 DMRs; 75.4%) rather than hypomethylated regions (8,280 DMRs; 25.3%). The authors selected metastatic PCa samples with confirmed ctDNA presence and TFx values greater than 15%, determined based on CNV analysis using the *ichorCNA* algorithm, to increase the likelihood of detecting tumor-specific methylation signals. Indeed, methylation levels were shown to correlate with the estimated TFx values derived from the CNV analysis[355]. In contrast to the study by Chen and colleagues, unsupervised clustering based on the methylation values of 900 DMRs with the highest absolute difference in methylation levels did not clearly separate metastatic PCa patients from controls[355]. My data differed from these two studies, as the methylation signals derived from the DMR analysis between metastatic PCa patients and non-metastatic PCa patients or cancer-free controls, respectively, were predominantly driven by a limited number of plasma or urine samples exhibiting increased or heterogeneous methylation signals. The majority of these samples also displayed distinct tumor-derived signal in the genomic analyses of chromosomal instability or CNVs, supporting the presence of ctDNA in these samples. Notably, these prominent plasma and urine samples identified in the methylation analysis did not originate from the same patients, emphasizing the presence of a global tumor-associated signal rather than a patient-specific pattern.

In plasma, the resulting DMRs from both comparisons (PCa M1 vs. controls or PCa M1 vs. PCa M0) were each distributed by approximately one third across intronic regions and distal intergenic regions, while 17–18% were located in promotor regions, with CpG islands representing the most common genomic landscape within promoter-associated DMRs. Approximately half of all significant hypermethylated DMRs with at least 2-fold higher methylation signal in metastatic PCa patients compared to either controls or non-metastatic PCa patients, respectively, were shared between both comparisons. Consequently, these hypermethylated DMRs identified in both comparisons were considered to represent informative regions for primary metastatic PCa in plasma cfDNA, as opposed to PCa without distant metastases and cancer-free controls. These regions were distributed 31% in distal intergenic regions, 25% in introns, and 23% in promoter regions, with CpG islands representing the most frequent genomic context overall. In contrast, only three shared hypomethylated regions were identified, located in two exons and one intron. Hypermethylation in CpG islands has been widely investigated in the context of carcinogenesis, and the pivotal role of promoter-associated CpG islands hypermethylation, leading to inactivation of tumor suppressor genes, has been highlighted[185,459]. In addition, the role of methylation changes in introns or distal regulatory sites and their impact on gene expression, in the context of tumorigenesis, has also been investigated[460-463]. Consistent with this, DMRs identified for tumor characterization in cfDNA have been shown to be frequently located in CpG islands[343,346,354,464], while also contributing to introns and (distal) intergenic regions[464,465].

The findings from the DMR analyses in urinary cfDNA within my cohort deviated from those observed in plasma cfDNA, revealing distinctly lower numbers of DMRs in the comparisons of metastatic PCa patients versus non-metastatic PCa patients and cancer-free controls, respectively, as well as differences in the genomic locations of the identified DMRs. Approximately 50% of all DMRs in both comparisons were located in introns, while promoter regions accounted for less than 10%. The distribution of CpG-related landscapes was more heterogeneous, with open sea regions representing the overall largest proportion. Similar to plasma cfDNA, only a limited number of informative regions were identified in urine samples from primary metastatic PCa compared to both PCa patients without distant metastases and cancer-free controls. These 16 shared hyper- and hypomethylated regions detected in both comparisons, with at least 2-fold higher or lower methylation signals in metastatic PCa patients compared to the other two groups, represented approximately 20% of all significant DMRs determined in the two comparisons. Half of these shared DMRs were located in introns and only one hypermethylated DMR mapped to a promoter region on chromosome 10 (chr 10: 21814201–21814500), associated with the *SKI/DACH domain containing 1 (SKIDA1)* gene. The role of the *SKIDA1* gene for tumor development has previously been investigated in ovarian cancer, where significant difference in *SKIDA1* methylation between tumor and normal tissue were identified, as well as significant inverse correlations between methylation and gene expression were reported[466].

When comparing the identified DMRs in plasma and urine across the two comparisons, metastatic PCa versus non-metastatic PCa or metastatic PCa versus controls, seven and eight shared regions, respectively, were identified, of which three were common across both comparisons and both liquid biopsy sources. These three regions were located in two introns on chromosomes 3 and 20, associated with the genes *SOX2-OT* and *ZBTB46*, respectively, as well as one exon on chromosome 7, associated with the gene *PTPRN2*. SOX2 overlapping transcript (SOX2-OT) is a long non-coding RNA, and one of its introns contains the *SRY-box transcription factor 2 (SOX2)* gene, whose role has previously been implicated in cancer progression across several tumor entities, including PCa[467,468]. Overexpression of SOX2-OT itself has also been reported to be associated with tumorigenesis and cell proliferation[469-471]. Epigenetic changes and altered gene expression of *PTPRN2* were related to metabolic diseases, chronic kidney disease and tumorigenesis[472-474]. AR signaling pathway-related upregulation of *ZBTB46* has been suggested to promote metastasis and epithelial-to-mesenchymal transition, and its overexpression has been associated with androgen-independent proliferation and neuroendocrine PCa differentiation[475,476]. In addition, the comparison between metastatic PCa patients and controls in my cohort identified one shared hypermethylated promoter region on chromosome 6 (chr 6:28367101–28367400) in both plasma and urinary cfDNA, associated with the *ZSCAN12* gene. For the comparison between metastatic PCa patients and non-metastatic PCa patients, another shared hypermethylated promoter region on chromosome 10 (chr 10:72200701–72201000) was identified in plasma and urine, associated with the *nodal growth differentiation factor (NODAL)* gene. *ZSCAN12* has previously been shown to be significantly hypermethylated in high-grade compared to low-grade bladder cancer based on tissue analysis, and it was proposed that methylation biomarkers, such as *ZSCAN12*, could be used for risk stratification in bladder cancer[477]. In PCa, cfMeDIP-seq-based methylation profiling of plasma cfDNA from mCRPC patient treated with enzalutamide or abiraterone identified shared CpG sites among patients in promoter regions of eight genes, including *ZSCAN12*, which were differentially methylated across longitudinal samples within an intrapatient comparison assessing treatment-related methylation changes[478]. Another study identified DMRs associated with smoking status in PCa patients and demonstrated a strong correlation to mRNA expression of *ZSCAN12*, as well as an association between a smoking-methylation score and adverse PCa outcomes[479]. The role of the NODAL protein and the corresponding *NODAL* gene in tumorigenesis and cancer progression has also been highlighted across several tumor entities, as it has been shown to be re-expressed in tumor cells, while it is not expressed in normal tissue[480-482]. Re-expression of the protein NODAL has also been associated to growth and tumor progression in PCa cells[483].

In conclusion, tumor-informative DMRs were identified in both plasma and urinary cfDNA from metastatic PCa patients compared to non-metastatic PCa patients and cancer-free controls, respectively. These identified DMRs comprised both exclusive and shared regions across the two liquid biopsy sources, as well as across the two comparisons of metastatic PCa patients versus the less aggressive PCa subcohort or cancer-free controls, respectively. The presence of shared DMRs in plasma and urinary cfDNA, albeit in limited number, highlighted their potential role as tumor-informative regions for risk stratification of metastatic PCa. The accurate distinction between metastatic and non-metastatic PCa at initial diagnosis is of high clinical relevance, as it directly impacts therapeutic decision-making, including the choice between local and systemic treatment strategies. Clinical staging for the detection of lymph node and distant metastases is primarily based on imaging modalities, including MRI with or without PET/CT, which may also yield false-positive or false-negative results. In this context, additional diagnostic tools, such as liquid biopsy–based analyses, which could complement imaging and support a more accurate classification of metastatic disease by providing complementary molecular information, would be highly beneficial.

The identified, predominantly hypermethylated, regions were related to several genes for which associations between altered expression or epigenetic changes and tumorigenesis or tumor progression in PCa or in other tumor entities have previously been reported. Additional candidate genes with known aberrant (hyper)methylation in PCa, such as *GSTP1*, *cysteine rich protein 3 (CRIP3)*, *APC*, *prostaglandin-endoperoxide synthase 2 (PTGS2)*, among others, have also been identified in PCa tissue, as well as in plasma or urinary cfDNA from PCa patients[183], but were not detected in my data. Beyond the shared significant DMRs identified in plasma and urinary cfDNA for the comparison of metastatic PCa versus non-metastatic PCa or cancer-free controls, respectively, several regions were mutually exclusive for either plasma or urine. These regions could either be less specific for (advanced) PCa, or harbor complementary molecular information relevant to the profiling of metastatic PCa. Definitive conclusions could not be drawn based on the available data, and further investigation on this question, ideally through integrated molecular profiling of matched plasma, urine, and PCa tissue samples in a larger cohort, would be highly informative.

The assessment of DMRs in this study cohort revealed several challenges and limitations in both liquid biopsy sources. The majority of liquid biopsy samples from both PCa patients and controls, with few exceptions, harbored overall low methylation levels, which limited the identification of cohort-specific methylation features or clear differences in methylation status between cohorts. These low methylation levels may be attributed to both biological reasons and technical aspects. The majority of PCa patients harbored an early-stage disease, in which epigenetic alterations may already be present, but ctDNA release was expected to be low. Only a small number of plasma and urine samples demonstrated tumor-derived signals based on the CNV analysis, indicating limited ctDNA abundance. On the other hand, epigenetic changes are assumed occur early in PCa development, even in the absence of genomic alterations, and therefore a higher sensitivity of the methylation analysis for ctDNA detection might be expected. However, the effective sensitivity of methylation analysis remains constrained by the proportion of ctDNA within total cfDNA. Plasma and urine samples exhibiting distinct methylation signal frequently showed genomic alterations based on the CNV and/or chromosomal instability analysis, indicating the presence of ctDNA captured across multiple analytical modalities. However, a larger number of PCa patients with detectable ctDNA would have been required to enable extraction of further tumor-informative methylation markers based on the DMR analysis. Accordingly, a commonly applied strategy is to perform methylation biomarker detection in a subcohort of tumor patients with confirmed ctDNA presence, for example based on prior genomic analyses, and to subsequently evaluate these identified markers on remaining samples with initially low or undetectable ctDNA levels[237,355].
The particular challenge of DMR detection in urine samples might have been related to the heterogeneity in the composition of urinary cfDNA and the influence of other physiological and pathological factors[282,421,425,426].

In addition to biological challenges, technical limitations could have further contributed to reduced sensitivity of the methylation analysis. As previously described in this discussion section, the efficacy and specificity of the methylation enrichment plays a crucial role for successful implementation of the methylation analysis. All plasma and urine samples in my cohort harbored a lower relative methylation enrichment score, serving as a methylation quality metric, than expected based on the literature. Although other methylation QC parameters indicated acceptable performance, a limited efficiency of the methylation enrichment procedure could not be entirely excluded. Insufficient methylation enrichment would result in a more homogeneous genome-wide coverage and, consequently, smaller differences between samples from different cohorts. The coverage of genome-wide regions also impacted the transformation from relative to absolute methylation signals, represented as β-values.

Relative methylation signals were expressed as read counts per genomic window and further normalized for CNVs, effective library size, zygosity and window size to NRPKMs[382,389]. For the transformation to absolute methylation signals, additional parameters had to be considered, including the abundance of (noisy) background reads and CpG-dependent read enrichment. The enrichment was modelled by a function, describing the sample-specific methylation enrichment as dependency of the CpG density, thereby considering the assumption of an inverse relationship between methylation and CpG density[382,389]. Small deviations or minor errors in these assumptions could therefore substantially bias the transformation to β-values. The distributions of β-values in the liquid biopsy samples were overall low in both PCa patients and controls. For example, in all identified DMRs between metastatic PCa patients and cancer-free controls in plasma cfDNA, the average β-values were 0.32 in tumor samples and 0.26 in control samples. The maximum absolute difference of β-values between tumor and control samples was 0.3. However, methylation levels between metastatic PCa patients and controls differed significantly, with up to 6.5-fold higher methylation levels in PCa patients compared to controls. The assessment of relative methylation changes, expressed as logFC of β-values, was therefore more sensitive in capturing subtle, but significant differences. The DMR analysis was also performed based on NRPKM values instead of β-values, in line with previous studies[344,484], but this approach further reduced the numbers of detectable DMRs in both plasma and urinary cfDNA. However, NRPKM values were successfully applied to filter out low-coverage regions, thereby reducing the number of genome-wide regions included in the DMR analysis to the most informative ones. Several filtering strategies based on read count thresholds have previously been proposed[312,343,346], but a minimum number of two NRPKMs in at least one sample for each analyzed genomic-region proved to be the most appropriate threshold in this cohort, after testing multiple thresholds and filtering algorithms.

4.5.2 Tumor-associated DNA methylation patterns in PCa tissue

Genome-wide methylation profiling was also performed on gDNA derived from eight PCa tissue samples, and DMRs were determined in comparison with matched buffy coat samples, in order to identify tumor-specific methylation markers from PCa tissue, thereby potentially overcoming the limitations of the tissue-agnostic approach applied to liquid biopsy samples. PCa tissue samples harbored sufficient tumor cell content, reaching up to 90%, as determined by histopathological assessment, suggesting the presence of distinct tumor-derived methylation signals. Buffy coat samples were used as matched normal references, as they predominantly contained gDNA from white blood cells, which was expected to represent the non-tumorous genetic background. However, DMRs derived from this comparison may partially reflect differences related to the DNA source (blood cells versus prostate tissue), in addition to tumor-specific methylation signals. Matched normal prostate tissue was not available for this cohort and, moreover, the definition of normal prostate tissue may pose challenges, since PCa often comprises multiple cancerous or precancerous lesions distributed throughout the prostate gland, and a field effect has also been reported in PCa[191,485,486]. The concept of field cancerization describes that histologically non-suspicious regions distant from the malignant tumor cells may already harbor cancer-associated molecular alterations, such as epigenetic changes[203,487,488]. Consequently, the comparison between PCa tissue and matched buffy coat samples, rather than matched normal prostate tissue, revealed both advantages and limitations.

In contrast to the DMR analysis performed on liquid biopsy samples, comparative methylation profiling between matched PCa tissue and buffy coat samples revealed a larger number of significant DMRs, including 18,021 hypermethylated and 14,125 hypomethylated regions in PCa tissue. After stringent filtering based on the logFC of β-values between PCa tissue and matched buffy coat samples, 6,015 hypermethylated DMRs (logFC > 2) and 1,364 hypomethylated DMRs (logFC < -2) remained. Similar to plasma cfDNA, the majority of these DMRs were located in introns or distal intergenic regions. Only 13% of hypermethylated DMRs and 6% of hypomethylated DMRs were associated with promoter regions. Hypermethylation was identified in promoter regions of the genes *RASSF1*, *PITX2* and *secreted frizzled related protein 2 (SFRP2)* which have previously been reported to demonstrate promoter hypermethylation in PCa[183]. Additionally, hypermethylation was detected in distal intergenic regions associated with *CCND2* and *FOXA1*, respectively, for which aberrant methylation has also been described previously in PCa tissue[183]. However, other well-established methylation markers in PCa, such as *GSTP1* or *APC*[183], were not detected in my PCa tissue methylation data.

Given the large number of identified DMRs in the PCa tissue methylation analysis, I subsequently focused on the extraction of tumor-informative methylation markers with potential biological relevance and diagnostic value for PCa detection. For this purpose, an independent external dataset published by Börno *et al.* [403] was used to validate and benchmark the findings. Börno and colleagues performed MeDIP-seq on 51 PCa tissue samples and 53 normal prostate tissue samples, and assessed DMRs in genome-wide 500-bp windows[403]. The tumors were staged as pT2–pT4, harbored Gleason scores ranging from 6 to 9, and 13 cases were lymph node-positive (N1), while the remaining tumors were lymph node-negative or of unknown nodal status. Molecular information on the *TMPRSS2-ERG* fusion status was available, with 17 tumors being fusion-positive and 20 fusion-negative. The comparative methylation analysis between all PCa tissue and normal prostate tissue samples revealed approximately 147,000 significant DMRs, of which 85,406 were hypermethylated and 61,308 hypomethylated. Börno and colleagues further demonstrated that 887 combinations of two out of the 110 most significant DMRs were sufficient to distinguish PCa tissue from normal prostate tissue, highlighting the biological relevance of these regions and their potential utility for PCa detection[403]. Although raw sequencing data or a comprehensive list of all identified DMRs were not publicly available, the authors published a data frame containing the 100 highest-ranked (TOP 100) DMRs, all of which were hypermethylated in PCa tissue. These TOP 100 hypermethylated DMRs (defined in 500-bp windows) were compared with the 6,015 hypermethylated DMRs (defined in 300-bp windows) with logFC > 2 of β-values between PCa tissue and matched buffy coat samples derived from my methylation analysis. This comparison revealed that 67 of the 300-bp windows overlapped with 55 of the 500-bp windows, indicating that 55% of the TOP 100 DMRs reported by Börno *et al.* were also significantly hypermethylated in PCa tissue samples from my cohort. Eight of these overlapping regions belonged to the top 10% of hypermethylated DMRs in my analysis, ranked by the largest differences in methylation signals between PCa tissue samples and matched buffy coat samples (i.e, logFC of β-values). Notably, significant hypomethylated regions identified in my PCa tissue methylation analysis did not overlap with any of the TOP 100 hypermethylated DMRs reported by Börno *et. al.*, underscoring the specificity and robustness of the hypermethylated tumor-associated regions. I performed the DMR analysis in PCa tissue samples using 300-bp windows, rather than 500-bp windows as applied in the external study, to ensure methodological consistency with the methylation analyses of liquid biopsy samples conducted in the present study.

One third of the overlapping hypermethylated regions were located in introns, and approximately another third (29%) were associated with promoter regions. The predominant CpG-associated landscape comprised CpG islands, and 20 of the 300-bp windows mapped to CpG islands within promoter regions, where hypermethylation has been linked to carcinogenesis[185,459]. These promoter regions were associated with the following genes: *cholinergic receptor muscarinic 1 (CHRM1)*, *HOXC4*, *transmembrane protein 106A (TMEM106A)*, *tubulin alpha 4b (TUBA4B)*, *ZFP64 zinc finger protein (ZFP64)*, *solute carrier family 2 member 2 (SLC2A2)*, *prominin 1 (PROM1)*, *dystonin (DST)*, *zinc finger and SCAN domain containing 12 (ZSCAN12)*, *carboxypeptidase vitellogenic like (CPVL)*, *adenylate cyclase 1 (ADCY1)*, and *tachykinin precursor 1 (TAC1)*.

Several of these genes or their encoded proteins have previously been implicated in the context of PCa tumorigenesis, progression, or aggressiveness. The role of *ZSCAN12* was already addressed earlier in the Discussion in the context of DMR analyses in liquid biopsy samples. The expression of *PROM1* has been correlated with the expression of *transmembrane serine protease 4 (TMPRSS4)*, which is known to be overexpressed in several tumor entities, including PCa, where it promotes invasion, proliferation, and metastasis in PCa[489]. CHRM1 is also known to be overexpressed in PCa tissue, and to be involved in several pathways promoting PCa progression, invasion and metastasis[490,491]. CHRM1 has been shown to play a pivotal role in resistance to docetaxel therapy in PCa, thereby representing a potential therapeutic target[490]. *HOXC4* is known to exhibit aberrant expression in PCa, which has been associated with tumor aggressiveness, and, in a comprehensive study based on TCGA methylation and RNA sequencing data, it was ranked as top transcription factor connected to more than 100 enhancers in PCa[492,493]. *TMEM106A, TAC1*, and *CPVL* were each identified in two studies as top-ranked hypermethylated genes associated with differentially methylated CpGs between PCa tissue and adjacent normal prostate tissue from patients with primary PCa undergoing prostatectomy[494,495]. *TUBA4B*, which is usually lowly expressed or absent in normal tissue, has been reported to show high expression in PCa, thereby promoting tumor aggressiveness[496]. *ZFP64* has been proposed as a methylation marker in colorectal cancer, but has not yet been comprehensively investigated in PCa[497,498]. Aberrant expression of *SLC2A* family genes has been reported in PCa, with elevated *SLC2A2* expression associated with neuroendocrinal characteristics in CRPC, while the gene expression in primary PCa tissue was low[499,500]. The protein DST has been shown to be significantly downregulated in primary PCa compared to benign tissue, and both protein abundance and *DST* gene expression inversely correlated with PCa aggressiveness[501]. ADCY1 promotes catalyzation of adenosine triphosphate (ATP) to cyclic adenosine monophosphate (AMP) and has several functions in the brain and nervous system, but has also been reported to be overexpressed in NSCLC tissue and hypermethylated in glioblastoma[502]. In PCa, *ADCY1* gene expression was shown to be downregulated in Lymph Node Carcinoma of the Prostate (LNCaP) cells in response to androgens[503].

In conclusion, the comparison between DMR analyses performed in this study and an independent external study, based on two MeDIP-seq datasets from PCa tissue samples compared to either matched buffy coat or normal prostate tissue samples, respectively, successfully identified tumor-informative methylation marker regions. These 67 hypermethylated regions were identified as shared DMRs across both independent cohorts, comprising predominantly primary localized PCa, supporting their potential as representative methylation markers for primary PCa and their applicability in liquid biopsy samples for the detection of tumor-derived signals.

The external methylation dataset published by Börno *et al.* offered several advantages as an appropriate validation cohort. First, the use of the same MeDIP-seq technology for genome-wide methylation profiling avoided methodological discrepancies, which could otherwise introduce technical bias in the analyses, when comparing data generated by different platforms. The majority of publicly available PCa methylation datasets were generated using base-level resolution approaches, such as data from methylation arrays or WGBS, which are difficult to directly compare with MeDIP-seq data that have a resolution of approximately 100–300 bp. Transformation of methylation data with base-level resolution to larger genomic windows by aggregating methylation signals would likely reduce accuracy and introduce additional bias, particularly in regions with high variability of β-values within individual 300 bp-windows or in genomic regions not represented on methylation arrays. Another important advantage of the dataset by Börno and colleagues was the availability of normal prostate tissue as reference, enabling a direct comparison between PCa tissue and normal prostate tissue, which was not feasible in my cohort, as discussed previously in this section. Although the limitations of the usage of normal prostate tissues as reference were also described, the concordance between DMRs identified from PCa tissue compared with matched buffy coat samples in my study and those identified between PCa tissue and normal prostate tissue samples in the external study increased the robustness and validity of the findings, as it reduced the potential impact of differences in the DNA sources (blood cells versus tissue). However, this confounding factor could not be entirely eliminated, as prostate tissue samples likely contain variable proportions of infiltrating immune cells[504,505], and comparisons between tumor tissue and normal tissues may not represent a pure contrast between malignant PCa cells and healthy prostate cells alone.

One substantial limitation of the presented approach was the lack of tumor-informative hypomethylated regions, as the results from my PCa tissue-based methylation analysis could only be compared and validated against the TOP 100 hypermethylated DMRs reported in the external data, since no hypomethylated DMRs from PCa tissue were published by Börno *et al.* [403]. The role of DNA hypomethylation in tumor evolution and progression has been controversially discussed, and both hyper- and hypomethylation are thought to co-occur early in tumorigenesis, as well as during tumor progression[194]. Hypomethylation has also been reported to occur as late(r) events in PCa, and has been associated with tumor progression and advanced or more aggressive disease[191,200-202]. Moreover, global or widespread hypomethylation may also be present at earlier disease stages and has been detected in non-tumorous tissue adjacent to tumor regions in the context of PCa field effect[203]. Consequently, a combined analysis of hyper- and hypomethylation would be desirable for the assessment of PCa aggressiveness. The limited suitability of the MeDIP-seq approach for the reliable detection and interpretation of DNA hypomethylation has previously been discussed in this section. Börno and colleagues identified hypomethylated regions in the PCa tissue-based methylation analysis, particularly within long interspersed nuclear elements (LINE) and long terminal repeats (LTR), but also reported that hypermethylated regions were more homogeneously detectable across tumor samples compared with hypomethylated regions[403]. They related this observation to the more frequent occurrence of locus-specific hypermethylation, whereas hypomethylation may occur more globally and less specifically across the genome[403]. Regional and gene-specific hypermethylation are known to occur early in PCa development, but also to play a pivotal role in tumor progression, with *GSTP1* hypermethylation representing the most frequent methylation biomarker[191,194]. Accordingly, hypermethylated regions alone may already provide sufficient information for improved PCa detection and risk stratification. In the methylation analysis by Börno and colleagues, *GSTP1* was ranked 108th among significantly different DMRs and was therefore not included in the provided list containing the TOP 100 DMRs[403]. Consequently, this marker was also not represented among the 67 PCa tissue-derived methylation marker regions. However, the selected regions covered several candidate genes which have already been associated to PCa development, progression or aggressiveness, as discussed above. Taken together, these findings support the potential of the identified hypermethylated regions as informative biomarkers for primary PCa, even in the absence of complementary hypomethylation signals.

4.5.3 Applicability of PCa tissue-derived methylation markers in liquid biopsies

The PCa tissue-derived 67 methylation marker regions were applied to the cfMeDIP-seq data based on plasma and urinary cfDNA from PCa patients and controls. When assessing the distribution of methylation signals in these 67 genomic regions across groups, a gradual increase in methylation signals from controls to localized PCa and further to advanced PC was evident in both plasma and urinary cfDNA, with significantly increased methylation levels (i.e., β-values) in advanced PCa patients compared to both other groups. This gradual increase in methylation signals and the significant differences between advanced PCa patients and cancer-free controls supported the potential of these PCa tissue-derived regions as methylation biomarkers for PCa detection. Moreover, the significant differences observed between advanced and localized PCa patients suggest their informative value for PCa risk stratification. Despite these trends, unsupervised clustering based on the β-values across the 67 hypermethylated regions did not result in a clear separation of the three cohorts. Instead, a limited number of plasma and urine samples showed markedly aberrant methylation signals. Notably, several of these samples also harbored multiple CNVs and/or increased CIA scores, indicating ctDNA presence and supporting the biological relevance of the observed methylation patterns. In a few cases, aberrant methylation signals were observed in the absence of detectable genomic alterations, underscoring the complementary value of combined epigenomic and genomic analyses for ctDNA detection.

The robustness and performance of the identified methylation biomarkers was further assessed in an external dataset from Chen and colleagues who performed cfMeDIP-seq on cfDNA from 30 plasma samples from a cohort with primary localized PCa and 103 plasma samples from three cohorts with mCRPC patients[354]. The plasma samples from mCRPC patients were collected within three different clinical trials, either at baseline and after PSA progression, or over the course of targeted AR inhibition treatment[354]. Using this external PCa methylation dataset obtained from cfMeDIP-seq, methodological consistency to was maintained, aiming to minimize potential technical bias which may occur, when comparing data generated by different platforms. For the same reasons, the raw sequencing data obtained from the external methylation dataset were processed with the same bioinformatic pipeline as the sequencing data from my PCa cohort, aiming to achieve best possible comparability. The distribution of methylation signals (i.e., β-values) across the 67 hypermethylated regions was evaluated for the plasma samples from localized PCa and mCRPC patients in the external cohort, while plasma samples from cancer-free controls were not available in the external cohort. Unsupervised clustering using both PCA and hierarchical clustering based on β-values across the 67 genomic regions presented a distinct separation of plasma samples from mCRPC patients compared with localized PCa patients, with elevated methylation signals observed in mCRPC patients, underscoring the potential of the selected genomic regions for the detection of PCa-specific tumor signals. The clear separation between the two stage-stratified cohorts was likely explainable by both biological reasons and technical aspects. Chen and colleagues also demonstrated a clear separation of plasma samples from localized PCa and mCRPC patients using PCA based on the top 10,000 most variable bins of global methylation patterns in their analysis[354]. As the external dataset comprised two biologically distinct cohorts representing primary localized PCa and mCRPC, pronounced genome-wide molecular differences detectable in plasma cfDNA were expected. However, additional confounding factors or batch effects could not be excluded, as the samples originated from four different cohorts, although they were sequenced and bioinformatically processed together.

To broaden the comparative framework, hierarchical clustering based on β-values across the 67 methylation marker regions was complemented by including plasma samples from localized PCa patients, advanced PCa patients, and cancer-free controls in my cohort. For the external cohort of mCRPC patients, predominantly high methylation signals together with heterogeneous patterns comprising both high and low methylation signals were observed. In contrast, low methylation signals and limited heterogeneity were observed for localized PCa patients in the external cohort. The majority of plasma samples from localized and advanced PCa patients in my own cohort also showed low methylation signals across the 67 genomic regions and clustered with the external cohort of localized PCa patients. However, a few plasma samples also showed increased and heterogeneous methylation levels and were located adjacent to or within the cluster from mCRPC patients. The low methylation signals observed in both the external cohort of localized PCa patients and in the majority of all PCa patients from my cohort may reflect very low or absent ctDNA abundance in these samples. Moreover, any present ctDNA could potentially harbor only a limited number of methylation events due to the early tumor stage, which may not be sufficiently captured by the 67 methylation biomarkers representing hypermethylated regions in PCa. Regional hypermethylation was known to increase during tumor progression[191,194], as does the expected ctDNA abundance. This was reflected by the increased methylation signals observed in plasma samples from the mCRPC cohort. Chen and colleagues also reported a significant and strong correlation between hypermethylated DMRs and ctDNA content in plasma samples from one of the cohorts with mCRPC patients[354]. In this context, some plasma samples from my cohort likewise exhibited increased and heterogeneous methylation levels, suggesting the presence of ctDNA and indicating a more aggressive PCa phenotype in these samples.

The hierarchical clustering of plasma samples from the external cohort was further complemented by including urine samples from my cohort to evaluate the tumor-informative content in urinary cfDNA, in comparison to the high methylation signal observed in the mCRPC samples and the decreased signals in localized PCa samples. Additionally, I aimed to investigate whether the methylation signals in the two liquid biopsy sources differed noticeably from each other, although the same biomarker regions were targeted. The combination of urine samples from my cohort with the plasma samples from the external cohort provided a similar overall picture compared to the combination of external and internal plasma samples, and similar conclusions regarding the representation of methylation signals in urinary cfDNA from localized and advanced PCa patients in my cohort could be drawn as for plasma cfDNA. The urine samples from both PCa patients and controls predominantly clustered with the plasma samples from localized PCa in the external cohort and revealed overall low methylation levels. However, several urine samples formed two additional separate clusters within the broader cluster of samples with low methylation levels, showing slightly increased β-values across a limited number of genomic regions, in addition to the low methylation signals in the remaining regions. One of these clusters was located adjacent to the cluster of plasma samples from mCRPC patients, and a few urine samples also clustered directly with the mCRPC samples. Notably, these urine samples did not originate from the same PCa patients in my cohort whose plasma samples also clustered with the external mCRPC samples. However, urine samples from localized and advanced PCa patients in my cohort overall exhibited higher methylation signals in the PCa tissue-derived methylation marker regions compared to corresponding plasma samples. This might be either due to a higher ctDNA content in urinary cfDNA, potentially related to the closer proximity of PCa tumors to the urinary tract with possible direct shedding of ctDNA[278,506], or to other biological or technical differences between the two liquid biopsy sources. The improved representation of methylation signals across the 67 PCa tissue-derived methylation marker regions in urinary cfDNA compared with plasma cfDNA contrasted with the findings from my DMR analysis, in which fewer DMRs were identified in urinary cfDNA than in plasma cfDNA. This discrepancy might be associated with a potentially higher abundance of ctDNA in urine compared to plasma, which was captured by the assessment of only 67 targeted genomic regions, whereas tumor-derived signals may have been more diluted by non-tumorous, non-informative background noise in urine when assessing many thousands of regions in the DMR analysis. While this dilution would have strongly impacted the DMR analysis, it could have been at least partially mitigated by the tissue-based approach, which targeted specific tumor-informative regions and reduced the number of non-specific regions. Silva and colleagues similarly applied a tissue-based approach on plasma and urine samples from a small subset of four treatment-naïve metastatic PCa patients and reported a better representation of tissue-specific biomarkers in urine samples[285]. By assessing differentially methylated probes between PCa tissue and matched normal tissue samples using a genome-wide array-based methylation analysis and applying them to liquid biopsy samples, they identified up to 39.4% of tumor tissue-specific methylation alterations in plasma and up to 64.14% in urine[285]. However, they also noted that the majority of methylation signals detected in plasma and urine from these metastatic PCa patients did not originate from tumor-derived DNA, but from other non-tumorous sources, including hematopoietic cells in both biofluids, as well as healthy bladder and kidney cells in urine[285].

4.5.4 CtDNA detection based on PCa tissue-derived methylation markers in liquid biopsies

The identified PCa tissue-derived DNA methylation markers were used to calculate a methylation score in plasma and urinary cfDNA, serving as a single methylation summary metric for each sample, synoptically representing its methylation status in the selected 67 regions. As these 67 genomic regions were known to be hypermethylated in PCa tissue and also revealed higher methylation levels in plasma and urine samples from PCa patients, elevated methylation scores were equally expected in individual liquid biopsy samples from PCa patients compared with control samples. The methylation score was intended to be compared and combined with the cfDNA features derived from the analyses of cfDNA fragmentation, chromosomal instability and CNVs. Additionally, it could prospectively be integrated as a single epigenomic cfDNA feature into predictive modeling for PCa detection and risk stratification.

The methylation score showed an increasing trend for localized PCa patients and advanced PCa patients in both liquid biopsy sources, with significantly increased values for advanced PCa compared to both localized PCa patients and controls in urinary cfDNA, but not in plasma cfDNA. Urine samples harbored higher methylation scores compared to matched plasma samples, consistent with previously reported findings. Interestingly, plasma and urine samples originating from the same PCa patients as the eight PCa tissue samples used for the identification of the PCa tissue-derived methylation marker regions did not show an increased methylation score, except for one urine sample. This could likely be associated with low ctDNA abundance in these liquid biopsy samples. However, this observation also underscored the tumor-informative value of the identified regions, which revealed increased methylation patterns in tumor samples across both the internal and external PCa cohorts, rather than being restricted to matched liquid biopsy samples from the same patients. The application of the methylation score for ctDNA detection achieved rates of 13% for localized PCa patients and 28% for advanced PCa patients in plasma, as well as 8% for localized PCa patients and 44% for advanced PCa patients in urine, all at 95% specificity. Complementary assessment of matched plasma and urine samples further increased the ctDNA detection rates in both disease stages to 16% in localized PCa patients and 56% in advanced PCa patients. The methylation analysis achieved the highest ctDNA detection rates for advanced PCa among all genomic and epigenomic analysis types, highlighting the tumor-informative value of the selected methylation biomarkers and the potentially increased sensitivity of epigenomic compared with genomic cfDNA analyses in this cohort. For localized PCa patients, ctDNA detection rates were comparable to those of the other analyses, which might be associated with limited ctDNA abundance and/or a limited number of regional epigenetic alterations in these samples that were not sufficiently captured by the tissue-based methylation approach based on only 67 hypermethylated regions. Notably, several PCa patients showed elevated methylation scores despite low PSA levels according to clinical assessments, analogous to findings from the CNV analysis, highlighting the potential of complementary liquid biopsy analyses in addition to PSA-based diagnostics for risk stratification, particularly in cases of low or intermediate PSA levels.

Overall, this dataset outlined the informative value of genome-wide methylation profiling in gDNA from PCa tissue samples and in cfDNA from liquid biopsy samples for molecular PCa characterization. In particular, this approach enabled improved ctDNA detection through the application of PCa tissue-derived methylation biomarkers to plasma and urinary cfDNA. In this approach, several methylation markers were combined to potentially increase the sensitivity for PCa detection compared with the achievable sensitivity when assessing only single biomarkers. Different studies have previously highlighted the advantages of this approach, namely simultaneously assessing several tumor-informed biomarkers for variant calling or methylation analysis in cfDNA, across different tumor entities[204,297,343,484,507-510]. Thereby, ctDNA detection has also been achieved in cases with expected low ctDNA abundance or in early-stage cancers[204,297,509], for which tumor-agnostic approaches might not be expedient. Notably, the majority of tumor-informed approaches for variant calling were based on an individual patient-level, capturing between ten and several thousand genomic variants derived from matched tissue and detected in the liquid biopsy samples[204,297,507-509]. In contrast, this tissue-based assessment in plasma and urinary cfDNA was based on the aggregation of β-values in 67 methylation marker regions into a single methylation score. Other studies employed a more sophisticated approach and combined informative markers into machine learning-based classifiers[343,510]. This type of complex analysis type was also considered but ultimately not implemented in the present study, due to the small sample size, which likely limited model performance. Moreover, the aim was to establish a single methylation metric for integrative analyses with other genomic and epigenomic cfDNA features, rather than implementing a model based on methylation biomarkers.

Two independent studies performing genome-wide methylation profiling of PCa by Chen *et al.* and Lleshi *et al.* [354,355] have already been presented in this Discussion, and both similarities and differences have been emphasized. These two studies followed the concept of simultaneously analyzing various methylation markers (i.e., previously identified DMRs) and integrating them into machine learning-based algorithms for PCa classification. Notably, they performed tumor-agnostic approaches in plasma cfDNA, since DMRs were either identified in plasma samples from mCRPC patients with expected, high tumor load[354], or in samples from metastatic primary PCa patients with previously confirmed ctDNA presence[355], respectively. Chen and colleagues established a machine learning classifier based on the top 150 hyper- and hypomethylated DMRs identified by comparing mCRPC and primary localized PCa samples in the training set, and achieved a median AUC of 1 and a median accuracy of 0.989 for this model in the test set[354]. Applying this predictive model, all localized PCa patients were correctly classified, and a significant difference in probability distribution between localized PCa patients and the healthy control cohort was achieved[346,354]. Lleshi and colleagues included primary localized and metastatic PCa patients, and cancer-free controls in the methylation analysis[355], and initially assessed DMRs in a training set including plasma cfDNA from metastatic PCa patients with known ctDNA presence compared with controls. Subsequently, they extracted 900 significant DMRs with the highest absolute difference in methylation levels between the two cohorts and included them in a machine learning-based modeling, achieving an AUC of 0.96 in the test set and a sensitivity of 95% at 100% specificity for tumor detection with a defined probability cut-off at

0.86. Furthermore, the model achieved an AUC of 0.74 for discriminating localized PCa patients from cancer-free controls[355]. The authors used the cfMeDIP-seq dataset from Chen and colleagues[354] to validate their machine learning model and were able to detect 51% of the localized PCa patients and 89% of the mCRPC patients[355]. The two studies identified methylation biomarkers in advanced PCa based on genome-wide methylation profiling, which also showed partial applicability for the characterization of localized PCa. In this context, Lleshi and colleagues hypothesized that molecular characteristics in metastatic PCa might also be present at localized tumor stages, albeit with varying or lower extent[355]. My PCa cohort only included a small proportion of advanced or even metastatic PCa patients, limiting the opportunity to perform a similar comprehensive methylation analysis of metastatic PCa patients as in the two presented studies. Consequently, I favored a tissue-based approach for methylation biomarker identification. These PCa tissue-derived methylation markers extracted from primary localized PCa tissue samples revealed increased signals in both localized and advanced PCa samples, supporting the hypothesis proposed by Lleshi and colleagues that localized and metastatic PCa potentially share molecular features.

In summary, the assessment of epigenomic features, reflected by aberrant DNA methylation, partially enabled tumor-derived signal detection in advanced PCa and even localized PCa, as well as risk stratification of aggressive disease, meeting previously defined objectives. The potential of the methylation analysis was clearly demonstrated, while limitations were also discussed, particularly with regard to the applied cfMeDIP-seq technology for the profiling of methylation-enriched DNA. The growing interest in various applications of epigenomic liquid biopsy analyses has resulted in the development of new technologies and has also fostered ongoing research in this field. Emerging technologies for methylation analysis include enzymatic cytosine conversion as an alternative to bisulfite conversion for whole-genome methylome profiling, thereby enabling the exploitation of the advantages of WGBS and the acquisition of methylation status at base-level resolution, while simultaneously avoiding limitations such as DNA degradation and imbalanced representation of AT- or GC-rich genomic regions observed with WGBS[336]. Enzymatic methylation sequencing has already been applied to plasma cfDNA, and it has been shown to serve as an appropriate data source for integrated multiparametric analyses of genome-wide cfDNA methylation, cfDNA fragmentation, and CNVs for cancer detection across seven different tumor entities (i.e., breast, colorectal, esophageal, gastric, liver, lung, or pancreatic cancers) with different tumor stages[337]. This comprehensive genomic and epigenomic analysis followed the same approach as my PhD project, with the intention of improving sensitivity for ctDNA detection through the combination of different data modalities. The authors established a classifier based on the combination of methylation, fragmentation, and CNV features, and achieved an AUC of 0.966 for overall tumor detection in the test cohort, as well as a sensitivity of 73% at 99% specificity for the detection of early-stage cancers. Another novel technology for genome-wide methylation profiling with single-base resolution has been developed, which also performs enzymatic conversion and introduces so-called 6-letter sequencing, offering the assessment of both methylation and hydroxymethylation, as well as the genomic analysis of both CNVs and SNVs based on a single sequencing dataset[511]. This method shows high potential for future genomic and epigenomic applications, but requires further validation before broad implementation in the field of liquid biopsy becomes feasible.

4.6 Added value of multimodal cfDNA analyses for ctDNA detection in newly diagnosed PCa patients

The previous sections discussed the capability for ctDNA detection in plasma and urine from patients with primary PCa, focusing on the assessment of specific features in cfDNA fragmentation, the presence of CNVs, and the extent of genome-wide chromosomal instability in cfDNA, as well as methylation levels in prostate-informative hypermethylated regions. In all genomic and epigenomic analyses, ctDNA was detectable in a number of liquid biopsy samples from both localized and advanced PCa patients, with higher ctDNA detection rates and/or higher signal intensities in advanced PCa. Furthermore, ctDNA detection rates were improved through the complementary analysis of plasma and urine in both localized and advanced PCa patients. However, ctDNA detection rates remained moderate to low, even for metastatic disease. In advanced PCa, the highest ctDNA detection rates were achieved based on the methylation analysis, reaching 56%. For localized PCa, all four analyses harbored similar ctDNA detection rates between 15% and 18%. Building on the findings of improved ctDNA detection based on the complementary analysis of plasma and urinary cfDNA, I explored the potential of an integrated approach, combining genomic and epigenomic analyses across both biofluids to further enhance ctDNA detection in localized and advanced PCa. This multimodal liquid biopsy approach increased ctDNA detection rates for all PCa patients to 45%, with a strong increase observed for the cohort of localized PCa patients to 42%. The detection rate for advanced PCa (56%) remained unchanged compared with the complementary assessment of methylation scores in plasma and urine, thereby yielding the highest detection rate among all individual analysis types.

PCa is known to shed limited amounts of ctDNA into the bloodstream compared to other tumor entities[250]. However, increased ctDNA levels in advanced tumor stages or rising levels during tumor progression are likely to increase the value of ctDNA analyses for late-stage PCa[264-266]. Zhang and colleagues reported, based on findings from a pan-cancer study on 13,333 blood samples from 11,525 Chinese patients across 41 different tumor types, that PCa samples revealed the second-highest ctDNA detectability (87.9% ctDNA-positive cases of all PCa samples) after small cell lung cancer (91.1%), at an overall sensitivity of 73.5% for ctDNA detection across all tumor entities[512]. However, the PCa cohort comprised less than 1% of all analyzed blood samples, and almost all PCa patients harbored late-stage disease. The successful application of cfDNA analyses, particularly in plasma samples from metastatic (castration-resistant) PCa patients, has been widely reported[306,351,513-516]. CNVs have also been identified in urinary cfDNA from CRPC patients[517], consistent with findings in matched plasma cfDNA[518]. However, ctDNA detection in newly diagnosed PCa remains challenging. Pope *et al.* analyzed plasma samples from 118 localized PCa patients prior to prostatectomy, applying the highly sensitive method INVAR (high-sensitivity integration of variant reads), which is based on patient-specific hybrid-capture sequencing combined with specific error-suppression and signal-enrichment methods for mutation analysis, and detected ctDNA in only 19 patients (16%)[519]. Another research group applied targeted sequencing for genomic alteration analysis on plasma cfDNA and matched prostate tissue samples obtained from 35 newly diagnosed, metastatic PCa patients prior to therapy[412]. The median ctDNA fraction was 11%, and *TP53* mutations or DNA repair defects were detected in 47% and 21% of the plasma samples, respectively, revealing high concordance (80%) with matched tissue samples. While the individual plasma and tissue analyses alone did not detect somatic alterations in 36% of patients, the combined analysis achieved detectability in 94% of the patients. Hennigan and colleagues did not detect ctDNA in any plasma samples from 112 localized PCa patients using a combined approach with lcWGS for CNV analysis and targeted sequencing for mutation analysis, but detected ctDNA in plasma from four out of seven pretreated metastatic PCa patients[263]. The authors reported that the majority of patients with localized

PCa had PSA levels below 10 ng/ml, and even the patient with the highest PSA level (43.6 ng/ml) did not show any detectable CNVs[263]. Interestingly, in my cohort, nine of 13 PCa patients with detectable ctDNA in either plasma or urine, and all PCa patients with detectable ctDNA in urine based on the CNV analyses, harbored PSA levels below 10 ng/ml. One additional metastatic PCa patient with detectable ctDNA based on the CNV analysis showed a PSA level of 11 ng/ml.

In addition to genomic analyses applied for PCa detection, different epigenetic features have been previously investigated to assess their potential for PCa characterization. Several studies have examined cfDNA integrity in plasma and urine samples from primary PCa patients, demonstrating that PCa patients could be distinguished from healthy individuals or patients with benign prostatic hyperplasia based on cfDNA fragmentation-related analyses[268,269,520-522]. However, the assessment of cfDNA integrity in plasma and urine has remained insufficient for reliable diagnostic or prognostic approaches in PCa, and similar controversial results have been reported for other tumor entities[523,524]. Depending on the design of the cfDNA fragmentation analysis, these approaches appeared to be less tumor-specific for ctDNA detection compared with the identification of genomic alterations, which provide a more definitive tumor-derived signal. In contrast, cfDNA methylation profiling has revealed promising results with regard to increased sensitivity and specificity for ctDNA detection in PCa. In this context, two studies performing genome-wide methylation profiling of plasma cfDNA in PCa using either cfMeDIP-seq or MBD-seq[354,355] have already been described in the Discussion. The studies developed machine learning-based classifiers, based on DMRs identified in plasma cfDNA from either mCRPC patients or primary metastatic PCa patients compared with localized PCa patients or cancer-free controls, respectively, and demonstrated good performances for PCa classification, even in localized PCa patients or samples with low ctDNA abundance[354,355]. From a molecular perspective, comparative analyses across non-metastatic and metastatic disease stages, or between localized PCa and mCRPC patients may provide valuable insights into tumor progression and therapy resistance, beyond their immediate diagnostic application. While the distinction between primary metastatic PCa and localized PCa or non-malignancy, respectively, is of clinical importance for risk stratification and treatment selection, the discrimination between mCRPC and localized PCa is of limited relevance for routine clinical decision-making. Silva and colleagues performed array-based methylation analysis in plasma cfDNA and urine sediment in a small subset of four metastatic treatment-naïve PCa patients, and identified between 7.19% and 64.14% of these tumor tissue-specific methylation alterations in plasma, as well as between 6.98% and 39.4% in urine[285]. The authors concluded that their epigenome-wide methylation analysis revealed lower sensitivities compared with previous studies that focused on single methylation markers. Ellinger and colleagues highlighted the potential of targeted epigenetic cfDNA analyses for localized PCa diagnosis[525]. They detected hypermethylation in CpG islands associated with *GSTP1*, *PTGS2*, *G protein-coupled receptor class C group 5 member A (GPRC5A, TIG1)*, and *reprimo, TP53 dependent G2 arrest mediator homolog (Reprimo, RPRM)*, in serum cfDNA from PCa patients by real-time PCR, with higher frequencies (ranging from a maximum of 42.3% (*GSTP1*) to a minimum of 1.2% (*Reprimo*)) compared to men with benign prostate hyperplasia (7.7% frequency for *GSTP1* and 0% for all other markers) or healthy controls (0% detection for all markers) [525]. Bastian and colleagues also identified *GSTP1* hypermethylation in serum cfDNA in 12% of localized PCa patients prior to prostatectomy and 28% of hormone-refractory metastatic PCa patients[347]. Payne and colleagues similarly assessed *GSTP1* promoter methylation, as well as three additional methylation markers in primary PCa, and reported sensitivities of 17–37% in plasma cfDNA and up to 47–74% in urinary cfDNA at 95% specificity[526]. Hypermethylation in CpG islands located in the promoter region of the *GSTP1* gene has been the most frequently reported epigenetic cfDNA biomarker in PCa, alongside other individual methylation markers or marker panels, demonstrating variable sensitivity and specificity for PCa detection[527]. The informative value of targeted cfDNA methylation analyses has also been highlighted in the context of a multi-cancer early detection test, which achieved, through the application of a targeted methylation assay and classifier, an overall sensitivity of 51.6% (16.8% [stage I]–90.1% [stage IV]) at 99.5% specificity in 2,823 plasma samples across more than 20 primary tumor entities, and an overall sensitivity of 11.2% specifically for primary PCa (total n = 420; stage I: 3.2%, stage II: 4.9%, stage III: 14%, stage IV: 83.3%)[352].

The comparison with published data in the literature, highlighting several examples for ctDNA detection in primary localized or advanced PCa, revealed that my findings in the genomic and epigenomic cfDNA analyses in newly diagnosed PCa, revealing relatively low to moderate ctDNA detection rates, were consistent with the reported challenges of ctDNA detection in primary PCa. However, I could show that increased detection rates were achieved through the complementary analysis of plasma and urinary cfDNA, which has only rarely been reported in this combined analytical context[285,517,518,526,528]. The potential of combined approaches for improved sensitivity and a more robust detection of limited tumor-specific signals in PCa, including the use of multigene or multi-analyte assays and the combination of new biomarkers with PSA testing, has previously been highlighted[78,80,348,529-533].

Analogous to the assessment of single analytes, multianalyte liquid biopsy assays have predominantly revealed improved ctDNA detection and higher diagnostic or prognostic potential in advanced PCa stages, especially in mCRPC. Different studies either combined the characterization of CTCs (enumeration and/or RNA analysis) with other analytes, such as cfDNA[534-537], or performed profiling of both cfDNA and cfRNA[538] to assess genomic alterations (SNVs, CNVs), alongside transcriptomic aberrations. Hodara *et al.* performed a comprehensive multiparametric approach, including the assessment of CTC counts, SNVs in both CTC DNA and matched cfDNA, CNVs in single CTCs, as well as AR expression profiling in cfRNA, intentionally focusing on a cohort with mCRPC to increase the likelihood of detecting CTCs or genomic alterations compared with the expected limited detectability in earlier disease stages[539]. Connell *et al.* addressed the challenge of performing an integrative multianalyte approach for cancer prediction using urine samples from 207 men with clinical suspicion for PCa[540]. They combined methylation analysis of six genes (*GSTP1, SFRP2, insulin like growth factor binding protein 3 (IGFBP3), insulin like growth factor binding protein 7 (IGFBP7), APC,* and *PTSG2*) in urinary cell pellets, cfRNA quantification from urinary extracellular vesicles (EVs), and various clinical parameters. They established the random forest model *ExoMeth*, combining selected clinical variables, methylation features, and transcript targets, and accurately predicted overall PCa (AUC = 0.91), as well as tumors with Gleason score ≥3 + 4 (AUC = 0.89) and Gleason score ≥ 4 + 3 (AUC = 0.81), potentially avoiding biopsies in 66% of cases without missing aggressive cancers[540]. The model outperformed single-analyte models, underscoring the value of multiparametric analyses. However, further validation in larger datasets is required prior to potential clinical application. It has also been demonstrated in other tumor entities that the integration of multiple analytes or algorithms enhances diagnostic and prognostic accuracy compared with single analyses, and such approaches have also been shown to be effective in patients with low ctDNA levels or low mutational burden.

CtDNA detection in renal cancer has been shown to be similarly challenging and contradictory to that observed in PCa[250,541,542], but the combined analysis of plasma and urinary cfDNA has shown potential for enhanced ctDNA detection in this urological tumor entity, as demonstrated by Smith and colleagues[204]. The authors applied CNV analysis based on lcWGS data and mutation analysis either using highly sensitive INtegration of VAriant Reads-TAilored PAnel Sequencing (INVAR-TAPAS) or panel-sequencing of 10 high-ranked renal cancer genes on both plasma and urinary cfDNA from two cohorts with different renal tumor subtypes and tumor stages[204]. The analyses in plasma and urine revealed both consistent and complementary results, achieving ctDNA detection in plasma and/or urine based on CNV analyses in 39.6% of the cohort, and through the combination of CNV and mutation analyses in an overall 56.3% of the cohort, including predominantly non-metastatic renal cancer patients and even three patients with oncocytomas[204].

Peneder and colleagues performed a combined genomic and epigenomic analysis based on WGS of plasma cfDNA from pediatric sarcoma patients, with a particular focus on Ewing Sarcoma, to evaluate CNVs, detect the *EWS-ETS* fusion oncogene using both WGS and droplet digital PCR, and characterize cfDNA fragmentation features[543]. This combined approach enabled ctDNA detection with high sensitivity and accurately distinguished between Ewing sarcoma patients and healthy controls by applying different machine learning-based classifiers, achieving an AUC of up to 0.97[543]. Burgener and colleagues also performed a multimodal genomic and epigenomic approach by combining CAncer Personalized Profiling by deep Sequencing (CAPP-seq) with cfMeDIP-seq for the analysis of somatic mutations, fragmentation, and methylation in plasma cfDNA from patients with locoregionally-confined head and neck squamous cell carcinoma[346]. This approach enabled molecular profiling of otherwise low-abundance ctDNA and revealed an association between detectable ctDNA and the risk of advanced disease[346]. Similarly, cfDNA profiling of methylation, CNVs and fragmentation was combined by Cheng and colleagues, applying shallow-depth WGBS on urinary cfDNA for tumor detection in newly diagnosed low-grade and high-grade non-muscle-invasive bladder cancer (NMIBC, n = 19 and 18, respectively) or high-grade muscle-invasive bladder cancer (n = 9) [270]. Through the integration of tumor contribution, as defined by methylation-based deconvolution, with the proportion of genome-wide bins harboring hypomethylation or CNVs, respectively, they achieved sensitivities of 84.2% and 91.9% for characterization of total NMIBC, and overall sensitivity of 93.5% at 95.8% specificity for the detection of low-grade NMIBC[270]. Notably, the hypomethylation feature showed the highest sensitivity among all individual analyses, followed by methylation-based deconvolution, with distinctly increased sensitivity compared with CNV analysis. The combination of all approaches considerably exceeded the sensitivity of the individual analyses[270]. This integrative genomic and epigenomic approach applied by Cheng and colleagues to urinary cfDNA in bladder cancer showed conceptual similarities to my PCa analyses. It highlighted both the potential of epigenomic analyses and the superiority of multiparametric analyses over single-analyte assessments. These findings supported the suggestion that similar approaches could prospectively be pursued for early PCa detection and risk stratification of aggressive disease. The presented studies provide examples of various multi-analyte approaches for the analysis of plasma or urinary cfDNA across multiple tumor entities, demonstrating their potential for tumor detection. Multimodal liquid biopsy analyses have emerged as a highly relevant research area, attracting increasing attention from scientists in the field. Emerging liquid biopsy technologies have demonstrated improved sensitivity and specificity for ctDNA detection[248]. However, the sensitivity of these approaches remains constrained by the limited number of tumor-derived DNA fragments or other tumor-related components present in body fluids[544]. By evaluating multiple analytes or combining different technologies, these limitations may be mitigated, as such approaches are less reliant on the presence of a single marker or few targeted components[544]. The increasing use of machine learning algorithms for the integration of various data modalities will likely further strengthen the value of multiparametric approaches and enhance their potential in the future.

4.7 Integrative modeling of clinical parameters and cfDNA features for PCa risk stratification

As a final objective of this study, I assessed the clinical characteristics of the PCa patients in relation to findings in the genomic and epigenomic cfDNA analyses. Within the advanced PCa cohort, ctDNA was more frequently detected in patients with distant metastases (67%) compared to lymph-node only disease (56%), consistent with the higher tumor burden associated with metastatic dissemination. Accordingly, the majority of ctDNA-positive, metastasized patients harbored both lymph node and distant metastases. Furthermore, higher PSA levels were observed among ctDNA-positive compared to the ctDNA-negative cases. Interestingly, two of the three ctDNA-negative, metastasized PCa patients harbored only bone metastases, which initially suggested reduced ctDNA shedding in the absence of nodal or visceral involvement, although ctDNA has previously been detected in tumor patients with bone metastases[545-547]. Additionally, three of the five ctDNA-negative, advanced PCa patients showed no lymph node signal on PET/CT imaging, but were subsequently confirmed as lymph-node positive based on the pathological assessment of lymph nodes after surgery. Tracer uptake in PET/CT imaging can reflect tumor burden and biological activity. Accordingly, the absence of detectable lymph node involvement within PET/CT imaging might indicate a lower overall metabolically-active tumor burden or reduced lower metabolic activity of the respective lymph node metastases, which could contribute to reduced ctDNA shedding. A previous study in advanced breast cancer investigated the association between ctDNA detectability and metabolic parameters derived from ^{18}Fluorodeoxyglucose-PET/CT, and reported significantly higher whole-body metabolic tumor volume and whole-body total lesion glycolysis in ctDNA-positive compared with ctDNA-negative patients[548]. In addition, moderate correlations were observed between ctDNA variant allele frequencies and metabolic parameters, leading the authors to conclude that ctDNA detection reflects metabolic tumor burden[548]. Previous PSMA-PET/CT analyses in PCa have reported significant correlations between ctDNA concentrations and PSMA-positive tumor volumes in a cohort of advanced PCa patients, as well as in the subcohort of CRPC patients, whereas such associations were absent or weak in hormone-sensitive PCa alone[545]. These findings were primarily attributed to differences in overall metastatic burden, particularly the presence of high-volume osseous metastases. Normalized ctDNA levels relative to total PSMA-positive tumor volume were significantly higher in disseminated disease driven by lymph node or bone metastases than in prostate-confined disease, but did not differ between lymph node and bone metastases. Based on these results, the authors concluded that ctDNA levels reflect global tumor burden rather than differences in ctDNA shedding rates between the metastatic sites[545]. Consistent findings were reported in a large cohort of 491 patients with mCRPC, in which ctDNA levels correlated strongly with serum and radiographic measures of tumor burden[547]. CtDNA was detectable in 90% of patients with liver metastases, compared with only 59% of patients with bone metastases and 57% of patients with lymph node metastases. Among patients diagnosed with bone metastases, a higher number of lesions was associated with increased ctDNA levels, with patients harboring ≥10 bone lesions exhibiting 2.7-fold higher ctDNA levels than those with less than 10 lesions. Based on these findings, the authors concluded that higher ctDNA levels were indicative of increased overall disease burden. In another mCRPC cohort (n = 112), ctDNA levels were significantly associated with burden of bone metastases[546]. CtDNA concentrations correlated with patterns of bone metastases (i.e., oligometastatic, polymetastatic, and widespread disease) and longitudinal changes in ctDNA levels reflected transitions between these patterns over the disease and therapy course. Progression to widespread bone disease and the presence of visceral metastases were associated with worse overall survival. In contrast, ctDNA levels were not associated with characteristics of lymph node metastases. Taken together, previous studies have consistently suggested that ctDNA detectability has been primarily driven by global tumor burden, with specific metastatic sites influencing ctDNA levels primarily when they represented a large proportion of total disease burden. With regard to my PCa cohort, these observations indicate that the apparent association between bone-only metastases or PET/CT-negative lymph node metastases and ctDNA negativity, respectively, likely reflected a lower overall tumor burden rather than reduced ctDNA shedding from bone lesions or lower metabolically-active, PET/CT-negative lymph node metastases, respectively. These findings were consistent with additional tumor characteristics observed among ctDNA-positive patients, as the majority of ctDNA-positive, advanced PCa cases harbored locally-advanced tumors and higher Gleason patterns, both indicative of higher tumor burden and tumor aggressiveness.

Despite the higher prevalence of locally-advanced tumor stages and higher Gleason grades in advanced PCa, cfDNA metrics were not significantly correlated with these clinical parameters. Similarly, no associations between ctDNA positivity and Gleason score or primary tumor size, respectively, were observed in the localized PCa cohort. CtDNA was also detectable in small tumors (T1c), including one case with a Gleason 6 pattern. Overall, no consistent pattern linking clinical characteristics to ctDNA positivity could be identified across localized and advanced PCa patients. These findings suggested that ctDNA detectability was not solely determined by conventional clinical parameters such as PSA level, Gleason score, or primary tumor stage, but likely reflected a complex interplay of tumor biology, metastatic pattern, and ctDNA shedding dynamics. This observed heterogeneity highlighted both the biological complexity of PCa and the need for multimodal approaches that integrate molecular and clinical information. Furthermore, technical limitations of the applied assays might have contributed to limited ctDNA detection rates in localized PCa patients, as well as to moderate sensitivities in advanced and metastatic PCa patients. Furthermore, the relatively small size of the advanced PCa cohort, which comprised only one quarter of the total study population and included distant metastases in only half of these patients, constrained the extent to which the analyses were representative of primary advanced or metastatic PCa.

With regard to PSA testing as an essential component of the diagnostic algorithm for PCa, ctDNA was detectable in patients with low PSA levels, down to 2.7 ng/ml, and ctDNA was present in 43% of localized PCa and 67% of advanced PCa cases with low to intermediate PSA levels (<10 ng/ml). These findings supported the added value of multimodal cfDNA analyses, as ctDNA detection rates in patients with PSA levels below 10 ng/ml exceeded those reported for single genomic ctDNA assays[263,545,549]. Previous studies reported ctDNA detection rates of maximum 40% in advanced and metastatic PCa with PSA < 10 ng/ml[545,549], and undetectable ctDNA in localized PCa patients with low PSA levels[263]. Based on their findings, Hennigan and colleagues concluded that ctDNA detectability was not directly proportional to PSA levels in localized PCa patients, even at higher PSA levels, whereas ctDNA detectability more closely reflects tumor burden and PSA levels in metastatic PCa patients, likely due to intrinsic biological differences between the disease stages (e.g., ctDNA shedding, turnover, and vascular proximity) [263]. Given the limitations of PSA testing[415,416,550,551], these findings highlight the clinical relevance of multimodal liquid biopsy analyses as a complementary tool to PSA testing to inform risk assessment within the diagnostic workflow.

To evaluate synergistic effects of clinical parameters and cfDNA features, I generated multivariable models to reflect different clinically relevant diagnostic scenarios in PCa. These scenarios simulated decision-making within the diagnostic workflow and compared models based on clinical parameters alone with models integrating cfDNA-derived information to evaluate the discriminative performance in distinguishing advanced from localized PCa patients or PCa patients from tumor-free controls. In the first scenario, models incorporated PSA level, T stage and Gleason score, which derived from clinical and pathological assessment of the tumor after prostate biopsy, combined with or without binary status of ctDNA positivity. The second clinical scenario reflected the diagnostic timepoint prior to PCa tissue biopsy, when only PSA levels and clinical assessment of the tumor (T stage) are available. A third scenario simulated early PCa screening, in which only the results from PSA testing are available and guide further diagnostic procedures, such as MRI or prostate biopsy. In this context, analyses were focused on PCa patients with PSA levels <10 ng/ml to reflect a clinically relevant setting of inconclusive or low PSA levels[63]. Across all three clinical scenarios, integration of binary ctDNA status did not improve the discriminative power to distinguish advanced from localized PCa patients. However, when quantitative cfDNA features were combined with PSA levels, a significant improvement in discriminating both advanced and localized PCa patients, as well as PCa patients and PSA-elevated controls, in men with PSA < 10ng/ml, was achieved. This was reflected by markedly increased ROC-AUC values and improved threshold-based classification performance, represented by higher F1 scores. These findings supported the previous observations in this study that fragmentation, methylation, and genomic instability-related cfDNA features captured complementary information on disease biology beyond PSA levels alone. The quantitative cfDNA features were shown to reveal distinct differences between the subgroups, especially between advanced PCa patients and controls. For the multivariable modeling, these quantitative cfDNA features were condensed into five parameters, including the highest per-patient values for TFx, CIA scores and methylation score, which were measured in both liquid biopsy sources, and the two different cfDNA fragmentation features in plasma (10-bp oscillation score) and urine (P163–169 bp). The application of per-patient maximum values enabled the integration of complementary information from matched plasma and urine into a single synoptic measure, thereby capturing the strongest signal per individual. Interestingly, the highest signal for all three cfDNA features was more frequently observed in plasma for localized PCa patients and in urine for advanced PCa patients. Given the different release mechanisms of ctDNA into the blood circulation and into the urinary tract[272,278,279], this might have been associated with differences in tumor extent and anatomical proximity of the tumors to the bladder or urethra. In advanced disease, larger tumor size and burden, as well as closer distance to bladder or urethra might facilitate more direct shedding into urine, whereas, in earlier stages and or in cases with larger distance to urethra, ctDNA might be primarily released into the blood circulation and reaches the urinary tract through glomerular filtration, resulting in lower urinary ctDNA fractions.

An unexpected observation was that predicted probabilities from the multivariable models were lower for advanced PCa or tumor samples than for localized disease or PSA-elevated controls, respectively, although most individual cfDNA features were elevated in advanced disease. This apparent contradiction did not indicate a biological inverse association but reflected the multivariable structure of the model and partial multicollinearity among the cfDNA features. Several cfDNA features showed negative coefficients in the multivariable models, despite positive univariate associations with disease stage, which suggested that the models distributed predictive weight across correlated parameters. Consequently, the predicted probabilities had to be interpreted as relative risk scores which ranked the risk of an individual relative to the others, rather than the absolute risk for PCa disease (e.g., predicted probability of 0.8 means 80% PCa risk). The lower predicted probabilities for advanced disease did not reduce the discriminative performance, as both ROC-AUC and F1 scores improved substantially in the multimodal models, and ROC-based metrics depend on relative ranking rather than absolute probability values. The lower PR-AUC observed for multimodal models compared with PSA-only models was not considered as an inferior overall performance, but likely reflected class imbalance and differences in score distribution. The ROC-AUC was therefore considered as the primary performance metrics, with complementary reporting of the threshold-based measures for completeness. Importantly, the multivariable analysis was conducted with the exploratory aim to evaluate whether cfDNA features might improve model discrimination beyond PSA alone, instead of developing a calibrated and prevalence-adjusted predictive model. Bootstrap resampling further supported the robustness of the multimodal models, with consistently high mean AUC estimates. The relatively wide bootstrap confidence intervals likely reflected sampling variability, associated with the small sample sizes and the inclusion of correlated predictors. Likelihood ratio tests confirmed a statistically significant improvement in model fit based on the integration of cfDNA features in the given dataset, but these results had to be interpreted cautiously and did not allow for generalizability.

Overall, the multivariable modeling approach provided promising insights into the added value of quantitative cfDNA features for PCa risk stratification, particularly in cases with inconclusive PSA results. However, the modeling approach was limited by the relatively small sample size, particularly for advanced PCa cases, which increased the risk of model overfitting. Furthermore, partial collinearity among cfDNA-derived features restricted the interpretation of the effects and the direct biological interpretation of individual regression coefficients, as well as contributed to model overfitting. Future analyses may involve alternative modeling strategies, such as penalized regression approaches or the aggregation of correlated cfDNA features into synoptic scores, to avoid collinearity and improve coefficient stability, consequently reducing overfitting and improving interpretability of the results. Model performance was assessed using internal validation only, and could not be confirmed by an independent test or validation cohort. Consequently, the findings should be considered as exploratory and hypothesis-generating, requiring validation in larger, independent cohorts with predefined training and validation sets to confirm their robustness and to determine the clinical utility of multimodal cfDNA-based prediction models. With regard to clinical applications, these preliminary findings emphasized the potential added value of liquid biopsy analyses in complementing current diagnostics, particularly in cases where PSA testing alone does not provide sufficient risk discrimination. In an early screening setting based on PSA testing and in patients with inconclusive PSA results, integration of cfDNA analyses might assist in determining the necessity for further imaging, such as mpMRI, or support decision-making in case of ambiguous imaging findings. Combined with imaging data, cfDNA results could guide decision for further diagnostic procedures, such as tissue biopsy, potentially reducing unnecessary procedures. For PSA >10 ng/ml, positive liquid biopsy results may provide confirmation of disease extent, aiding in differentiating aggressive localized PCa from advanced or metastatic disease, and could thereby further guide risk stratification toward either localized or systemic treatment decisions.

5 Conclusion and future outlook

This PhD project investigated the potential of multimodal liquid biopsy analyses, combining genomic and epigenomic profiling of plasma and urinary cfDNA, to improve ctDNA detection in localized and advanced PCa at initial diagnosis. Addressing known challenges in cfDNA-based analyses in PCa, particularly in early-stage PCa, this work was motivated by the clinical need of identifying molecular biomarkers and diagnostic strategies which complement current diagnostic tools to inform clinical decision-making and improve PCa risk stratification.

The results demonstrated the benefit of integrating four cfDNA analysis types across two biofluids within a multimodal approach for comprehensive genomic and epigenomic tumor characterization in newly diagnosed PCa. The complementary assessment of fragmentation, CNVs, chromosomal instability and methylation in plasma and in urinary cfDNA achieved substantially increased ctDNA detection rates compared with analyses based on single parameters or single biofluids. This proof-of-concept study introduced the first reported multimodal framework combining genomic and epigenomic cfDNA analyses in both plasma and urine from PCa patients at initial diagnosis. Within this setting, the resulting sensitivities exceeded those previously reported in the literature, particularly for localized PCa, underscoring the potential of multimodal profiling for early molecular tumor characterization. Genomic and epigenomic cfDNA profiling revealed robust differences between PCa patients and cancer-free controls, with increasing tumor-associated signals observed in advanced disease. Identified genomic alterations and methylation patterns in plasma and urinary cfDNA were consistent with findings from PCa tissue analyses performed in this study, as well as with known molecular characteristics of primary PCa reported in the literature. CfDNA methylation results were further supported by comparison with an independent, external PCa tissue dataset. While differences were most pronounced in advanced PCa, selected cfDNA features, particularly related to methylation and chromosomal instability, also distinguished advanced from localized disease. Importantly, ctDNA was also detectable in both localized and advanced PCa patients with low to intermediate PSA levels below 10 ng/mL, a clinically relevant group in which current screening and diagnostic strategies often provide limited risk stratification. In this setting, multivariable models integrating PSA levels with quantitative cfDNA features demonstrated improved discrimination between advanced and localized PCa, as well as between PCa patients and PSA-elevated controls. Overall, these findings support the potential of cfDNA-based genomic and epigenomic markers to complement existing diagnostics and contribute to improved PCa risk stratification at initial diagnosis.

Despite these encouraging findings, several limitations have to be acknowledged. The overall cohort size was moderate, with a limited representation of advanced and metastatic PCa. Although the multimodal approach improved ctDNA detection, overall detection rates remained moderate to low, reflecting both biological factors, as well as current technical limitations of cfDNA-based assays. The multivariable modeling results should be regarded as exploratory and were affected by the moderate cohort size, partial multicollinearity among cfDNA features, limited biological interpretability of individual model parameters, and the risk of overfitting. Further studies with larger, independent cohorts are required to confirm the findings of this study, improve statistical power, and enable the development of robust and generalizable predictive models.

From a translational perspective, liquid biopsy assays offer a minimally invasive assessment of biomarkers with clinically feasible turnaround times. The lcWGS- and cfMeDIP-seq workflows applied in this study could be implemented within approximately two to three weeks at estimated per-patient costs below 400 €, with costs expected to further decrease as sequencing technologies continue to advance. These characteristics support the practical feasibility of integrating liquid biopsy analyses into clinical workflows. Prospective clinical trials investigating the applicability of multimodal liquid biopsy analyses integrated into diagnostic workflows will be essential to evaluate their performance and applicability in the clinical setting, as well as the acceptance by physicians and patients. Such trials represent a critical next step toward the clinical implementation of liquid biopsy analyses, as increasingly emphasized by the research community[216,248,552-554]. Future perspectives also include the combination of multimodal liquid biopsy data with additional clinical information, laboratory parameters, imaging modalities, and molecular datasets to achieve comprehensive, multimodal representations of PCa biology. Machine learning–based approaches may facilitate the integration and interpretation of these complex data layers, offering promising opportunities to further refine risk stratification and guide personalized treatment decisions.

To finally conclude, this proof-of-concept study demonstrated that the integration of multiple genomic and epigenomic cfDNA biomarkers in plasma and urine enhanced molecular characterization and ctDNA detection in newly diagnosed PCa. With further technological advancements, multimodal liquid biopsy analyses could complement established clinical diagnostics, and hold promise for supporting risk stratification and informing clinical decision-making, particularly in cases with inconclusive or low PSA levels. The proposed multimodal liquid biopsy strategies may also be extended to other urological malignancies beyond PCa, supporting broader advances in precision oncology.

6 References

1 Bray, F. *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **74**, 229-263 (2024). https://doi.org/10.3322/caac.21834

2 Zhou, C. K. *et al.* Prostate cancer incidence in 43 populations worldwide: An analysis of time trends overall and by age group. *Int J Cancer* **138**, 1388-1400 (2016). https://doi.org/10.1002/ijc.29894

3 Center, M. M. *et al.* International variation in prostate cancer incidence and mortality rates. *Eur Urol* **61**, 1079-1092 (2012). https://doi.org/10.1016/j.eururo.2012.02.054

4 Siegel, R. L., Miller, K. D., Wagle, N. S. & Jemal, A. Cancer statistics, 2023. *CA Cancer J Clin* **73**, 17-48 (2023). https://doi.org/10.3322/caac.21763

5 Jemal, A. *et al.* Prostate Cancer Incidence and PSA Testing Patterns in Relation to USPSTF Screening Recommendations. *JAMA* **314**, 2054-2061 (2015). https://doi.org/10.1001/jama.2015.14905

6 Moyer, V. A. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* **157**, 120-134 (2012). https://doi.org/10.7326/0003-4819-157-2-201207170-00459

7 Smailyte, G. & Aleknaviciene, B. Incidence of prostate cancer in Lithuania after introduction of the Early Prostate Cancer Detection Programme. *Public Health* **126**, 1075-1077 (2012). https://doi.org/10.1016/j.puhe.2012.01.016

8 Culp, M. B., Soerjomataram, I., Efstathiou, J. A., Bray, F. & Jemal, A. Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates. *Eur Urol* **77**, 38-52 (2020). https://doi.org/10.1016/j.eururo.2019.08.005

9 Ronckers, C. *et al.* Krebs in Deutschland für 2019/2020. *Robert Koch-Institut and Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V.*, 164 (2023). https://doi.org/10.25646/11357

10 Statistisches Bundesamt (Destatis). *Sterbefälle durch Krebs insgesamt 2023.* Gesundheitsberichterstattung des Bundes (2024). https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Todesursachen/Tabellen/sterbefaelle-krebs-insgesamt.html (accessed 28 November 2024)

11 Brierley, J. D., Gospodarowicz, M. K. & Wittekind, C. (eds). TNM classification of malignant tumours. 8th edn. 272 pp. John Wiley & Sons, Oxford (2016).

12 Johns, L. E. & Houlston, R. S. A systematic review and meta-analysis of familial prostate cancer risk. *BJU Int* **91**, 789-794 (2003). https://doi.org/10.1046/j.1464-410x.2003.04232.x

13 Zeegers, M. P., Jellema, A. & Ostrer, H. Empiric risk of prostate carcinoma for relatives of patients with prostate carcinoma: a meta-analysis. *Cancer* **97**, 1894-1903 (2003). https://doi.org/10.1002/cncr.11262

14 Salinas, C. A., Tsodikov, A., Ishak-Howard, M. & Cooney, K. A. Prostate cancer in young men: an important clinical entity. *Nat Rev Urol* **11**, 317-323 (2014). https://doi.org/10.1038/nrurol.2014.91

15 Houlahan, K. E. *et al.* A polygenic two-hit hypothesis for prostate cancer. *J Natl Cancer Inst* **115**, 468-472 (2023). https://doi.org/10.1093/jnci/djad001

16 Mucci, L. A. *et al.* Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA* **315**, 68-76 (2016). https://doi.org/10.1001/jama.2015.17703

17 Seibert, T. M. *et al.* Genetic Risk Prediction for Prostate Cancer: Implications for Early Detection and Prevention. *Eur Urol* **83**, 241-248 (2023). https://doi.org/10.1016/j.eururo.2022.12.021

18 Oh, M. *et al.* The association of BRCA1 and BRCA2 mutations with prostate cancer risk, frequency, and mortality: A meta-analysis. *Prostate* **79**, 880-895 (2019). https://doi.org/10.1002/pros.23795

19 Russo, J. & Giri, V. N. Germline testing and genetic counselling in prostate cancer. *Nat Rev Urol* **19**, 331-343 (2022). https://doi.org/10.1038/s41585-022-00580-7

20 Attard, G. *et al.* Prostate cancer. *Lancet* **387**, 70-82 (2016). https://doi.org/10.1016/s0140-6736(14)61947-4

21 Andersson, S. O. *et al.* Lifestyle factors and prostate cancer risk: a case-control study in Sweden. *Cancer Epidemiol Biomarkers Prev* **5**, 509-513 (1996).

22 Andersson, S. O. *et al.* Body size and prostate cancer: a 20-year follow-up study among 135006 Swedish construction workers. *J Natl Cancer Inst* **89**, 385-389 (1997). https://doi.org/10.1093/jnci/89.5.385

23 Wilson, K. M. & Mucci, L. A. Diet and Lifestyle in Prostate Cancer. *Adv Exp Med Biol* **1210**, 1-27 (2019). https://doi.org/10.1007/978-3-030-32656-2_1

24 Giovannucci, E., Rimm, E. B., Stampfer, M. J., Colditz, G. A. & Willett, W. C. Height, body weight, and risk of prostate cancer. *Cancer Epidemiol Biomarkers Prev* **6**, 557-563 (1997).

25 MacInnis, R. J. & English, D. R. Body size and composition and prostate cancer risk: systematic review and meta-regression analysis. *Cancer Causes Control* **17**, 989-1003 (2006). https://doi.org/10.1007/s10552-006-0049-z

26 Dennis, L. K., Dawson, D. V. & Resnick, M. I. Vasectomy and the risk of prostate cancer: a meta-analysis examining vasectomy status, age at vasectomy, and time since vasectomy. *Prostate Cancer Prostatic Dis* **5**, 193-203 (2002). https://doi.org/10.1038/sj.pcan.4500586

27 Fan, C.-Y. *et al.* Lower Urinary Tract Infection and Subsequent Risk of Prostate Cancer: A Nationwide Population-Based Cohort Study. *PLOS ONE* **12**, e0168254 (2017). https://doi.org/10.1371/journal.pone.0168254

28 Shrestha, E. *et al.* Profiling the Urinary Microbiome in Men with Positive versus Negative Biopsies for Prostate Cancer. *J Urol* **199**, 161-171 (2018). https://doi.org/10.1016/j.juro.2017.08.001

29 Bengtsen, M. B., Farkas, D. K., Borre, M., Sørensen, H. T. & Nørgaard, M. Acute urinary retention and risk of cancer: population based Danish cohort study. *BMJ* **375**, n2305 (2021). https://doi.org/10.1136/bmj.n2305

30 Siddiqui, M. M. *et al.* Vasectomy and risk of aggressive prostate cancer: a 24-year follow-up study. *J Clin Oncol* **32**, 3033-3038 (2014). https://doi.org/10.1200/jco.2013.54.8446

31 EAU Guidelines Office. EAU Guidelines. Edn. presented at the EAU Annual Congress Paris 2024. ISBN 978-94-92671-23-3. EAU Guidelines Office, Arnhem, The Netherlands (2024). https://uroweb.org/guidelines/prostate-cancer

32 Merriel, S. W. D., Funston, G. & Hamilton, W. Prostate Cancer in Primary Care. *Adv Ther* **35**, 1285-1294 (2018). https://doi.org/10.1007/s12325-018-0766-1

33 Heuveling, J. Früherkennung von Prostatakrebs. Deutsche Krebsgesellschaft (2021). https://www.krebsgesellschaft.de/onko-internetportal/basis-informationen-krebs/krebsarten/prostatakrebs/frueherkennung.html (accessed 28 November 2024)

34 Carvalhal, G. F., Smith, D. S., Mager, D. E., Ramos, C. & Catalona, W. J. Digital rectal examination for detecting prostate cancer at prostate specific antigen levels of 4 ng./ml. or less. *J Urol* **161**, 835-839 (1999).

35 Gosselaar, C., Roobol, M. J., Roemeling, S. & Schröder, F. H. The role of the digital rectal examination in subsequent screening visits in the European randomized study of screening for prostate cancer (ERSPC), Rotterdam. *Eur Urol* **54**, 581-588 (2008). https://doi.org/10.1016/j.eururo.2008.03.104

36 Cui, T., Kovell, R. C. & Terlecki, R. P. Is it time to abandon the digital rectal examination? Lessons from the PLCO Cancer Screening Trial and peer-reviewed literature. *Curr Med Res Opin* **32**, 1663-1669 (2016). https://doi.org/10.1080/03007995.2016.1198312

37 Mistry, K. & Cable, G. Meta-Analysis of Prostate-Specific Antigen and Digital Rectal Examination as Screening Tests for Prostate Carcinoma. *J Am Board Fam Pract* **16**, 95-101 (2003). https://doi.org/10.3122/jabfm.16.2.95

38 Halpern, J. A. *et al.* Prognostic Significance of Digital Rectal Examination and Prostate Specific Antigen in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Arm. *J Urol* **197**, 363-368 (2017). https://doi.org/10.1016/j.juro.2016.08.092

39 Arsov, C. *et al.* A randomized trial of risk-adapted screening for prostate cancer in young men—Results of the first screening round of the PROBASE trial. *Int J Cancer* **150**, 1861-1869 (2022). https://doi.org/10.1002/ijc.33940

40 Deutsche Krebsgesellschaft, Deutsche Krebshilfe & AWMF. Leitlinienprogramm Onkologie: S3-Leitlinie Prostatakarzinom, Langversion 7.0. (2024). https://www.leitlinienprogramm-onkologie.de/leitlinien/prostatakarzinom/

41 Deutsche Krebsgesellschaft, Deutsche Krebshilfe & AWMF. Leitlinienprogramm Onkologie: S3-Leitlinie Prostatakarzinom, Langversion 8.0. (2025). https://www.leitlinienprogramm-onkologie.de/leitlinien/prostatakarzinom/

42 Ilic, D., Neuberger, M. M., Djulbegovic, M. & Dahm, P. Screening for prostate cancer. *Cochrane Database Syst Rev* **2013**, Cd004720 (2013). https://doi.org/10.1002/14651858.CD004720.pub3

43 Ilic, D. *et al.* Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis. *BMJ* **362**, k3519 (2018). https://doi.org/10.1136/bmj.k3519

44 Schröder, F. H. *et al.* Prostate-cancer mortality at 11 years of follow-up. *N Engl J Med* **366**, 981-990 (2012). https://doi.org/10.1056/NEJMoa1113135

45 Schröder, F. H. *et al.* Screening for Prostate Cancer Decreases the Risk of Developing Metastatic Disease: Findings from the European Randomized Study of Screening for Prostate Cancer (ERSPC). *Eur Urol* **62**, 745-752 (2012). https://doi.org/10.1016/j.eururo.2012.05.068

46 Hugosson, J. *et al.* A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer. *Eur Urol* **76**, 43-51 (2019). https://doi.org/10.1016/j.eururo.2019.02.009

47 Van Poppel, H. *et al.* Serum PSA-based early detection of prostate cancer in Europe and globally: past, present and future. *Nat Rev Urol* **19**, 562-572 (2022). https://doi.org/10.1038/s41585-022-00638-6

48 Stenman, U. H., Leinonen, J., Zhang, W. M. & Finne, P. Prostate-specific antigen. *Semin Cancer Biol* **9**, 83-93 (1999). https://doi.org/10.1006/scbi.1998.0086

49 Lilja, H. *et al.* Prostate-specific antigen in serum occurs predominantly in complex with alpha 1-antichymotrypsin. *Clin Chem* **37**, 1618-1625 (1991).

50 Jung, K. *et al.* A gap between total prostate-specific antigen and the sum of free prostate-specific antigen plus alpha1-antichymotrypsin-prostate- specific antigen in patients with prostate carcinoma but not in those with benign prostate hyperplasia. *Clin Chem* **45**, 422-424 (1999).

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

52 Oremek, G. M. & Seiffert, U. B. Physical activity releases prostate-specific antigen (PSA) from the prostate gland into blood and increases serum PSA concentrations. *Clin Chem* **42**, 691-695 (1996).

53 Kovac, E. *et al.* Association of Baseline Prostate-Specific Antigen Level With Long-term Diagnosis of Clinically Significant Prostate Cancer Among Patients Aged 55 to 60 Years: A Secondary Analysis of a Cohort in the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. *JAMA Netw Open* **3**, e1919284 (2020). https://doi.org/10.1001/jamanetworkopen.2019.19284

54 Vickers, A. J. *et al.* The Relationship between Prostate-Specific Antigen and Prostate Cancer Risk: The Prostate Biopsy Collaborative Group. *Clin Cancer Res* **16**, 4374-4381 (2010). https://doi.org/10.1158/1078-0432.Ccr-10-1328

55 Thompson, I. M. *et al.* Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. *N Engl J Med* **350**, 2239-2246 (2004). https://doi.org/10.1056/NEJMoa031918

56 Lin, K., Lipsitz, R., Miller, T. & Janakiraman, S. Benefits and harms of prostate-specific antigen screening for prostate cancer: an evidence update for the U.S. Preventive Services Task Force. *Ann Intern Med* **149**, 192-199 (2008). https://doi.org/10.7326/0003-4819-149-3-200808050-00009

57 Loeb, S. *et al.* Overdiagnosis and overtreatment of prostate cancer. *Eur Urol* **65**, 1046-1055 (2014). https://doi.org/10.1016/j.eururo.2013.12.062

58 Aizer, A. A. *et al.* Cost implications and complications of overtreatment of low-risk prostate cancer in the United States. *J Natl Compr Canc Netw* **13**, 61-68 (2015). https://doi.org/10.6004/jnccn.2015.0009

59 Wilt, T. J. *et al.* Radical prostatectomy versus observation for localized prostate cancer. *N Engl J Med* **367**, 203-213 (2012). https://doi.org/10.1056/NEJMoa1113162

60 Benson, M. C., McMahon, D. J., Cooner, W. H. & Olsson, C. A. An algorithm for prostate cancer detection in a patient population using prostate-specific antigen and prostate-specific antigen density. *World J Urol* **11**, 206-213 (1993). https://doi.org/10.1007/BF00185070

61 Falagario, U. G. *et al.* Combined Use of Prostate-specific Antigen Density and Magnetic Resonance Imaging for Prostate Biopsy Decision Planning: A Retrospective Multi-institutional Study Using the Prostate Magnetic Resonance Imaging Outcome Database (PROMOD). *Eur Urol Oncol* **4**, 971-979 (2021). https://doi.org/10.1016/j.euo.2020.08.014

62 Catalona, W. J. *et al.* Comparison of percent free PSA, PSA density, and age-specific PSA cutoffs for prostate cancer detection and staging. *Urology* **56**, 255-260 (2000). https://doi.org/10.1016/S0090-4295(00)00637-3

63 Aminsharifi, A. *et al.* Prostate Specific Antigen Density as a Predictor of Clinically Significant Prostate Cancer When the Prostate Specific Antigen is in the Diagnostic Gray Zone: Defining the Optimum Cutoff Point Stratified by Race and Body Mass Index. *J Urol* **200**, 758-764 (2018). https://doi.org/10.1016/j.juro.2018.05.016

64 Nordström, T., Akre, O., Aly, M., Grönberg, H. & Eklund, M. Prostate-specific antigen (PSA) density in the diagnostic algorithm of prostate cancer. *Prostate Cancer Prostatic Dis* **21**, 57-63 (2018). https://doi.org/10.1038/s41391-017-0024-7

65 Washino, S. *et al.* Combination of prostate imaging reporting and data system (PI-RADS) score and prostate-specific antigen (PSA) density predicts biopsy outcome in prostate biopsy naïve patients. *BJU Int* **119**, 225-233 (2017). https://doi.org/10.1111/bju.13465

66 Yusim, I., Krenawi, M., Mazor, E., Novack, V. & Mabjeesh, N. J. The use of prostate specific antigen density to predict clinically significant prostate cancer. *Sci Rep* **10**, 20015 (2020). https://doi.org/10.1038/s41598-020-76786-9

67 Bruno, S. M. *et al.* PSA Density Help to Identify Patients With Elevated PSA Due to Prostate Cancer Rather Than Intraprostatic Inflammation: A Prospective Single Center Study. *Front Oncol* **11** (2021). https://doi.org/10.3389/fonc.2021.693684

68 Wei, J. T. *et al.* Early Detection of Prostate Cancer: AUA/SUO Guideline Part II: Considerations for a Prostate Biopsy. *J Urol* **210**, 54-63 (2023). https://doi.org/10.1097/Ju.0000000000003492

69 Bryant, R. J. *et al.* Predicting high-grade cancer at ten-core prostate biopsy using four kallikrein markers measured in blood in the ProtecT study. *J Natl Cancer Inst* **107**, djv095 (2015). https://doi.org/10.1093/jnci/djv095

70 Catalona, W. J. *et al.* A multicenter study of [-2]pro-prostate specific antigen combined with prostate specific antigen and free prostate specific antigen for prostate cancer detection in the 2.0 to 10.0 ng/ml prostate specific antigen range. *J Urol* **185**, 1650-1655 (2011). https://doi.org/10.1016/j.juro.2010.12.032

71 Nordström, T. *et al.* Comparison Between the Four-kallikrein Panel and Prostate Health Index for Predicting Prostate Cancer. *Eur Urol* **68**, 139-146 (2015). https://doi.org/10.1016/j.eururo.2014.08.010

72 Klein, E. A. *et al.* Clinical validation of IsoPSA, a single parameter, structure-focused assay for improved detection of prostate cancer: A prospective, multicenter study. *Urol Oncol* **40**, 408.e409-408.e418 (2022). https://doi.org/10.1016/j.urolonc.2022.06.002

73 Morote, J. *et al.* Improving the Early Detection of Clinically Significant Prostate Cancer in Men in the Challenging Prostate Imaging-Reporting and Data System 3 Category. *Eur Urol Open Sci* **37**, 38-44 (2022). https://doi.org/10.1016/j.euros.2021.12.009

74 Nordström, T. *et al.* Prostate cancer screening using a combination of risk-prediction, MRI, and targeted prostate biopsies (STHLM3-MRI): a prospective, population-based, randomised, open-label, non-inferiority trial. *Lancet Oncol* **22**, 1240-1249 (2021). https://doi.org/10.1016/s1470-2045(21)00348-x

75 Grönberg, H. *et al.* Prostate Cancer Diagnostics Using a Combination of the Stockholm3 Blood Test and Multiparametric Magnetic Resonance Imaging. *Eur Urol* **74**, 722-728 (2018). https://doi.org/10.1016/j.eururo.2018.06.022

76 Durand, X., Moutereau, S., Xylinas, E. & de la Taille, A. Progensa™ PCA3 test for prostate cancer. *Expert Rev Mol Diagn* **11**, 137-144 (2011). https://doi.org/10.1586/erm.10.122

77 Maggi, M. *et al.* SelectMDx and Multiparametric Magnetic Resonance Imaging of the Prostate for Men Undergoing Primary Prostate Biopsy: A Prospective Assessment in a Multi-Institutional Study. *Cancers (Basel)* **13**, 2047 (2021). https://doi.org/10.3390/cancers13092047

78 Van Neste, L. *et al.* Detection of High-grade Prostate Cancer Using a Urinary Molecular Biomarker-Based Risk Score. *Eur Urol* **70**, 740-748 (2016). https://doi.org/10.1016/j.eururo.2016.04.012

79 Tomlins, S. A. *et al.* Urine TMPRSS2:ERG Plus PCA3 for Individualized Prostate Cancer Risk Assessment. *Eur Urol* **70**, 45-53 (2016). https://doi.org/10.1016/j.eururo.2015.04.039

80 McKiernan, J. *et al.* A Novel Urine Exosome Gene Expression Assay to Predict High-grade Prostate Cancer at Initial Biopsy. *JAMA Oncol* **2**, 882-889 (2016). https://doi.org/10.1001/jamaoncol.2016.0097

81 Dickinson, L. *et al.* Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: recommendations from a European consensus meeting. *Eur Urol* **59**, 477-494 (2011). https://doi.org/10.1016/j.eururo.2010.12.009

82 Villers, A., Lemaitre, L., Haffner, J. & Puech, P. Current status of MRI for the diagnosis, staging and prognosis of prostate cancer: implications for focal therapy and active surveillance. *Curr Opin Urol* **19**, 274-282 (2009). https://doi.org/10.1097/MOU.0b013e328329a2ed

83 Eldred-Evans, D. *et al.* Population-Based Prostate Cancer Screening With Magnetic Resonance Imaging or Ultrasonography: The IP1-PROSTAGRAM Study. *JAMA Oncol* **7**, 395-402 (2021). https://doi.org/10.1001/jamaoncol.2020.7456

84 Radtke, J. P. *et al.* Multiparametric Magnetic Resonance Imaging (MRI) and MRI-Transrectal Ultrasound Fusion Biopsy for Index Tumor Detection: Correlation with Radical Prostatectomy Specimen. *Eur Urol* **70**, 846-853 (2016). https://doi.org/10.1016/j.eururo.2015.12.052

85 Bratan, F. *et al.* Influence of imaging and histological factors on prostate cancer detection and localisation on multiparametric MRI: a prospective study. *Eur Radiol* **23**, 2019-2029 (2013). https://doi.org/10.1007/s00330-013-2795-0

86 Drost, F. H. *et al.* Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer. *Cochrane Database Syst Rev* **4**, Cd012663 (2019). https://doi.org/10.1002/14651858.CD012663.pub2

87 Bonekamp, D. *et al.* Radiomic Machine Learning for Characterization of Prostate Lesions with MRI: Comparison to ADC Values. *Radiology* **289**, 128-137 (2018). https://doi.org/10.1148/radiol.2018173064

88 Greer, M. D. *et al.* Computer-aided diagnosis prior to conventional interpretation of prostate mpMRI: an international multi-reader study. *Eur Radiol* **28**, 4407-4417 (2018). https://doi.org/10.1007/s00330-018-5374-6

89 Weinreb, J. C. *et al.* PI-RADS Prostate Imaging – Reporting and Data System: 2015, Version 2. *Eur Urol* **69**, 16-40 (2016). https://doi.org/10.1016/j.eururo.2015.08.052

90 Turkbey, B. *et al.* Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. *Eur Urol* **76**, 340-351 (2019). https://doi.org/10.1016/j.eururo.2019.02.033

91 Oerther, B. *et al.* Cancer detection rates of the PI-RADSv2.1 assessment categories: systematic review and meta-analysis on lesion level and patient level. *Prostate Cancer Prostatic Dis* **25**, 256-263 (2022). https://doi.org/10.1038/s41391-021-00417-1

92 Pradere, B. *et al.* Nonantibiotic Strategies for the Prevention of Infectious Complications following Prostate Biopsy: A Systematic Review and Meta-Analysis. *J Urol* **205**, 653-663 (2021). https://doi.org/10.1097/ju.0000000000001399

93 Raaijmakers, R., Kirkels, W. J., Roobol, M. J., Wildhagen, M. F. & Schrder, F. H. Complication rates and risk factors of 5802 transrectal ultrasound-guided sextant biopsies of the prostate within a population-based screening program. *Urology* **60**, 826-830 (2002). https://doi.org/10.1016/s0090-4295(02)01958-1

94 Rosario, D. J. *et al.* Short term outcomes of prostate biopsy in men tested for cancer by prostate specific antigen: prospective evaluation within ProtecT study. *BMJ* **344**, d7894 (2012). https://doi.org/10.1136/bmj.d7894

95 Rouvière, O. *et al.* Use of prostate systematic and targeted biopsy on the basis of multiparametric MRI in biopsy-naive patients (MRI-FIRST): a prospective, multicentre, paired diagnostic study. *Lancet Oncol* **20**, 100-109 (2019). https://doi.org/10.1016/s1470-2045(18)30569-2

96 van der Leest, M. *et al.* Head-to-head Comparison of Transrectal Ultrasound-guided Prostate Biopsy Versus Multiparametric Prostate Resonance Imaging with Subsequent Magnetic Resonance-guided Biopsy in Biopsy-naïve Men with Elevated Prostate-specific Antigen: A Large Prospective Multicenter Clinical Study. *Eur Urol* **75**, 570-578 (2019). https://doi.org/10.1016/j.eururo.2018.11.023

97 D'Amico, A. V. *et al.* Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* **280**, 969-974 (1998). https://doi.org/10.1001/jama.280.11.969

98 Amin, M. B. *et al.* (eds). AJCC cancer staging manual. 8th edn. 1032 pp. Springer, Cham (2016). ISBN 978-3-319-40617-6.

99 Hövels, A. M. *et al.* The diagnostic accuracy of CT and MRI in the staging of pelvic lymph nodes in patients with prostate cancer: a meta-analysis. *Clin Radiol* **63**, 387-395 (2008). https://doi.org/10.1016/j.crad.2007.05.022

100 Luining, W. I. *et al.* Targeting PSMA Revolutionizes the Role of Nuclear Medicine in Diagnosis and Treatment of Prostate Cancer. *Cancers (Basel)* **14**, 1169 (2022). https://doi.org/10.3390/cancers14051169

101 Van Damme, J. *et al.* Comparison of (68)Ga-Prostate Specific Membrane Antigen (PSMA) Positron Emission Tomography Computed Tomography (PET-CT) and Whole-Body Magnetic Resonance Imaging (WB-MRI) with Diffusion Sequences (DWI) in the Staging of Advanced Prostate Cancer. *Cancers (Basel)* **13**, 5286 (2021). https://doi.org/10.3390/cancers13215286

102 Meijer, D. *et al.* External Validation and Addition of Prostate-specific Membrane Antigen Positron Emission Tomography to the Most Frequently Used Nomograms for the Prediction of Pelvic Lymph-node Metastases: an International Multicenter Study. *Eur Urol* **80**, 234-242 (2021). https://doi.org/10.1016/j.eururo.2021.05.006

103 Wu, H. *et al.* Diagnostic Performance of (68)Gallium Labelled Prostate-Specific Membrane Antigen Positron Emission Tomography/Computed Tomography and Magnetic Resonance Imaging for Staging the Prostate Cancer with Intermediate or High Risk Prior to Radical Prostatectomy: A Systematic Review and Meta-analysis. *World J Mens Health* **38**, 208-219 (2020). https://doi.org/10.5534/wjmh.180124

104 van Kalmthout, L. W. M. *et al.* Prospective Validation of Gallium-68 Prostate Specific Membrane Antigen-Positron Emission Tomography/Computerized Tomography for Primary Staging of Prostate Cancer. *J Urol* **203**, 537-545 (2020). https://doi.org/10.1097/ju.0000000000000531

105 Corfield, J., Perera, M., Bolton, D. & Lawrentschuk, N. (68)Ga-prostate specific membrane antigen (PSMA) positron emission tomography (PET) for primary staging of high-risk prostate cancer: a systematic review. *World J Urol* **36**, 519-527 (2018). https://doi.org/10.1007/s00345-018-2182-1

106 Hofman, M. S. *et al.* Prostate-specific membrane antigen PET-CT in patients with high-risk prostate cancer before curative-intent surgery or radiotherapy (proPSMA): a prospective, randomised, multicentre study. *Lancet* **395**, 1208-1216 (2020). https://doi.org/10.1016/s0140-6736(20)30314-7

107 Shen, G., Deng, H., Hu, S. & Jia, Z. Comparison of choline-PET/CT, MRI, SPECT, and bone scintigraphy in the diagnosis of bone metastases in patients with prostate cancer: a meta-analysis. *Skeletal Radiol* **43**, 1503-1513 (2014). https://doi.org/10.1007/s00256-014-1903-9

108 Lin, Y. *et al.* When to perform bone scintigraphy in patients with newly diagnosed prostate cancer? a retrospective study. *BMC Urol* **17**, 41 (2017). https://doi.org/10.1186/s12894-017-0229-z

109 van Leenders, G. *et al.* The 2019 International Society of Urological Pathology (ISUP) Consensus Conference on Grading of Prostatic Carcinoma. *Am J Surg Pathol* **44**, e87-e99 (2020). https://doi.org/10.1097/pas.0000000000001497

110 Kane, C. J., Eggener, S. E., Shindel, A. W. & Andriole, G. L. Variability in Outcomes for Patients with Intermediate-risk Prostate Cancer (Gleason Score 7, International Society of Urological Pathology Gleason Group 2-3) and Implications for Risk Stratification: A Systematic Review. *Eur Urol Focus* **3**, 487-497 (2017). https://doi.org/10.1016/j.euf.2016.10.010

111 Zumsteg, Z. S. *et al.* Unification of favourable intermediate-, unfavourable intermediate-, and very high-risk stratification criteria for prostate cancer. *BJU Int* **120**, E87-e95 (2017). https://doi.org/10.1111/bju.13903

112 Preisser, F. *et al.* Intermediate-risk Prostate Cancer: Stratification and Management. *Eur Urol Oncol* **3**, 270-280 (2020). https://doi.org/10.1016/j.euo.2020.03.002

113 Gnanapragasam, V. J. *et al.* The Cambridge Prognostic Groups for improved prediction of disease mortality at diagnosis in primary non-metastatic prostate cancer: a validation study. *BMC Med* **16**, 31 (2018). https://doi.org/10.1186/s12916-018-1019-5

114 Mohler, J. L. *et al.* Prostate cancer, version 2.2019, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* **17**, 479-505 (2019). https://doi.org/10.6004/jnccn.2019.0023

115 Humphrey, P. A. Histological variants of prostatic carcinoma and their significance. *Histopathology* **60**, 59-74 (2012). https://doi.org/10.1111/j.1365-2559.2011.04039.x

116 WHO Classification of Tumours Editorial Board (eds). WHO Classification of Tumours: Urinary and Male Genital Tumours. 5th edn. ISBN 9283245121. International Agency for Research on Cancer, Lyon (2022).

117 Haffner, M. C. *et al.* Genomic and phenotypic heterogeneity in prostate cancer. *Nat Rev Urol* **18**, 79-92 (2021). https://doi.org/10.1038/s41585-020-00400-w

118 Ali, A. *et al.* Prostate zones and cancer: lost in transition? *Nat Rev Urol* **19**, 101-115 (2022). https://doi.org/10.1038/s41585-021-00524-7

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

120 Bostwick, D. G., Liu, L., Brawer, M. K. & Qian, J. High-grade prostatic intraepithelial neoplasia. *Rev Urol* **6**, 171-179 (2004).

121 Zhou, M. High-grade prostatic intraepithelial neoplasia, PIN-like carcinoma, ductal carcinoma, and intraductal carcinoma of the prostate. *Mod Pathol* **31**, 71-79 (2018). https://doi.org/10.1038/modpathol.2017.138

122 Netto, G. J. & Epstein, J. I. Widespread high-grade prostatic intraepithelial neoplasia on prostatic needle biopsy: a significant likelihood of subsequently diagnosed adenocarcinoma. *Am J Surg Pathol* **30**, 1184-1188 (2006). https://doi.org/10.1097/01.pas.0000213324.97294.54

123 Borboroglu, P. G., Sur, R. L., Roberts, J. L. & Amling, C. L. Repeat biopsy strategy in patients with atypical small acinar proliferation or high grade prostatic intraepithelial neoplasia on initial prostate needle biopsy. *J Urol* **166**, 866-870 (2001).

124 Jiang, Z. *et al.* Alpha-methylacyl-CoA racemase: a multi-institutional study of a new prostate cancer marker. *Histopathology* **45**, 218-225 (2004). https://doi.org/10.1111/j.1365-2559.2004.01930.x

125 Epstein, J. I. *et al.* The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *Am J Surg Pathol* **40**, 244-252 (2016). https://doi.org/10.1097/pas.0000000000000530

126 Epstein, J. I., Allsbrook, W. C., Jr., Amin, M. B. & Egevad, L. L. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *Am J Surg Pathol* **29**, 1228-1242 (2005). https://doi.org/10.1097/01.pas.0000173646.99337.b1

127 Gleason, D. F. Classification of prostatic carcinomas. *Cancer Chemother Rep* **50**, 125-128 (1966).

128 Willemse, P. M. *et al.* Systematic Review of Active Surveillance for Clinically Localised Prostate Cancer to Develop Recommendations Regarding Inclusion of Intermediate-risk Disease, Biopsy Characteristics at Inclusion and Monitoring, and Surveillance Repeat Biopsy Strategy. *Eur Urol* **81**, 337-346 (2022). https://doi.org/10.1016/j.eururo.2021.12.007

129 Tosoian, J. J. *et al.* Active Surveillance of Grade Group 1 Prostate Cancer: Long-term Outcomes from a Large Prospective Cohort. *Eur Urol* **77**, 675-682 (2020). https://doi.org/10.1016/j.eururo.2019.12.017

130 Hamdy, F. C. *et al.* 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. *N Engl J Med* **375**, 1415-1424 (2016). https://doi.org/doi:10.1056/NEJMoa1606220

131 Lam, T. B. L. *et al.* EAU-EANM-ESTRO-ESUR-SIOG Prostate Cancer Guideline Panel Consensus Statements for Deferred Treatment with Curative Intent for Localised Prostate Cancer from an International Collaborative Study (DETECTIVE Study). *Eur Urol* **76**, 790-813 (2019). https://doi.org/10.1016/j.eururo.2019.09.020

132 Stolzenburg, J. U. *et al.* A comparison of outcomes for interfascial and intrafascial nerve-sparing radical prostatectomy. *Urology* **76**, 743-748 (2010). https://doi.org/10.1016/j.urology.2010.03.089

133 Walz, J. *et al.* A Critical Analysis of the Current Knowledge of Surgical Anatomy of the Prostate Related to Optimisation of Cancer Control and Preservation of Continence and Erection in Candidates for Radical Prostatectomy: An Update. *Eur Urol* **70**, 301-311 (2016). https://doi.org/10.1016/j.eururo.2016.01.026

134 Haglind, E. *et al.* Urinary Incontinence and Erectile Dysfunction After Robotic Versus Open Radical Prostatectomy: A Prospective, Controlled, Nonrandomised Trial. *Eur Urol* **68**, 216-225 (2015). https://doi.org/10.1016/j.eururo.2015.02.029

135 Ficarra, V. *et al.* Systematic review and meta-analysis of studies reporting urinary continence recovery after robot-assisted radical prostatectomy. *Eur Urol* **62**, 405-417 (2012). https://doi.org/10.1016/j.eururo.2012.05.045

136 Briganti, A. *et al.* Complications and other surgical outcomes associated with extended pelvic lymphadenectomy in men with localized prostate cancer. *Eur Urol* **50**, 1006-1013 (2006). https://doi.org/10.1016/j.eururo.2006.08.015

137 Fossati, N. *et al.* The Benefits and Harms of Different Extents of Lymph Node Dissection During Radical Prostatectomy for Prostate Cancer: A Systematic Review. *Eur Urol* **72**, 84-109 (2017). https://doi.org/10.1016/j.eururo.2016.12.003

138 Ramsay, C. *et al.* Systematic review and economic modelling of the relative clinical benefit and cost-effectiveness of laparoscopic surgery and robotic surgery for removal of the prostate in men with localised prostate cancer. *Health Technol Assess* **16**, 1-313 (2012). https://doi.org/10.3310/hta16410

139 Novara, G. *et al.* Systematic review and meta-analysis of perioperative outcomes and complications after robot-assisted radical prostatectomy. *Eur Urol* **62**, 431-452 (2012). https://doi.org/10.1016/j.eururo.2012.05.044

140 Van den Broeck, T. *et al.* Prognostic Value of Biochemical Recurrence Following Treatment with Curative Intent for Prostate Cancer: A Systematic Review. *Eur Urol* **75**, 967-987 (2019). https://doi.org/10.1016/j.eururo.2018.10.011

141 Merseburger AS, K. L., Krause BJ, Böhmer D, Perner S, von Amsberg G. The treatment of metastatic, hormone-sensitive prostatic carcinoma. *Dtsch Arztebl Int*, 622–632 (2022). https://doi.org/10.3238/arztebl.m2022.0294

142 Sweeney, C. J. *et al.* Chemohormonal Therapy in Metastatic Hormone-Sensitive Prostate Cancer. *N Engl J Med* **373**, 737-746 (2015). https://doi.org/10.1056/NEJMoa1503747

143 Fizazi, K. *et al.* Abiraterone acetate plus prednisone in patients with newly diagnosed high-risk metastatic castration-sensitive prostate cancer (LATITUDE): final overall survival analysis of a randomised, double-blind, phase 3 trial. *Lancet Oncol* **20**, 686-700 (2019). https://doi.org/10.1016/S1470-2045(19)30082-8

144 Petrylak, D. Therapeutic options in androgen-independent prostate cancer: building on docetaxel. *BJU Int* **96**, 41-46 (2005). https://doi.org/10.1111/j.1464-410X.2005.05946.x

145 Finianos, A., Gupta, K., Clark, B., Simmens, S. J. & Aragon-Ching, J. B. Characterization of Differences Between Prostate Cancer Patients Presenting With De Novo Versus Primary Progressive Metastatic Disease. *Clin Genitourin Cancer* **S1558-7673**, 30247-30241 (2017). https://doi.org/10.1016/j.clgc.2017.08.006

146 Phillips, R. *et al.* Outcomes of Observation vs Stereotactic Ablative Radiation for Oligometastatic Prostate Cancer: The ORIOLE Phase 2 Randomized Clinical Trial. *JAMA Oncol* **6**, 650-659 (2020). https://doi.org/10.1001/jamaoncol.2020.0147

147 Ost, P. *et al.* Surveillance or Metastasis-Directed Therapy for Oligometastatic Prostate Cancer Recurrence: A Prospective, Randomized, Multicenter Phase II Trial. *J Clin Oncol* **36**, 446-453 (2018). https://doi.org/10.1200/jco.2017.75.4853

148 Coleman, R. *et al.* Bone health in cancer: ESMO Clinical Practice Guidelines†. *Ann Oncol* **31**, 1650-1663 (2020). https://doi.org/10.1016/j.annonc.2020.07.019

149 Weinfurt, K. P. *et al.* The significance of skeletal-related events for the health-related quality of life of patients with metastatic prostate cancer. *Ann Oncol* **16**, 579-584 (2005). https://doi.org/10.1093/annonc/mdi122

150 Kyriakopoulos, C. E. *et al.* Chemohormonal Therapy in Metastatic Hormone-Sensitive Prostate Cancer: Long-Term Survival Analysis of the Randomized Phase III E3805 CHAARTED Trial. *J Clin Oncol* **36**, 1080-1087 (2018). https://doi.org/10.1200/jco.2017.75.3657

151 James, N. D. *et al.* Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial. *The Lancet* **387**, 1163-1177 (2016). https://doi.org/10.1016/S0140-6736(15)01037-5

152 James, N. D. *et al.* Abiraterone acetate plus prednisolone for metastatic patients starting hormone therapy: 5-year follow-up results from the STAMPEDE randomised trial (NCT00268476). *Int J Cancer* **151**, 422-434 (2022). https://doi.org/10.1002/ijc.34018

153 Davis, I. D. *et al.* Updated overall survival outcomes in ENZAMET (ANZUP 1304), an international, cooperative group trial of enzalutamide in metastatic hormone-sensitive prostate cancer (mHSPC). *J Clin Oncol* **40**, LBA5004-LBA5004 (2022). https://doi.org/10.1200/JCO.2022.40.17_suppl.LBA5004

154 Chi, K. N. *et al.* Apalutamide in Patients With Metastatic Castration-Sensitive Prostate Cancer: Final Survival Analysis of the Randomized, Double-Blind, Phase III TITAN Study. *J Clin Oncol* **39**, 2294-2303 (2021). https://doi.org/10.1200/jco.20.03488

155 Fizazi, K. *et al.* LBA5 A phase III trial with a 2x2 factorial design in men with de novo metastatic castration-sensitive prostate cancer: Overall survival with abiraterone acetate plus prednisone in PEACE-1. *Ann Oncol* **32**, S1299 (2021). https://doi.org/10.1016/j.annonc.2021.08.2099

156 Smith, M. R. *et al.* Darolutamide and Survival in Metastatic, Hormone-Sensitive Prostate Cancer. *N Engl J Med* **386**, 1132-1142 (2022). https://doi.org/10.1056/NEJMoa2119115

157 Epstein, J. I. *et al.* Proposed morphologic classification of prostate cancer with neuroendocrine differentiation. *Am J Surg Pathol* **38**, 756-767 (2014). https://doi.org/10.1097/pas.0000000000000208

158 Beltran, H. *et al.* Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat Med* **22**, 298-305 (2016). https://doi.org/10.1038/nm.4045

159 Aggarwal, R. *et al.* Clinical and Genomic Characterization of Treatment-Emergent Small-Cell Neuroendocrine Prostate Cancer: A Multi-institutional Prospective Study. *J Clin Oncol* **36**, 2492-2503 (2018). https://doi.org/10.1200/jco.2017.77.6880

160 Surintrspanont, J. & Zhou, M. Prostate Pathology: What is New in the 2022 WHO Classification of Urinary and Male Genital Tumors? *Pathologica* **115**, 41-56 (2022). https://doi.org/10.32074/1591-951x-822

161 Abeshouse, A. *et al.* The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011-1025 (2015). https://doi.org/10.1016/j.cell.2015.10.025

162 Taylor, B. S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11-22 (2010). https://doi.org/10.1016/j.ccr.2010.05.026

163 Chakravarthi, B. V., Nepal, S. & Varambally, S. Genomic and Epigenomic Alterations in Cancer. *Am J Pathol* **186**, 1724-1735 (2016). https://doi.org/10.1016/j.ajpath.2016.02.023

164 Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-158 (2007). https://doi.org/10.1038/nature05610

165 Cancer-associated genes. In: McKinnell, R. G. et al. (eds). *The Biological Basis of Cancer.* 2nd edn. Cambridge University Press, Cambridge, pp. 145–194 (2006). https://doi.org/10.1017/CBO9780511816642.007

166 Weinberg, R. A. Tumor suppressor genes. *Science* **254**, 1138-1146 (1991). https://doi.org/10.1126/science.1659741

167 Sager, R. Tumor suppressor genes: the puzzle and the promise. *Science* **246**, 1406-1412 (1989). https://doi.org/10.1126/science.2574499

168 Helleday, T., Petermann, E., Lundin, C., Hodgson, B. & Sharma, R. A. DNA repair pathways as targets for cancer therapy. *Nat Rev Cancer* **8**, 193-204 (2008). https://doi.org/10.1038/nrc2342

169 Alhmoud, J. F., Woolley, J. F., Al Moustafa, A.-E. & Malki, M. I. DNA Damage/Repair Management in Cancers. *Cancers (Basel)* **12**, 1050 (2020). https://doi.org/10.3390/cancers12041050

170 Mitra, A. *et al.* Prostate cancer in male BRCA1 and BRCA2 mutation carriers has a more aggressive phenotype. *Br J Cancer* **98**, 502-507 (2008). https://doi.org/10.1038/sj.bjc.6604132

171 Dall'Era, M. A. *et al.* Germline and somatic DNA repair gene alterations in prostate cancer. *Cancer* **126**, 2980-2985 (2020). https://doi.org/10.1002/cncr.32908

172 Arora, K. & Barbieri, C. E. Molecular Subtypes of Prostate Cancer. *Curr Oncol Rep* **20**, 58 (2018). https://doi.org/10.1007/s11912-018-0707-9

173 Mehra, R. *et al.* Comprehensive assessment of TMPRSS2 and ETS family gene aberrations in clinically localized prostate cancer. *Mod Pathol* **20**, 538-544 (2007). https://doi.org/10.1038/modpathol.3800769

174 Tomlins, S. A. *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-648 (2005). https://doi.org/10.1126/science.1117679

175 Mosquera, J.-M. *et al.* Characterization of TMPRSS2-ERG fusion high-grade prostatic intraepithelial neoplasia and potential clinical implications. *Clin Cancer Res* **14**, 3380-3385 (2008). https://doi.org/10.1158/1078-0432.CCR-07-5194

176 Yoshimoto, M. *et al.* Interphase FISH analysis of PTEN in histologic sections shows genomic deletions in 68% of primary prostate cancer and 23% of high-grade prostatic intra-epithelial neoplasias. *Cancer Genet Cytogenet* **169**, 128-137 (2006). https://doi.org/10.1016/j.cancergencyto.2006.04.003

177 Shah, R. B., Bentley, J., Jeffery, Z. & DeMarzo, A. M. Heterogeneity of PTEN and ERG expression in prostate cancer on core needle biopsies: implications for cancer risk stratification and biomarker sampling. *Hum Pathol* **46**, 698-706 (2015). https://doi.org/10.1016/j.humpath.2015.01.008

178 Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215-1228 (2015). https://doi.org/10.1016/j.cell.2015.05.001

179 Han, B. *et al.* Fluorescence in situ hybridization study shows association of PTEN deletion with ERG rearrangement during prostate cancer progression. *Mod Pathol* **22**, 1083-1093 (2009). https://doi.org/10.1038/modpathol.2009.69

180 Berger, M. F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470**, 214-220 (2011). https://doi.org/10.1038/nature09744

181 Quigley, D. A. *et al.* Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. *Cell* **174**, 758-769.e759 (2018). https://doi.org/10.1016/j.cell.2018.06.039

182 Barbieri, C. E. *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet* **44**, 685-689 (2012). https://doi.org/10.1038/ng.2279

183 Eickelschulte, S. *et al.* Biomarkers for the Detection and Risk Stratification of Aggressive Prostate Cancer. *Cancers (Basel)* **14**, 6094 (2022). https://doi.org/10.3390/cancers14246094

184 Hamid, A. A. *et al.* Compound Genomic Alterations of TP53, PTEN, and RB1 Tumor Suppressors in Localized and Metastatic Prostate Cancer. *Eur Urol* **76**, 89-97 (2019). https://doi.org/10.1016/j.eururo.2018.11.045

185 Laird, P. W. The power and the promise of DNA methylation markers. *Nat Rev Cancer* **3**, 253-266 (2003). https://doi.org/10.1038/nrc1045

186 Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* **3**, 662-673 (2002). https://doi.org/10.1038/nrg887

187 Ramsahoye, B. H. *et al.* Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S A* **97**, 5237-5242 (2000). https://doi.org/10.1073/pnas.97.10.5237

188 Cross, S. H. & Bird, A. P. CpG islands and genes. *Curr Opin Genet Dev* **5**, 309-314 (1995). https://doi.org/10.1016/0959-437X(95)80044-1

189 Plass, C. Cancer epigenomics. *Hum Mol Genet* **11**, 2479-2488 (2002). https://doi.org/10.1093/hmg/11.20.2479

190 Rechache, N. S. *et al.* DNA methylation profiling identifies global methylation differences and markers of adrenocortical tumors. *J Clin Endocrinol Metab* **97**, E1004-1013 (2012). https://doi.org/10.1210/jc.2011-3298

191 Massie, C. E., Mills, I. G. & Lynch, A. G. The importance of DNA methylation in prostate cancer development. *J Steroid Biochem Mol Biol* **166**, 1-15 (2017). https://doi.org/10.1016/j.jsbmb.2016.04.009

192 Verma, M. & Srivastava, S. Epigenetics in cancer: implications for early detection and prevention. *Lancet Oncol* **3**, 755-763 (2002). https://doi.org/10.1016/S1470-2045(02)00932-4

193 Hur, K. *et al.* Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. *Gut* **63**, 635 - 646 (2013). https://doi.org/10.1136/gutjnl-2012-304219

194 Ehrlich, M. DNA Hypomethylation In Cancer Cells. *Epigenomics* **1**, 239-259 (2009). https://doi.org/10.2217/epi.09.33

195 Sandoval, J. *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **6**, 692-702 (2011). https://doi.org/10.4161/epi.6.6.16196

196 Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* **41**, 178-186 (2009). https://doi.org/10.1038/ng.298

197 Moss, J. *et al.* Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat Commun* **9**, 5068 (2018). https://doi.org/10.1038/s41467-018-07466-6

198 Sun, K. *et al.* Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc Natl Acad Sci U S A* **112**, E5503-E5512 (2015). https://doi.org/10.1073/pnas.1508736112

199 Guo, S. *et al.* Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat Genet* **49**, 635-642 (2017). https://doi.org/10.1038/ng.3805

200 Yegnasubramanian, S. *et al.* DNA hypomethylation arises later in prostate cancer progression than CpG island hypermethylation and contributes to metastatic tumor heterogeneity. *Cancer Res* **68**, 8954-8967 (2008). https://doi.org/10.1158/0008-5472.Can-07-6088

201 Cho, N. Y. *et al.* Hypermethylation of CpG island loci and hypomethylation of LINE-1 and Alu repeats in prostate adenocarcinoma and their relationship to clinicopathological features. *J Pathol* **211**, 269-277 (2007). https://doi.org/10.1002/path.2106

202 Zelic, R. *et al.* Global DNA hypomethylation in prostate cancer development and progression: a systematic review. *Prostate Cancer Prostatic Dis* **18**, 1-12 (2015). https://doi.org/10.1038/pcan.2014.45

203 Yang, B. *et al.* Methylation Profiling Defines an Extensive Field Defect in Histologically Normal Prostate Tissues Associated with Prostate Cancer. *Neoplasia* **15**, 399-IN313 (2013). https://doi.org/10.1593/neo.13280

204 Smith, C. G. *et al.* Comprehensive characterization of cell-free tumor DNA in plasma and urine of patients with renal tumors. *Genome Med* **12**, 23 (2020). https://doi.org/10.1186/s13073-020-00723-8

205 Mouliere, F. *et al.* Fragmentation patterns and personalized sequencing of cell-free DNA in urine and plasma of glioma patients. *EMBO Mol Med* **13**, e12881 (2021). https://doi.org/10.15252/emmm.202012881

206 Swarup, N. *et al.* Multi-faceted attributes of salivary cell-free DNA as liquid biopsy biomarkers for gastric cancer detection. *Biomark Res* **11**, 90 (2023). https://doi.org/10.1186/s40364-023-00524-2

207 Mithani, S. K. *et al.* Mitochondrial resequencing arrays detect tumor-specific mutations in salivary rinses of patients with head and neck cancer. *Clin Cancer Res* **13**, 7335-7340 (2007). https://doi.org/10.1158/1078-0432.CCR-07-0220

208 Sriram, K. B. *et al.* Pleural fluid cell-free DNA integrity index to identify cytologically negative malignant pleural effusions including mesotheliomas. *BMC Cancer* **12**, 428 (2012). https://doi.org/10.1186/1471-2407-12-428

209 Mahmood, K. *et al.* High Yield of Pleural Cell-Free DNA for Diagnosis of Oncogenic Mutations in Lung Adenocarcinoma. *Chest* **164**, 252-261 (2023). https://doi.org/10.1016/j.chest.2023.01.019

210 Koc, A. *et al.* cfDNA in exhaled breath condensate (EBC) and contamination by ambient air: toward volatile biopsies. *J Breath Res* **13**, 036006 (2019). https://doi.org/10.1088/1752-7163/ab17ff

211 Nagai, K. *et al.* Genetic Analyses of Cell-Free DNA in Pancreatic Juice or Bile for Diagnosing Pancreatic Duct and Biliary Tract Strictures. *Diagnostics (Basel)* **12**, 2704 (2022). https://doi.org/10.3390/diagnostics12112704

212 Shen, N. *et al.* Bile cell-free DNA as a novel and powerful liquid biopsy for detecting somatic variants in biliary tract cancer. *Oncol Rep* **42**, 549-560 (2019). https://doi.org/10.3892/or.2019.7177

213 Mouliere, F. *et al.* Detection of cell-free DNA fragmentation and copy number alterations in cerebrospinal fluid from glioma patients. *EMBO Mol Med* **10**, e9323 (2018). https://doi.org/10.15252/emmm.201809323

214 Wang, Y. *et al.* Detection of tumor-derived DNA in cerebrospinal fluid of patients with primary tumors of the brain and spinal cord. *Proc Natl Acad Sci U S A* **112**, 9704-9709 (2015). https://doi.org/10.1073/pnas.1511694112

215 Michela, B. Liquid Biopsy: A Family of Possible Diagnostic Tools. *Diagnostics (Basel)* **11**, 1391 (2021). https://doi.org/10.3390/diagnostics11081391

216 Alix-Panabières, C., Marchetti, D. & Lang, J. E. Liquid biopsy: from concept to clinical application. *Sci Rep* **13**, 21685 (2023). https://doi.org/10.1038/s41598-023-48501-x

217 Wan, J. C. M. *et al.* Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* **17**, 223-238 (2017). https://doi.org/10.1038/nrc.2017.7

218 Gahlawat, A. W., Witte, T., Sinn, P. & Schott, S. Circulating cf-miRNA as a more appropriate surrogate liquid biopsy marker than cfDNA for ovarian cancer. *Sci Rep* **13**, 5503 (2023). https://doi.org/10.1038/s41598-023-32243-x

219 Grosgeorges, M. *et al.* A straightforward method to quantify circulating mRNAs as biomarkers of colorectal cancer. *Sci Rep* **13**, 2739 (2023). https://doi.org/10.1038/s41598-023-29948-4

220 Haupts, A. *et al.* Comparative analysis of nuclear and mitochondrial DNA from tissue and liquid biopsies of colorectal cancer patients. *Sci Rep* **11**, 16745 (2021). https://doi.org/10.1038/s41598-021-95006-6

221 Meddeb, R. *et al.* Quantifying circulating cell-free DNA in humans. *Sci Rep* **9**, 5220 (2019). https://doi.org/10.1038/s41598-019-41593-4

222 Yu, W. *et al.* Exosome-based liquid biopsies in cancer: opportunities and challenges. *Ann Oncol* **32**, 466-477 (2021). https://doi.org/10.1016/j.annonc.2021.01.074

223 Vitale, S. R. *et al.* Detection of tumor-derived extracellular vesicles in plasma from patients with solid cancer. *BMC Cancer* **21**, 315 (2021). https://doi.org/10.1186/s12885-021-08007-z

224 Best, M. G. *et al.* Swarm Intelligence-Enhanced Detection of Non-Small-Cell Lung Cancer Using Tumor-Educated Platelets. *Cancer Cell* **32**, 238-252.e239 (2017). https://doi.org/10.1016/j.ccell.2017.07.004

225 Best, Myron G. *et al.* RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. *Cancer Cell* **28**, 666-676 (2015). https://doi.org/10.1016/j.ccell.2015.09.018

226 Alix-Panabières, C. & Pantel, K. Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy. *Cancer Discov* **6**, 479-491 (2016). https://doi.org/10.1158/2159-8290.Cd-15-1483

227 Rieckmann, L. M. *et al.* Diagnostic leukapheresis reveals distinct phenotypes of NSCLC circulating tumor cells. *Mol Cancer* **23**, 93 (2024). https://doi.org/10.1186/s12943-024-01984-2

228 Overman, M. J. *et al.* Use of Research Biopsies in Clinical Trials: Are Risks and Benefits Adequately Discussed? *J Clin Oncol* **31**, 17-22 (2013). https://doi.org/10.1200/jco.2012.43.1718

229 McGranahan, N. & Swanton, C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* **27**, 15-26 (2015). https://doi.org/10.1016/j.ccell.2014.12.001

230 Ilié, M. & Hofman, P. Pros: Can tissue biopsy be replaced by liquid biopsy? *Transl Lung Cancer Res* **5**, 420-423 (2016). https://doi.org/10.21037/tlcr.2016.08.06

231 Yates, L. R. *et al.* Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* **21**, 751-759 (2015). https://doi.org/10.1038/nm.3886

232 Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883-892 (2012). https://doi.org/10.1056/NEJMoa1113205

233 De Mattos-Arruda, L. *et al.* Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. *Ann Oncol* **25**, 1729-1735 (2014). https://doi.org/10.1093/annonc/mdu239

234 Lebofsky, R. *et al.* Circulating tumor DNA as a non-invasive substitute to metastasis biopsy for tumor genotyping and personalized medicine in a prospective trial across all tumor types. *Mol Oncol* **9**, 783-790 (2015). https://doi.org/10.1016/j.molonc.2014.12.003

235 Dietz, S. *et al.* Global DNA methylation reflects spatial heterogeneity and molecular evolution of lung adenocarcinomas. *Int J Cancer* **144**, 1061-1072 (2019). https://doi.org/10.1002/ijc.31939

236 Murtaza, M. *et al.* Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat Commun* **6**, 8760 (2015). https://doi.org/10.1038/ncomms9760

237 Janke, F. *et al.* Longitudinal monitoring of cell-free DNA methylation in ALK-positive non-small cell lung cancer patients. *Clin Epigenetics* **14**, 163 (2022). https://doi.org/10.1186/s13148-022-01387-4

238 Dietz, S. *et al.* Longitudinal therapy monitoring of ALK-positive lung cancer by combined copy number and targeted mutation profiling of cell-free DNA. *EBioMedicine* **62**, 103103 (2020). https://doi.org/10.1016/j.ebiom.2020.103103

239 Riediger, A. L. *et al.* Mutation analysis of circulating plasma DNA to determine response to EGFR tyrosine kinase inhibitor therapy of lung adenocarcinoma patients. *Sci Rep* **6**, 33505 (2016). https://doi.org/10.1038/srep33505

240 Diehl, F. *et al.* Circulating mutant DNA to assess tumor dynamics. *Nat Med* **14**, 985-990 (2008). https://doi.org/10.1038/nm.1789

241 Newman, A. M. *et al.* An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med* **20**, 548-554 (2014). https://doi.org/10.1038/nm.3519

242 Dawson, S. J. *et al.* Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* **368**, 1199-1209 (2013). https://doi.org/10.1056/NEJMoa1213261

243 Forshew, T. *et al.* Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* **4**, 136ra168 (2012). https://doi.org/10.1126/scitranslmed.3003726

244 Mandel, P. & Metais, P. [Nuclear Acids In Human Blood Plasma]. *C R Seances Soc Biol Fil* **142**, 241-243 (1948).

245 Jahr, S. *et al.* DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res* **61**, 1659-1665 (2001).

246 Lo, Y. M. D. *et al.* Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus. *Sci Transl Med* **2**, 61ra91-61ra91 (2010). https://doi.org/doi:10.1126/scitranslmed.3001720

247 Diaz, L. A., Jr. & Bardelli, A. Liquid biopsies: genotyping circulating tumor DNA. *J Clin Oncol* **32**, 579-586 (2014). https://doi.org/10.1200/JCO.2012.45.2011

248 Alix-Panabières, C. & Pantel, K. Liquid biopsy: from discovery to clinical implementation. *Mol Oncol* **15**, 1617-1621 (2021). https://doi.org/10.1002/1878-0261.12997

249 Thierry, A. R., El Messaoudi, S., Gahan, P. B., Anker, P. & Stroun, M. Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Rev* **35**, 347-376 (2016). https://doi.org/10.1007/s10555-016-9629-x

250 Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* **6**, 224ra224 (2014). https://doi.org/10.1126/scitranslmed.3007094

251 Beiter, T., Fragasso, A., Hudemann, J., Niess, A. M. & Simon, P. Short-term treadmill running as a model for studying cell-free DNA kinetics in vivo. *Clin Chem* **57**, 633-636 (2011). https://doi.org/10.1373/clinchem.2010.158030

252 Atamaniuk, J. *et al.* Increased concentrations of cell-free plasma DNA after exhaustive exercise. *Clin Chem* **50**, 1668-1670 (2004). https://doi.org/10.1373/clinchem.2004.034553

253 Yuwono, N. L., Warton, K. & Ford, C. E. The influence of biological and lifestyle factors on circulating cell-free DNA in blood plasma. *eLife* **10**, e69679 (2021). https://doi.org/10.7554/eLife.69679

254 Angeles, A. K. *et al.* Liquid Biopsies beyond Mutation Calling: Genomic and Epigenomic Features of Cell-Free DNA in Cancer. *Cancers (Basel)* **13**, 5615 (2021). https://doi.org/10.3390/cancers13225615

255 Diaz, L. A., Jr. *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537-540 (2012). https://doi.org/10.1038/nature11219

256 Thress, K. S. *et al.* Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat Med* **21**, 560-562 (2015). https://doi.org/10.1038/nm.3854

257 Russo, M. *et al.* Tumor Heterogeneity and Lesion-Specific Response to Targeted Therapy in Colorectal Cancer. *Cancer Discov* **6**, 147-153 (2016). https://doi.org/10.1158/2159-8290.Cd-15-1283

258 Chaudhuri, A. A. *et al.* Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling. *Cancer Discov* **7**, 1394-1403 (2017). https://doi.org/10.1158/2159-8290.Cd-17-0716

259 Angeles, A. K. *et al.* Early identification of disease progression in ALK-rearranged lung cancer using circulating tumor DNA analysis. *NPJ Precis Oncol* **5**, 100 (2021). https://doi.org/10.1038/s41698-021-00239-3

260 Ashley, C. W. *et al.* High-Sensitivity Mutation Analysis of Cell-Free DNA for Disease Monitoring in Endometrial Cancer. *Clin Cancer Res* **29**, 410-421 (2023). https://doi.org/10.1158/1078-0432.Ccr-22-1134

261 Khier, S. & Lohan, L. Kinetics of circulating cell-free DNA for biomedical applications: critical appraisal of the literature. *Future Sci OA* **4**, Fso295 (2018). https://doi.org/10.4155/fsoa-2017-0140

262 García-Pardo, M., Makarem, M., Li, J. J. N., Kelly, D. & Leighl, N. B. Integrating circulating-free DNA (cfDNA) analysis into clinical practice: opportunities and challenges. *Br J Cancer* **127**, 592-602 (2022). https://doi.org/10.1038/s41416-022-01776-9

263 Hennigan, S. T. *et al.* Low Abundance of Circulating Tumor DNA in Localized Prostate Cancer. *JCO Precis Oncol* **3**, 1-13 (2019). https://doi.org/10.1200/po.19.00176

264 Ponti, G., Maccaferri, M., Percesepe, A., Tomasi, A. & Ozben, T. Liquid biopsy with cell free DNA: new horizons for prostate cancer. *Crit Rev Clin Lab Sci* **58**, 60-76 (2021). https://doi.org/10.1080/10408363.2020.1803789

265 Liu, H., Gao, Y., Vafaei, S., Gu, X. & Zhong, X. The Prognostic Value of Plasma Cell-Free DNA Concentration in the Prostate Cancer: A Systematic Review and Meta-Analysis. *Front Oncol* **11**, 599602 (2021). https://doi.org/10.3389/fonc.2021.599602

266 Schwarzenbach, H. *et al.* Cell-free Tumor DNA in Blood Plasma As a Marker for Circulating Tumor Cells in Prostate Cancer. *Clin Cancer Res* **15**, 1032-1038 (2009). https://doi.org/10.1158/1078-0432.CCR-08-1910

267 Souza, A. G. *et al.* Cell-free DNA promotes malignant transformation in non-tumor cells. *Sci Rep* **10**, 21674 (2020). https://doi.org/10.1038/s41598-020-78766-5

268 Feng, J. *et al.* Plasma cell-free DNA and its DNA integrity as biomarker to distinguish prostate cancer from benign prostatic hyperplasia in patients with increased serum prostate-specific antigen. *Int Urol Nephrol* **45**, 1023-1028 (2013). https://doi.org/10.1007/s11255-013-0491-2

269 Ellinger, J. *et al.* Noncancerous PTGS2 DNA fragments of apoptotic origin in sera of prostate cancer patients qualify as diagnostic and prognostic indicators. *Int J Cancer* **122**, 138-143 (2008). https://doi.org/10.1002/ijc.23057

270 Cheng, T. H. T. *et al.* Noninvasive Detection of Bladder Cancer by Shallow-Depth Genome-Wide Bisulfite Sequencing of Urinary Cell-Free DNA for Methylation and Copy Number Profiling. *Clin Chem* **65**, 927-936 (2019). https://doi.org/10.1373/clinchem.2018.301341

271 Husain, H. *et al.* Monitoring Daily Dynamics of Early Tumor Response to Targeted Therapy by Detecting Circulating Tumor DNA in Urine. *Clin Cancer Res* **23**, 4716-4723 (2017). https://doi.org/10.1158/1078-0432.Ccr-17-0454

272 Stejskal, P. *et al.* Circulating tumor nucleic acids: biology, release mechanisms, and clinical relevance. *Mol Cancer* **22**, 15 (2023). https://doi.org/10.1186/s12943-022-01710-w

273 Heitzer, E., Auinger, L. & Speicher, M. R. Cell-Free DNA and Apoptosis: How Dead Cells Inform About the Living. *Trends Mol Med* **26**, 519-528 (2020). https://doi.org/10.1016/j.molmed.2020.01.012

274 Mouliere, F. *et al.* Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med* **10**, eaat4921 (2018). https://doi.org/10.1126/scitranslmed.aat4921

275 Mouliere, F. *et al.* High fragmentation characterizes tumour-derived circulating DNA. *PLoS One* **6**, e23418 (2011). https://doi.org/10.1371/journal.pone.0023418

276 Underhill, H. R. *et al.* Fragment Length of Circulating Tumor DNA. *PLoS Genet* **12**, e1006162 (2016). https://doi.org/10.1371/journal.pgen.1006162

277 Su, Y. H. *et al.* Human urine contains small, 150 to 250 nucleotide-sized, soluble DNA derived from the circulation and may be useful in the detection of colorectal cancer. *J Mol Diagn* **6**, 101-107 (2004). https://doi.org/10.1016/S1525-1578(10)60497-7

278 Oshi, M. *et al.* Urine as a Source of Liquid Biopsy for Cancer. *Cancers (Basel)* **13**, 2652 (2021). https://doi.org/10.3390/cancers13112652

279 Dermody, S. M., Bhambhani, C., Swiecicki, P. L., Brenner, J. C. & Tewari, M. Trans-Renal Cell-Free Tumor DNA for Urine-Based Liquid Biopsy of Cancer. *Front Genet* **13**, 879108 (2022). https://doi.org/10.3389/fgene.2022.879108

280 Cheng, T. H. T. *et al.* Genomewide bisulfite sequencing reveals the origin and time-dependent fragmentation of urinary cfDNA. *Clin Biochem* **50**, 496-501 (2017). https://doi.org/10.1016/j.clinbiochem.2017.02.017

281 Eskra, J. N., Rabizadeh, D., Pavlovich, C. P., Catalona, W. J. & Luo, J. Approaches to urinary detection of prostate cancer. *Prostate Cancer Prostatic Dis* **22**, 362-381 (2019). https://doi.org/10.1038/s41391-019-0127-4

282 Markus, H. *et al.* Analysis of recurrently protected genomic regions in cell-free DNA found in urine. *Sci Transl Med* **13**, eaaz3088 (2021). https://doi.org/10.1126/scitranslmed.aaz3088

283 Chaudhuri, A. A. *et al.* Emerging Roles of Urine-Based Tumor DNA Analysis in Bladder Cancer Management. *JCO Precis Oncol* **4** (2020). https://doi.org/10.1200/po.20.00060

284 Loyfer, N. *et al.* A DNA methylation atlas of normal human cell types. *Nature* **613**, 355-364 (2023). https://doi.org/10.1038/s41586-022-05580-6

285 Silva, R. *et al.* Evaluating liquid biopsies for methylomic profiling of prostate cancer. *Epigenetics* **15**, 715-727 (2020). https://doi.org/10.1080/15592294.2020.1712876

286 Adalsteinsson, V. A. *et al.* Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* **8**, 1324 (2017). https://doi.org/10.1038/s41467-017-00965-y

287 Lipson, E. J. *et al.* Circulating tumor DNA analysis as a real-time method for monitoring tumor burden in melanoma patients undergoing treatment with immune checkpoint blockade. *J Immunother Cancer* **2**, 42 (2014). https://doi.org/10.1186/s40425-014-0042-0

288 van der Pol, Y. & Mouliere, F. Toward the Early Detection of Cancer by Decoding the Epigenetic and Environmental Fingerprints of Cell-Free DNA. *Cancer Cell* **36**, 350-368 (2019). https://doi.org/10.1016/j.ccell.2019.09.003

289 Hellmann, M. D. *et al.* Circulating Tumor DNA Analysis to Assess Risk of Progression after Long-term Response to PD-(L)1 Blockade in NSCLC. *Clin Cancer Res* **26**, 2849-2858 (2020). https://doi.org/10.1158/1078-0432.Ccr-19-3418

290 Newman, A. M. *et al.* Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat Biotechnol* **34**, 547-555 (2016). https://doi.org/10.1038/nbt.3520

291 Kurtz, D. M. *et al.* Enhanced detection of minimal residual disease by targeted sequencing of phased variants in circulating tumor DNA. *Nat Biotechnol* **39**, 1537-1547 (2021). https://doi.org/10.1038/s41587-021-00981-w

292 Widman, A. J. *et al.* Ultrasensitive plasma-based monitoring of tumor burden using machine-learning-guided signal enrichment. *Nat Med* **30**, 1655-1666 (2024). https://doi.org/10.1038/s41591-024-03040-4

293 Cohen, J. D. *et al.* Detection of low-frequency DNA variants by targeted sequencing of the Watson and Crick strands. *Nat Biotechnol* **39**, 1220-1227 (2021). https://doi.org/10.1038/s41587-021-00900-z

294 Gydush, G. *et al.* Massively parallel enrichment of low-frequency alleles enables duplex sequencing at low depth. *Nat Biomed Eng* **6**, 257-266 (2022). https://doi.org/10.1038/s41551-022-00855-9

295 Black, J. R. M. *et al.* Ultrasensitive ctDNA detection for preoperative disease stratification in early-stage lung adenocarcinoma. *Nat Med* **31**, 70-76 (2025). https://doi.org/10.1038/s41591-024-03216-y

296 Elliott, M. J. *et al.* Ultrasensitive Detection and Monitoring of Circulating Tumor DNA Using Structural Variants in Early-Stage Breast Cancer. *Clin Cancer Res* **31**, 1520-1532 (2025). https://doi.org/10.1158/1078-0432.Ccr-24-3472

297 Wan, J. C. M. *et al.* ctDNA monitoring using patient-specific sequencing and integration of variant reads. *Sci Transl Med* **12**, eaaz8084 (2020). https://doi.org/10.1126/scitranslmed.aaz8084

298 Elazezy, M. & Joosse, S. A. Techniques of using circulating tumor DNA as a liquid biopsy component in cancer management. *Comput Struct Biotechnol J* **16**, 370-378 (2018). https://doi.org/10.1016/j.csbj.2018.10.002

299 McDonald, B. R. *et al.* Personalized circulating tumor DNA analysis to detect residual disease after neoadjuvant therapy in breast cancer. *Sci Transl Med* **11**, eaax7392 (2019). https://doi.org/10.1126/scitranslmed.aax7392

300 Christensen, E. *et al.* Early Detection of Metastatic Relapse and Monitoring of Therapeutic Efficacy by Ultra-Deep Sequencing of Plasma Cell-Free DNA in Patients With Urothelial Bladder Carcinoma. *J Clin Oncol* **37**, 1547-1557 (2019). https://doi.org/10.1200/jco.18.02052

301 Janke, F. *et al.* Early circulating tumor DNA changes predict outcomes in head and neck cancer patients under re-radiotherapy. *Int J Cancer* **156**, 853-864 (2024). https://doi.org/10.1002/ijc.35152

302 Wan, J. C. M. *et al.* Genome-wide mutational signatures in low-coverage whole genome sequencing of cell-free DNA. *Nat Commun* **13**, 4953 (2022). https://doi.org/10.1038/s41467-022-32598-1

303 Bruhm, D. C. *et al.* Single-molecule genome-wide mutation profiles of cell-free DNA for non-invasive detection of cancer. *Nat Genet* **55**, 1301-1310 (2023). https://doi.org/10.1038/s41588-023-01446-3

304 Hollizeck, S. *et al.* Unravelling mutational signatures with plasma circulating tumour DNA. *Nat Commun* **15**, 9876 (2024). https://doi.org/10.1038/s41467-024-54193-2

305 Heitzer, E. *et al.* Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med* **5**, 30 (2013). https://doi.org/10.1186/gm434

306 Wyatt, A. W. *et al.* Concordance of Circulating Tumor DNA and Matched Metastatic Tissue Biopsy in Prostate Cancer. *J Natl Cancer Inst* **109**, djx118 (2017). https://doi.org/10.1093/jnci/djx118

307 Kohli, M. *et al.* Clinical and genomic insights into circulating tumor DNA-based alterations across the spectrum of metastatic hormone-sensitive and castrate-resistant prostate cancer. *EBioMedicine* **54**, 102728 (2020). https://doi.org/10.1016/j.ebiom.2020.102728

308 Espiritu, S. M. G. *et al.* The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical Aggression. *Cell* **173**, 1003-1013.e1015 (2018). https://doi.org/10.1016/j.cell.2018.03.029

309 Romanel, A. *et al.* Plasma AR and abiraterone-resistant prostate cancer. *Sci Transl Med* **7**, 312re310 (2015). https://doi.org/10.1126/scitranslmed.aac9511

310 Lau, E. *et al.* Detection of ctDNA in plasma of patients with clinically localised prostate cancer is associated with rapid disease progression. *Genome Med* **12**, 72 (2020). https://doi.org/10.1186/s13073-020-00770-1

311 Loyfer, N. *et al.* A human DNA methylation atlas reveals principles of cell type-specific methylation and identifies thousands of cell type-specific regulatory elements. *bioRxiv*, 2022.2001.2024.477547 (2022). https://doi.org/10.1101/2022.01.24.477547

312 Shen, S. Y. *et al.* Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* **563**, 579-583 (2018). https://doi.org/10.1038/s41586-018-0703-0

313 Song, C.-X. *et al.* 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res* **27**, 1231-1242 (2017). https://doi.org/10.1038/cr.2017.106

314 Li, W. *et al.* 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. *Cell Research* **27**, 1243-1257 (2017). https://doi.org/10.1038/cr.2017.121

315 Guler, G. D. *et al.* Plasma cell-free DNA hydroxymethylation profiling reveals anti-PD-1 treatment response and resistance biology in non-small cell lung cancer. *J Immunother Cancer* **12**, e008028 (2024). https://doi.org/10.1136/jitc-2023-008028

316 Cristiano, S. *et al.* Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385-389 (2019). https://doi.org/10.1038/s41586-019-1272-6

317 Jiang, P. *et al.* Plasma DNA End-Motif Profiling as a Fragmentomic Marker in Cancer, Pregnancy, and Transplantation. *Cancer Discov* **10**, 664-673 (2020). https://doi.org/10.1158/2159-8290.CD-19-0622

318 Jiang, P. *et al.* Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. *Proc Natl Acad Sci U S A* **115**, E10925-E10933 (2018). https://doi.org/10.1073/pnas.1814616115

319 Jiang, P. *et al.* Detection and characterization of jagged ends of double-stranded DNA in plasma. *Genome Res* **30**, 1144-1153 (2020). https://doi.org/10.1101/gr.261396.120

320 Snyder, Matthew W., Kircher, M., Hill, Andrew J., Daza, Riza M. & Shendure, J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164**, 57-68 (2016). https://doi.org/10.1016/j.cell.2015.11.050

321 Ulz, P. *et al.* Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat Genet* **48**, 1273-1278 (2016). https://doi.org/10.1038/ng.3648

322 Sun, K. *et al.* Size-tagged preferred ends in maternal plasma DNA shed light on the production mechanism and show utility in noninvasive prenatal testing. *Proc Natl Acad Sci U S A* **115**, E5106-E5114 (2018). https://doi.org/10.1073/pnas.1804134115

323 Jiang, P. *et al.* Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci U S A* **112**, E1317-1325 (2015). https://doi.org/10.1073/pnas.1500076112

324 Hellwig, S. *et al.* Automated size selection for short cell-free DNA fragments enriches for circulating tumor DNA and improves error correction during next generation sequencing. *PLoS One* **13**, e0197333 (2018). https://doi.org/10.1371/journal.pone.0197333

325 Serpas, L. *et al.* Dnase1l3 deletion causes aberrations in length and end-motif frequencies in plasma DNA. *Proc Natl Acad Sci U S A* **116**, 641-649 (2019). https://doi.org/10.1073/pnas.1815031116

326 Chandrananda, D., Thorne, N. P. & Bahlo, M. High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA. *BMC Med Genomics* **8**, 29 (2015). https://doi.org/10.1186/s12920-015-0107-z

327 Han, D. S. C. *et al.* The Biology of Cell-free DNA Fragmentation and the Roles of DNASE1, DNASE1L3, and DFFB. *Am J Hum Genet* **106**, 202-214 (2020). https://doi.org/10.1016/j.ajhg.2020.01.008

328 Zhou, Z. *et al.* Jagged Ends of Urinary Cell-Free DNA: Characterization and Feasibility Assessment in Bladder Cancer Detection. *Clin Chem* **67**, 621-630 (2021). https://doi.org/10.1093/clinchem/hvaa325

329 Doebley, A.-L. *et al.* A framework for clinical cancer subtyping from nucleosome profiling of cell-free DNA. *Nat Commun* **13**, 7475 (2022). https://doi.org/10.1038/s41467-022-35076-w

330 Sun, K. *et al.* Orientation-aware plasma cell-free DNA fragmentation analysis in open chromatin regions informs tissue of origin. *Genome Res* **29**, 418-427 (2019). https://doi.org/10.1101/gr.242719.118

331 Huang, J. & Wang, L. Cell-Free DNA Methylation Profiling Analysis-Technologies and Bioinformatics. *Cancers (Basel)* **11**, 1741 (2019). https://doi.org/10.3390/cancers11111741

332 Galardi, F. *et al.* Cell-Free DNA-Methylation-Based Methods and Applications in Oncology. *Biomolecules* **10**, 1677 (2020). https://doi.org/10.3390/biom10121677

333 Shen, S. Y., Burgener, J. M., Bratman, S. V. & De Carvalho, D. D. Preparation of cfMeDIP-seq libraries for methylome profiling of plasma cell-free DNA. *Nat Protoc* **14**, 2749-2780 (2019). https://doi.org/10.1038/s41596-019-0202-2

334 Leal, A., Sidransky, D. & Brait, M. Tissue and Cell-Free DNA-Based Epigenomic Approaches for Cancer Detection. *Clin Chem* **66**, 105-116 (2020). https://doi.org/10.1373/clinchem.2019.303594

335 Liu, Y. *et al.* Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat Biotechnol* **37**, 424-429 (2019). https://doi.org/10.1038/s41587-019-0041-2

336 Vaisvila, R. *et al.* Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Res* **31**, 1280-1289 (2021). https://doi.org/10.1101/gr.266551.120

337 Bie, F. *et al.* Multimodal analysis of cell-free DNA whole-methylome sequencing for cancer detection and localization. *Nat Commun* **14**, 6042 (2023). https://doi.org/10.1038/s41467-023-41774-w

338 Taiwo, O. *et al.* Methylome analysis using MeDIP-seq with low DNA concentrations. *Nat Protoc* **7**, 617-636 (2012). https://doi.org/10.1038/nprot.2012.012

339 Brinkman, A. B. *et al.* Whole-genome DNA methylation profiling using MethylCap-seq. *Methods* **52**, 232-236 (2010). https://doi.org/10.1016/j.ymeth.2010.06.012

340 Aberg, K. A., Chan, R. F., Xie, L., Shabalin, A. A. & van den Oord, E. Methyl-CpG-Binding Domain Sequencing: MBD-seq. *Methods Mol Biol* **1708**, 171-189 (2018). https://doi.org/10.1007/978-1-4939-7481-8_10

341 Huang, J., Soupir, A. C. & Wang, L. Cell-free DNA methylome profiling by MBD-seq with ultra-low input. *Epigenetics* **17**, 239-252 (2022). https://doi.org/10.1080/15592294.2021.1896984

342 Conway, A.-M. *et al.* A cfDNA methylation-based tissue-of-origin classifier for cancers of unknown primary. *Nat Commun* **15**, 3292 (2024). https://doi.org/10.1038/s41467-024-47195-7

343 Chemi, F. *et al.* cfDNA methylome profiling for detection and subtyping of small cell lung cancers. *Nat Cancer* **3**, 1260-1270 (2022). https://doi.org/10.1038/s43018-022-00415-9

344 Nuzzo, P. V. *et al.* Detection of renal cell carcinoma using plasma and urine cell-free DNA methylomes. *Nat Med* **26**, 1041-1043 (2020). https://doi.org/10.1038/s41591-020-0933-1

345 Nassiri, F. *et al.* Detection and discrimination of intracranial tumors using plasma cell-free DNA methylomes. *Nat Med* **26**, 1044-1047 (2020). https://doi.org/10.1038/s41591-020-0932-2

346 Burgener, J. M. *et al.* Tumor-Naive Multimodal Profiling of Circulating Tumor DNA in Head and Neck Squamous Cell Carcinoma. *Clin Cancer Res* **27**, 4230-4244 (2021). https://doi.org/10.1158/1078-0432.CCR-21-0110

347 Bastian, P. J. *et al.* Preoperative serum DNA GSTP1 CpG island hypermethylation and the risk of early prostate-specific antigen recurrence following radical prostatectomy. *Clin Cancer Res* **11**, 4037-4043 (2005). https://doi.org/10.1158/1078-0432.Ccr-04-2446

348 O'Reilly, E. *et al.* epiCaPture: A Urine DNA Methylation Test for Early Detection of Aggressive Prostate Cancer. *JCO Precis Oncol* **2019**, PO.18.00134 (2019). https://doi.org/10.1200/po.18.00134

349 Zhao, F. *et al.* Urinary DNA Methylation Biomarkers for Noninvasive Prediction of Aggressive Disease in Patients with Prostate Cancer on Active Surveillance. *J Urol* **197**, 335-341 (2017). https://doi.org/10.1016/j.juro.2016.08.081

350 Cairns, P. *et al.* Molecular detection of prostate cancer in urine by GSTP1 hypermethylation. *Clin Cancer Res* **7**, 2727-2730 (2001).

351 Hendriks, R. J. *et al.* Epigenetic markers in circulating cell-free DNA as prognostic markers for survival of castration-resistant prostate cancer patients. *Prostate* **78**, 336-342 (2018). https://doi.org/10.1002/pros.23477

352 Klein, E. A. *et al.* Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann Oncol* **32**, 1167-1177 (2021). https://doi.org/10.1016/j.annonc.2021.05.806

353 Liu, M. C., Oxnard, G. R., Klein, E. A., Swanton, C. & Seiden, M. V. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol* **31**, 745-759 (2020). https://doi.org/10.1016/j.annonc.2020.02.011

354 Chen, S. *et al.* The cell-free DNA methylome captures distinctions between localized and metastatic prostate tumors. *Nat Commun* **13**, 6467 (2022). https://doi.org/10.1038/s41467-022-34012-2

355 Lleshi, E. *et al.* Prostate cancer detection through unbiased capture of methylated cell-free DNA. *iScience* **27**, 110330 (2024). https://doi.org/10.1016/j.isci.2024.110330

356 Elicit: The AI Research Assistant (2024). https://elicit.com (accessed 25 November 2024)

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

358 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012). https://doi.org/10.1038/nmeth.1923

359 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, 10-12 (2011). https://doi.org/10.14806/ej.17.1.200

360 Andrews, S. FastQC: a quality control tool for high throughput sequence data. (Babraham Bioinformatics, Babraham Institute, Cambridge, UK, 2010). https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

361 Lai, D., Ha, G., Shah, S. HMMcopy: Copy number prediction with correction for GC and mappability bias for HTS data. R package version 1.32.0. (2021). https://doi.org/10.18129/B9.bioc.HMMcopy

362 Lai, D. HMM Copy Utils. GitHub repository (Vancouver, Canada, 2011). https://github.com/shahcompbio/hmmcopy_utils

363 Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048 (2016). https://doi.org/10.1093/bioinformatics/btw354

364 Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat Biotechnol* **35**, 316-319 (2017). https://doi.org/10.1038/nbt.3820

365 Broad Institute. *Picard Toolkit.* GitHub repository (Cambridge, MA, USA, 2019). https://broadinstitute.github.io/picard

366 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009). https://doi.org/10.1093/bioinformatics/btp352

367 Kurtzer, G. M., Sochat, V. & Bauer, M. W. . Singularity: Scientific containers for mobility of compute. *PLoS ONE* **12**, e0177459 (2017). https://doi.org/10.1371/journal.pone.0177459

368 R Core Team. *R: A Language and Environment for Statistical Computing.* (R Foundation for Statistical Computing, Vienna, Austria, 2020)

369 Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**, 491-499 (2017). https://doi.org/10.1101/gr.209601.116

370 Davison, A. C. H., David Victor *Bootstrap Methods and Their Applications*. Vol. 1 (Cambridge University Press, 1997).

371 Zhu, L. J. *et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010). https://doi.org/10.1186/1471-2105-11-237

372 Dowle, M. & Srinivasan, A. data.table: Extension of data.frame. R package version 1.14.4 (2022). https://doi.org/10.32614/CRAN.package.data.table

373 Wickham, H., François, R., Henry, L., Müller, K. & Vaughan, D. dplyr: A grammar of data manipulation. R package version 1.1.1 (2023). https://doi.org/10.32614/CRAN.package.dplyr

374 Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol* **9**, e1003118 (2013). https://doi.org/10.1371/journal.pcbi.1003118

375 Kassambara, A. ggcorrplot: Visualization of a correlation matrix using ggplot2. R package version 0.1.4.1 (2023). http://www.sthda.com/english/wiki/ggcorrplot-visualization-of-a-correlation-matrix-using-ggplot2

376 Tang, Y. a. H., Masaaki and Li, Wenxuan. ggfortify: Unified Interface to Visualize Statistical Results of Popular R Packages. *The R Journal* **8**, 474-485 (2016). https://doi.org/ 10.32614/CRAN.package.ggfortify

377 Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (Springer-Verlag, New York, 2016). https://doi.org/10.32614/CRAN.package.ggplot2

378 Kassambara, A. ggpubr: 'ggplot2' based publication ready plots. R package version 0.6.0 (2023). https://doi.org/10.32614/CRAN.package.ggpubr

379 Mohammadi, R. & Burke, K. liver: Eating the liver of data science. R package version 1.15 (2023). https://doi.org/10.32614/CRAN.package.liver

380 Zeileis, A. H., Torsten. Diagnostic Checking in Regression Relationships. *R News* **2**, 7–10 (2002). https://doi.org/10.32614/CRAN.package.lmtest

381 Bengtsson, H. matrixStats: Functions that apply to rows and columns of matrices (and to vectors). R package version 0.61.0 (2021). https://doi.org/10.32614/CRAN.package.matrixStats

382 Lienhard, M., Grimm, C., Morkel, M., Herwig, R. & Chavez, L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics* **30**, 284-286 (2014). https://doi.org/10.1093/bioinformatics/btt650

383 Yan, Y. MLmetrics: Machine learning evaluation metrics. R package version 1.1.3 (2024). https://doi.org/10.32614/CRAN.package.MLmetrics

384 Kolde, R. pheatmap: Pretty Heatmaps. R package version 1.0.12 (2018). https://doi.org/10.32614/CRAN.package.pheatmap

385 Sievert, C. Interactive Web-Based Data Visualization with R, plotly, and shiny. *Chapman and Hall/CRC* (2019). https://plotly-r.com

386 Lee, S., Cook, D. & Lawrence, M. plyranges: a grammar of genomic data transformation. *Genome Biol* **20**, 4 (2019). https://doi.org/10.1186/s13059-018-1597-8

387 Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011). https://doi.org/10.1186/1471-2105-12-77

388 Grau, J., Grosse, I. & Keilwagen, J. PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R. *Bioinformatics* **31**, 2595-2597 (2015). https://doi.org/10.1093/bioinformatics/btv153

389 Lienhard, M. *et al.* QSEA-modelling of genome-wide DNA methylation from sequencing enrichment experiments. *Nucleic Acids Res* **45**, e44 (2017). https://doi.org/10.1093/nar/gkw1193

390 Neuwirth, E. RColorBrewer: ColorBrewer palettes. R package version 1.1-3 (2022). https://doi.org/10.32614/CRAN.package.RColorBrewer

391 Wickham, H., Hester, J. & Bryan, J. readr: Read rectangular text data. R package version 2.1.2 (2022). https://doi.org/10.32614/CRAN.package.readr

392 Wickham, H. & Bryan, J. readxl: Read Excel files. R package version 1.3.1 (2019). https://doi.org/10.32614/CRAN.package.readxl

393 Gel, B. *et al.* regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**, 289-291 (2015). https://doi.org/10.1093/bioinformatics/btv562

394 Kassambara, A. rstatix: Pipe-friendly framework for basic statistical tests. R package version 0.7.2 (2023). https://doi.org/10.32614/CRAN.package.rstatix

395 Wickham, H. *et al.* Welcome to the tidyverse. *J Open Source Softw* **4**, 1686 (2019). https://doi.org/10.21105/joss.01686

396 Zeileis, A. & Grothendieck, G. zoo: S3 Infrastructure for Regular and Irregular Time Series. *J Stat Softw* **14**, 1 - 27 (2005). https://doi.org/10.18637/jss.v014.i06

397 Görtz, M. *et al.* The Value of Prostate-specific Antigen Density for Prostate Imaging-Reporting and Data System 3 Lesions on Multiparametric Magnetic Resonance Imaging: A Strategy to Avoid Unnecessary Prostate Biopsies. *Eur Urol Focus* **7**, 325-331 (2021). https://doi.org/10.1016/j.euf.2019.11.012

398 Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* **38**, 276-278 (2020). https://doi.org/10.1038/s41587-020-0439-x

399 Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354 (2019). https://doi.org/10.1038/s41598-019-45839-z

400 Chen, Z. *et al.* Chromosomal instability of circulating tumor DNA reflect therapeutic responses in advanced gastric cancer. *Cell Death Dis* **10**, 697 (2019). https://doi.org/10.1038/s41419-019-1907-4

401 Liu, H. *et al.* MALBAC-based chromosomal imbalance analysis: a novel technique enabling effective non-invasive diagnosis and monitoring of bladder cancer. *BMC Cancer* **18**, 659 (2018). https://doi.org/10.1186/s12885-018-4571-7

402 Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25 (2010). https://doi.org/10.1186/gb-2010-11-3-r25

403 Börno, S. T. *et al.* Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation. *Cancer Discov* **2**, 1024-1035 (2012). https://doi.org/10.1158/2159-8290.CD-12-0041

404 Fraser, M. *et al.* Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541**, 359-364 (2017). https://doi.org/10.1038/nature20788

405 Zhao, S. G. *et al.* The DNA methylation landscape of advanced prostate cancer. *Nat Genet* **52**, 778-789 (2020). https://doi.org/10.1038/s41588-020-0648-8

406 Annala, M. *et al.* Circulating Tumor DNA Genomics Correlate with Resistance to Abiraterone and Enzalutamide in Prostate Cancer. *Cancer Discov* **8**, 444-457 (2018). https://doi.org/10.1158/2159-8290.Cd-17-0937

407 Weir, I. *Spearman's Rank Correlation- Introduction*, <https://statstutor.ac.uk/resources/uploaded/spearmans.pdf> (

408 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).

409 Kung, J. Y. Elicit. *J Can Health Libr Assoc* **44**, 15-18 (2023).

410 Riediger, A. L. *et al.* Integrative genomic and epigenomic profiling in plasma and urinary cell-free DNA improves early risk stratification of newly diagnosed prostate cancer. *medRxiv*, 2025.2004.2013.25325732 (2025). https://doi.org/10.1101/2025.04.13.25325732 (under revision)

411 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001). https://doi.org/10.1038/35057062

412 Vandekerkhove, G. *et al.* Circulating Tumor DNA Abundance and Potential Utility in De Novo Metastatic Prostate Cancer. *Eur Urol* **75**, 667-675 (2019). https://doi.org/10.1016/j.eururo.2018.12.042

413 Zhu, X. Y. *et al.* Risk-Based Prostate Cancer Screening. *Eur Urol* **61**, 652-661 (2012). https://doi.org/10.1016/j.eururo.2011.11.029

414 Fleshner, N. E. & Lawrentschuk, N. Risk of Developing Prostate Cancer in the Future: Overview of Prognostic Biomarkers. *Urology* **73**, 21-27 (2009). https://doi.org/10.1016/j.urology.2009.02.022

415 Schröder, F. H. *et al.* Screening and Prostate-Cancer Mortality in a Randomized European Study. *N Engl J Med* **360**, 1320-1328 (2009). https://doi.org/10.1056/NEJMoa0810084

416 Van Poppel, H. *et al.* Prostate-specific Antigen Testing as Part of a Risk-Adapted Early Detection Strategy for Prostate Cancer: European Association of Urology Position and Recommendations for 2021. *Eur Urol* **80**, 703-711 (2021). https://doi.org/10.1016/j.eururo.2021.07.024

417 Travis, R. C. *et al.* Serum androgens and prostate cancer among 643 cases and 643 controls in the European Prospective Investigation into Cancer and Nutrition. *Int J Cancer* **121**, 1331-1338 (2007). https://doi.org/10.1002/ijc.22814

418 Greytak, S. R. *et al.* Harmonizing Cell-Free DNA Collection and Processing Practices through Evidence-Based Guidance. *Clin Cancer Res* **26**, 3104-3109 (2020). https://doi.org/10.1158/1078-0432.CCR-19-3015

419 Bronkhorst, A. J., Aucamp, J. & Pretorius, P. J. Cell-free DNA: Preanalytical variables. *Clin Chim Acta* **450**, 243-253 (2015). https://doi.org/10.1016/j.cca.2015.08.028

420 Ntzifa, A. & Lianidou, E. Pre-analytical conditions and implementation of quality control steps in liquid biopsy analysis. *Crit Rev Clin Lab Sci* **60**, 573-594 (2023). https://doi.org/10.1080/10408363.2023.2230290

421 Augustus, E. *et al.* The art of obtaining a high yield of cell-free DNA from urine. *PLoS One* **15**, e0231058 (2020). https://doi.org/10.1371/journal.pone.0231058

422 Lee, E. Y., Lee, E. J., Yoon, H., Lee, D. H. & Kim, K. H. Comparison of Four Commercial Kits for Isolation of Urinary Cell-Free DNA and Sample Storage Conditions. *Diagnostics (Basel)* **10**, 234 (2020). https://doi.org/10.3390/diagnostics10040234

423 Streleckiene, G., Reid, H. M., Arnold, N., Bauerschlag, D. & Forster, M. Quantifying cell free DNA in urine: comparison between commercial kits, impact of gender and inter-individual variation. *Biotechniques* **64**, 225-230 (2018). https://doi.org/10.2144/btn-2018-0003

424 Pos, Z. *et al.* Technical and Methodological Aspects of Cell-Free Nucleic Acids Analyzes. *Int J Mol Sci* **21**, 8634 (2020). https://doi.org/10.3390/ijms21228634

425 Vorsters, A. *et al.* Optimization of HPV DNA detection in urine by improving collection, storage, and extraction. *Eur J Clin Microbiol Infect Dis* **33**, 2005-2014 (2014). https://doi.org/10.1007/s10096-014-2147-2

426 Brisuda, A. *et al.* Urinary Cell-Free DNA Quantification as Non-Invasive Biomarker in Patients with Bladder Cancer. *Urol Int* **96**, 25-31 (2016). https://doi.org/10.1159/000438828

427 Henao Diaz, E., Yachnin, J., Grönberg, H. & Lindberg, J. The In Vitro Stability of Circulating Tumour DNA. *PLoS One* **11**, e0168153 (2016). https://doi.org/10.1371/journal.pone.0168153

428 Chang, H. W. *et al.* Urinary cell-free DNA as a potential tumor marker for bladder cancer. *Int J Biol Markers* **22**, 287-294 (2007). https://doi.org/10.1177/172460080702200408

429 Diefenbach, R. J., Lee, J. H., Kefford, R. F. & Rizos, H. Evaluation of commercial kits for purification of circulating free DNA. *Cancer Genet* **228-229**, 21-27 (2018). https://doi.org/10.1016/j.cancergen.2018.08.005

430 Trigg, R. M., Martinson, L. J., Parpart-Li, S. & Shaw, J. A. Factors that influence quality and yield of circulating-free DNA: A systematic review of the methodology literature. *Heliyon* **4**, e00699 (2018). https://doi.org/10.1016/j.heliyon.2018.e00699

431 Hudecova, I. *et al.* Characteristics, origin, and potential for cancer diagnostics of ultrashort plasma cell-free DNA. *Genome Res* **32**, 215-227 (2021). https://doi.org/10.1101/gr.275691.121

432 Dietz, S. *et al.* Serial liquid biopsies for detection of treatment failure and profiling of resistance mechanisms in KLC1-ALK-rearranged lung cancer. *Cold Spring Harb Mol Case Stud* **5**, a004630 (2019). https://doi.org/10.1101/mcs.a004630

433 Chandrananda, D., Thorne, N. P. & Bahlo, M. High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA. *BMC Med Genomics* **8**, 29 (2015). https://doi.org/10.1186/s12920-015-0107-z

434 Ganesamoorthy, D. *et al.* Whole genome deep sequencing analysis of cell-free DNA in samples with low tumour content. *BMC Cancer* **22**, 85 (2022). https://doi.org/10.1186/s12885-021-09160-1

435 Stutheit-Zhao, E. Y. *et al.* Early Changes in Tumor-Naive Cell-Free Methylomes and Fragmentomes Predict Outcomes in Pembrolizumab-Treated Solid Tumors. *Cancer Discov* **14**, 1048-1063 (2024). https://doi.org/10.1158/2159-8290.Cd-23-1060

436 Zhitnyuk, Y. V. *et al.* Deep cfDNA fragment end profiling enables cancer detection. *Mol Cancer* **21**, 26 (2022). https://doi.org/10.1186/s12943-021-01491-8

437 Diehl, F. *et al.* Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A* **102**, 16368-16373 (2005). https://doi.org/10.1073/pnas.0507904102

438 Markus, H. *et al.* Refined characterization of circulating tumor DNA through biological feature integration. *Sci Rep* **12**, 1928 (2022). https://doi.org/10.1038/s41598-022-05606-z

439 Chen, E. *et al.* Cell-free DNA concentration and fragment size as a biomarker for prostate cancer. *Sci Rep* **11**, 5040 (2021). https://doi.org/10.1038/s41598-021-84507-z

440 Russo, I. *et al.* Toward Personalised Liquid Biopsies for Urothelial Carcinoma: Characterisation of ddPCR and Urinary cfDNA for the Detection of the TERT 228 G>A/T Mutation. *Bladder Cancer* **4**, 41-48 (2018). https://doi.org/10.3233/BLC-170152

441 Ponti, G. *et al.* Quick assessment of cell-free DNA in seminal fluid and fragment size for early non-invasive prostate cancer diagnosis. *Clin Chim Acta* **497**, 76-80 (2019). https://doi.org/10.1016/j.cca.2019.07.011

442 Kroeze, A. *et al.* Systemic Inflammation Induces Release of Cell-Free DNA from Hematopoietic and Parenchymal Cells in Mice and Human. *Blood* **132**, 1107-1107 (2018). https://doi.org/10.1182/blood-2018-99-114468

443 Lu, T. & Li, J. Clinical applications of urinary cell-free DNA in cancer: current insights and promising future. *Am J Cancer Res* **7**, 2318-2332 (2017).

444 Burnham, P. *et al.* Urinary cell-free DNA is a versatile analyte for monitoring infections of the urinary tract. *Nat Commun* **9**, 2412 (2018). https://doi.org/10.1038/s41467-018-04745-0

445 García Moreira, V., Prieto García, B., de la Cera Martínez, T. & Álvarez Menéndez, F. V. Elevated transrenal DNA (cell-free urine DNA) in patients with urinary tract infection compared to healthy controls. *Clin Biochem* **42**, 729-731 (2009). https://doi.org/10.1016/j.clinbiochem.2008.12.021

446 Zhou, Q. *et al.* Epigenetic analysis of cell-free DNA by fragmentomic profiling. *Proc Natl Acad Sci U S A* **119**, e2209852119 (2022). https://doi.org/10.1073/pnas.2209852119

447 Noë, M. *et al.* DNA methylation and gene expression as determinants of genome-wide cell-free DNA fragmentation. *Nat Commun* **15**, 6690 (2024). https://doi.org/10.1038/s41467-024-50850-8

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

449 Leary, R. J. *et al.* Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci Transl Med* **4**, 162ra154 (2012). https://doi.org/10.1126/scitranslmed.3004742

450 Greene, S. B. *et al.* Chromosomal Instability Estimation Based on Next Generation Sequencing and Single Cell Genome Wide Copy Number Variation Analysis. *PLoS One* **11**, e0165089 (2016). https://doi.org/10.1371/journal.pone.0165089

451 Schonhoft, J. D. *et al.* Morphology-Predicted Large-Scale Transition Number in Circulating Tumor Cells Identifies a Chromosomal Instability Biomarker Associated with Poor Outcome in Castration-Resistant Prostate Cancer. *Cancer Res* **80**, 4892-4903 (2020). https://doi.org/10.1158/0008-5472.Can-20-1216

452 Oza, V. H., Fisher, J. L., Darji, R. & Lasseigne, B. N. CINmetrics: an R package for analyzing copy number aberrations as a measure of chromosomal instability. *PeerJ* **11**, e15244 (2023). https://doi.org/10.7717/peerj.15244

453 Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non–Small-Cell Lung Cancer. *N Engl J Med* **376**, 2109-2121 (2017). https://doi.org/10.1056/NEJMoa1616288

454 Miller, E. T. *et al.* Chromosomal instability in untreated primary prostate cancer as an indicator of metastatic potential. *BMC Cancer* **20**, 398 (2020). https://doi.org/10.1186/s12885-020-06817-1

455 Sansregret, L., Vanhaesebroeck, B. & Swanton, C. Determinants and clinical implications of chromosomal instability in cancer. *Nat Rev Clin Oncol* **15**, 139-150 (2018). https://doi.org/10.1038/nrclinonc.2017.198

456 McGranahan, N., Burrell, R. A., Endesfelder, D., Novelli, M. R. & Swanton, C. Cancer chromosomal instability: therapeutic and diagnostic challenges. *EMBO Rep* **13**, 528-538 (2012). https://doi.org/10.1038/embor.2012.61

457 Taluri, S., Oza, V. H., Soelter, T. M., Fisher, J. L. & Lasseigne, B. N. Inferring chromosomal instability from copy number aberrations as a measure of chromosomal instability across human cancers. *bioRxiv* **2023.05.24.542174** (2023). https://doi.org/10.1101/2023.05.24.542174

458 Taluri, S., Oza, V. H., Soelter, T. M., Fisher, J. L. & Lasseigne, B. N. Inferring chromosomal instability from copy number aberrations as a measure of chromosomal instability across human cancers. *Cancer Reports* **6**, e1902 (2023). https://doi.org/10.1002/cnr2.1902

459 Kaplun, D. S., Kaluzhny, D. N., Prokhortchouk, E. B. & Zhenilo, S. V. DNA Methylation: Genomewide Distribution, Regulatory Mechanism and Therapy Target. *Acta Naturae* **14**, 4-19 (2022). https://doi.org/10.32607/actanaturae.11822

460 Nam, A. R. *et al.* Alternative methylation of intron motifs is associated with cancer-related gene expression in both canine mammary tumor and human breast cancer. *Clin Epigenetics* **12**, 110 (2020). https://doi.org/10.1186/s13148-020-00888-4

461 Anastasiadi, D., Esteve-Codina, A. & Piferrer, F. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin* **11**, 37 (2018). https://doi.org/10.1186/s13072-018-0205-1

462 Aran, D., Sabato, S. & Hellman, A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol* **14**, R21 (2013). https://doi.org/10.1186/gb-2013-14-3-r21

463 Akalin, A. *et al.* Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet* **8**, e1002781 (2012). https://doi.org/10.1371/journal.pgen.1002781

464 Hu, S. *et al.* Accurate detection of early-stage lung cancer using a panel of circulating cell-free DNA methylation biomarkers. *Biomark Res* **11**, 45 (2023). https://doi.org/10.1186/s40364-023-00486-5

465 Qi, J. *et al.* Plasma cell-free DNA methylome-based liquid biopsy for accurate gastric cancer detection. *Cancer Sci* **115**, 3426-3438 (2024). https://doi.org/10.1111/cas.16284

466 Pharoah, P. D. P. *et al.* GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat Genet* **45**, 362-370 (2013). https://doi.org/10.1038/ng.2564

467 Kar, S., Niharika, Roy, A. & Patra, S. K. Overexpression of SOX2 Gene by Histone Modifications: SOX2 Enhances Human Prostate and Breast Cancer Progression by Prevention of Apoptosis and Enhancing Cell Proliferation. *Oncology* **101**, 591-608 (2023). https://doi.org/10.1159/000531195

468 Wuebben, E. L. & Rizzino, A. The dark side of SOX2: cancer - a comprehensive overview. *Oncotarget* **8**, 44917-44943 (2017). https://doi.org/10.18632/oncotarget.16570

469 Liu, S., Xu, B. & Yan, D. Enhanced expression of long non-coding RNA Sox2ot promoted cell proliferation and motility in colorectal cancer. *Minerva Med* **107**, 279-286 (2016).

470 Saghaeian Jazi, M., Samaei, N. M., Ghanei, M., Shadmehr, M. B. & Mowla, S. J. Overexpression of the non-coding SOX2OT variants 4 and 7 in lung tumors suggests an oncogenic role in lung cancer. *Tumour Biol* **37**, 10329-10338 (2016). https://doi.org/10.1007/s13277-016-4901-9

471 Dehghani-Samani, M. *et al.* Correlations between Overexpression of SOX2OT Long Non-coding RNA and Susceptibility to Breast Cancer. *Comb Chem High Throughput Screen* **23**, 981-987 (2020). https://doi.org/10.2174/1386207323666200514075042

472 Smyth, L. J., McKay, G. J., Maxwell, A. P. & McKnight, A. J. DNA hypermethylation and DNA hypomethylation is present at different loci in chronic kidney disease. *Epigenetics* **9**, 366-376 (2014). https://doi.org/10.4161/epi.27161

473 Lee, S. The association of genetically controlled CpG methylation (cg158269415) of protein tyrosine phosphatase, receptor type N2 (PTPRN2) with childhood obesity. *Sci Rep* **9**, 4855 (2019). https://doi.org/10.1038/s41598-019-40486-w

474 Gentilini, D. *et al.* Epigenome-wide association study in hepatocellular carcinoma: Identification of stochastic epigenetic mutations through an innovative statistical approach. *Oncotarget* **8**, 41890-41902 (2017). https://doi.org/10.18632/oncotarget.17462

475 Chen, W. Y. *et al.* Inhibition of the androgen receptor induces a novel tumor promoter, ZBTB46, for prostate cancer metastasis. *Oncogene* **36**, 6213-6224 (2017). https://doi.org/10.1038/onc.2017.226

476 Chen, W. Y. *et al.* Androgen deprivation-induced ZBTB46-PTGS1 signaling promotes neuroendocrine differentiation of prostate cancer. *Cancer Lett* **440-441**, 35-46 (2019). https://doi.org/10.1016/j.canlet.2018.10.004

477 Olkhov-Mitsel, E. *et al.* Epigenome-Wide DNA Methylation Profiling Identifies Differential Methylation Biomarkers in High-Grade Bladder Cancer. *Transl Oncol* **10**, 168-177 (2017). https://doi.org/10.1016/j.tranon.2017.01.001

478 Peter, M. R. *et al.* Dynamics of the cell-free DNA methylome of metastatic prostate cancer during androgen-targeting treatment. *Epigenomics* **12**, 1317-1332 (2020). https://doi.org/10.2217/epi-2020-0173

479 Shui, I. M. *et al.* Prostate tumor DNA methylation is associated with cigarette smoking and adverse prostate cancer outcomes. *Cancer* **122**, 2168-2177 (2016). https://doi.org/10.1002/cncr.30045

480 Strizzi, L., Hardy, K. M., Kirschmann, D. A., Ahrlund-Richter, L. & Hendrix, M. J. C. Nodal Expression and Detection in Cancer: Experience and Challenges. *Cancer Res* **72**, 1915-1920 (2012). https://doi.org/10.1158/0008-5472.Can-11-3419

481 Wu, T. *et al.* Nodal promotes colorectal cancer survival and metastasis through regulating SCD1-mediated ferroptosis resistance. *Cell Death Dis* **14**, 229 (2023). https://doi.org/10.1038/s41419-023-05756-6

482 Lee, C. C. *et al.* Nodal promotes growth and invasion in human gliomas. *Oncogene* **29**, 3110-3123 (2010). https://doi.org/10.1038/onc.2010.55

483 Lawrence, M. G. *et al.* Reactivation of embryonic nodal signaling is associated with tumor progression and promotes the growth of prostate cancer cells. *Prostate* **71**, 1198-1209 (2011). https://doi.org/10.1002/pros.21335

484 Berchuck, J. E. *et al.* Detecting Neuroendocrine Prostate Cancer Through Tissue-Informed Cell-Free DNA Methylation Analysis. *Clin Cancer Res* **28**, 928-938 (2022). https://doi.org/10.1158/1078-0432.Ccr-21-3762

485 Nonn, L., Ananthanarayanan, V. & Gann, P. H. Evidence for field cancerization of the prostate. *Prostate* **69**, 1470-1479 (2009). https://doi.org/10.1002/pros.20983

486 Singhal, U. *et al.* Integrative multi-region molecular profiling of primary prostate cancer in men with synchronous lymph node metastasis. *Nat Commun* **15**, 4341 (2024). https://doi.org/10.1038/s41467-024-48629-y

487 Yang, B. *et al.* Validation of an epigenetic field of susceptibility to detect significant prostate cancer from non-tumor biopsies. *Clin Epigenetics* **11**, 168 (2019). https://doi.org/10.1186/s13148-019-0771-5

488 Truong, M. *et al.* Using the epigenetic field defect to detect prostate cancer in biopsy negative patients. *J Urol* **189**, 2335-2341 (2013). https://doi.org/10.1016/j.juro.2012.11.074

489 Lee, Y. *et al.* TMPRSS4 promotes cancer stem–like properties in prostate cancer cells through upregulation of SOX2 by SLUG and TWIST1. *J Exp Clin Cancer Res* **40**, 372 (2021). https://doi.org/10.1186/s13046-021-02147-7

490 Wang, J. *et al.* Cholinergic signaling via muscarinic M1 receptor confers resistance to docetaxel in prostate cancer. *Cell Rep Med* **5**, 20 (2024). https://doi.org/10.1016/j.xcrm.2023.101388

491 Yin, Q. Q., Xu, L. H., Zhang, M. & Xu, C. Muscarinic acetylcholine receptor M1 mediates prostate cancer cell migration and invasion through hedgehog signaling. *Asian J Androl* **20**, 608-614 (2018). https://doi.org/10.4103/aja.aja_55_18

492 Luo, Z. & Farnham, P. J. Genome-wide analysis of HOXC4 and HOXC6 regulated genes and binding sites in prostate cancer cells. *PLoS One* **15**, e0228590 (2020). https://doi.org/10.1371/journal.pone.0228590

493 Rhie, S. K. *et al.* Identification of activated enhancers and linked transcription factors in breast, prostate, and kidney tumors by tracing enhancer networks using epigenetic traits. *Epigenetics Chromatin* **9**, 50 (2016). https://doi.org/10.1186/s13072-016-0102-4

494 Geybels, M. S. *et al.* Epigenomic profiling of DNA methylation in paired prostate cancer versus adjacent benign tissue. *Prostate* **75**, 1941-1950 (2015). https://doi.org/10.1002/pros.23093

495 Reynolds, S. R., Zhang, Z., Salas, L. A. & Christensen, B. C. Tumor microenvironment deconvolution identifies cell-type-independent aberrant DNA methylation and gene expression in prostate cancer. *Clin Epigenetics* **16**, 5 (2024). https://doi.org/10.1186/s13148-023-01609-3

496 Gao, S. *et al.* TUBB4A interacts with MYH9 to protect the nucleus during cell migration and promotes prostate cancer via GSK3β/β-catenin signalling. *Nat Commun* **13**, 2792 (2022). https://doi.org/10.1038/s41467-022-30409-1

497 Naumov, V. A. *et al.* Genome-scale analysis of DNA methylation in colorectal cancer using Infinium HumanMethylation450 BeadChips. *Epigenetics* **8**, 921-934 (2013). https://doi.org/10.4161/epi.25577

498 Janssens, K. *et al.* Epigenome-wide methylation analysis of colorectal carcinoma, adenoma and normal tissue reveals novel biomarkers addressing unmet clinical needs. *Clin Epigenetics* **15**, 111 (2023). https://doi.org/10.1186/s13148-023-01516-7

499 Huang, H. *et al.* Expressions of glucose transporter genes are diversely attenuated and significantly associated with prostate cancer progression. *Am J Clin Exp Urol* **11**, 578-593 (2023).

500 Liu, Y. *et al.* Pan-cancer analysis of SLC2A family genes as prognostic biomarkers and therapeutic targets. *Heliyon* **10**, e29655 (2024). https://doi.org/10.1016/j.heliyon.2024.e29655

501 Heidegger, I. *et al.* Prediction of Clinically Significant Prostate Cancer by a Specific Collagen-related Transcriptome, Proteome, and Urinome Signature. *Eur Urol Oncol* **8**, 652-662 (2024). https://doi.org/10.1016/j.euo.2024.05.014

502 Zou, T. *et al.* A perspective profile of ADCY1 in cAMP signaling with drug-resistance in lung cancer. *J Cancer* **10**, 6848-6857 (2019). https://doi.org/10.7150/jca.36614

503 Munkley, J. *et al.* Androgen-regulation of the protein tyrosine phosphatase PTPRR activates ERK1/2 signalling in prostate cancer cells. *BMC Cancer* **15**, 9 (2015). https://doi.org/10.1186/s12885-015-1012-8

504 Domínguez Conde, C. *et al.* Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**, eabl5197 (2022). https://doi.org/10.1126/science.abl5197

505 Wu, Z. *et al.* The Landscape of Immune Cells Infiltrating in Prostate Cancer. *Front Oncol* **10**, 517637 (2020). https://doi.org/10.3389/fonc.2020.517637

506 Tivey, A., Church, M., Rothwell, D., Dive, C. & Cook, N. Circulating tumour DNA — looking beyond the blood. *Nat Rev Clin Oncol* **19**, 600-612 (2022). https://doi.org/10.1038/s41571-022-00660-y

507 Watanabe, K. *et al.* Tumor-Informed Approach Improved ctDNA Detection Rate in Resected Pancreatic Cancer. *Int J Mol Sci* **23**, 11521 (2022). https://doi.org/10.3390/ijms231911521

508 Chen, K. *et al.* Individualized tumor-informed circulating tumor DNA analysis for postoperative monitoring of non-small cell lung cancer. *Cancer Cell* **41**, 1749-1762.e1746 (2023). https://doi.org/10.1016/j.ccell.2023.08.010

509 Reinert, T. *et al.* Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients With Stages I to III Colorectal Cancer. *JAMA Oncol* **5**, 1124-1131 (2019). https://doi.org/10.1001/jamaoncol.2019.0528

510 Dillinger, T. *et al.* Identification of tumor tissue-derived DNA methylation biomarkers for the detection and therapy response evaluation of metastatic castration resistant prostate cancer in liquid biopsies. *Mol Cancer* **21**, 7 (2022). https://doi.org/10.1186/s12943-021-01445-0

511 Füllgrabe, J. *et al.* Simultaneous sequencing of genetic and epigenetic bases in DNA. *Nat Biotechnol* **41**, 1457-1464 (2023). https://doi.org/10.1038/s41587-022-01652-0

512 Zhang, Y. *et al.* Pan-cancer circulating tumor DNA detection in over 10,000 Chinese patients. *Nat Commun* **12**, 11 (2021). https://doi.org/10.1038/s41467-020-20162-8

513 Mayrhofer, M. *et al.* Cell-free DNA profiling of metastatic prostate cancer reveals microsatellite instability, structural rearrangements and clonal hematopoiesis. *Genome Med* **10**, 85 (2018). https://doi.org/10.1186/s13073-018-0595-5

514 Ritch, E. *et al.* Identification of Hypermutation and Defective Mismatch Repair in ctDNA from Metastatic Prostate Cancer. *Clin Cancer Res* **26**, 1114-1125 (2020). https://doi.org/10.1158/1078-0432.Ccr-19-1623

515 Conteduca, V. *et al.* Androgen receptor gene status in plasma DNA associates with worse outcome on enzalutamide or abiraterone for castration-resistant prostate cancer: a multi-institution correlative biomarker study. *Ann Oncol* **28**, 1508-1516 (2017). https://doi.org/10.1093/annonc/mdx155

516 Sumiyoshi, T. *et al.* Clinical utility of androgen receptor gene aberrations in circulating cell-free DNA as a biomarker for treatment of castration-resistant prostate cancer. *Sci Rep* **9**, 4030 (2019). https://doi.org/10.1038/s41598-019-40719-y

517 Xia, Y. *et al.* Copy number variations in urine cell free DNA as biomarkers in advanced prostate cancer. *Oncotarget* **7**, 35818-35831 (2016). https://doi.org/10.18632/oncotarget.9027

518 Xia, S. *et al.* Plasma genetic and genomic abnormalities predict treatment response and clinical outcome in advanced prostate cancer. *Oncotarget* **6**, 16411-16421 (2015). https://doi.org/10.18632/oncotarget.3845

519 Pope, B. *et al.* Ultrasensitive Detection of Circulating Tumour DNA enriches for Patients with a Greater Risk of Recurrence of Clinically Localised Prostate Cancer. *Eur Urol* **85**, 407-410 (2024). https://doi.org/10.1016/j.eururo.2024.01.002

520 Salvi, S. *et al.* Urine Cell-Free DNA Integrity Analysis for Early Detection of Prostate Cancer Patients. *Dis Markers* **2015**, 574120 (2015). https://doi.org/10.1155/2015/574120

521 Casadio, V. *et al.* Urine cell-free DNA integrity as a marker for early prostate cancer diagnosis: a pilot study. *Biomed Res Int* **2013**, 270457 (2013). https://doi.org/10.1155/2013/270457

522 Fawzy, A., Sweify, K. M., El-Fayoumy, H. M. & Nofal, N. Quantitative analysis of plasma cell-free DNA and its DNA integrity in patients with metastatic prostate cancer using ALU sequence. *J Egypt Natl Canc Inst* **28**, 235-242 (2016). https://doi.org/10.1016/j.jnci.2016.08.003

523 Jung, K., Fleischhacker, M. & Rabien, A. Cell-free DNA in the blood as a solid tumor biomarker—A critical appraisal of the literature. *Clin Chim Acta* **411**, 1611-1624 (2010). https://doi.org/10.1016/j.cca.2010.07.032

524 Casadio, V. *et al.* Cell-Free DNA Integrity Analysis in Urine Samples. *J Vis Exp* (2017). https://doi.org/10.3791/55049

525 Ellinger, J. *et al.* CpG island hypermethylation in cell-free serum DNA identifies patients with localized prostate cancer. *Prostate* **68**, 42-49 (2008). https://doi.org/10.1002/pros.20651

526 Payne, S. R. *et al.* DNA methylation biomarkers of prostate cancer: confirmation of candidates and evidence urine is the most sensitive body fluid for non-invasive detection. *Prostate* **69**, 1257-1269 (2009). https://doi.org/10.1002/pros.20967

527 Constâncio, V., Nunes, S. P., Henrique, R. & Jerónimo, C. DNA Methylation-Based Testing in Liquid Biopsies as Detection and Prognostic Biomarkers for the Four Major Cancer Types. *Cells* **9** (2020). https://doi.org/10.3390/cells9030624

528 Chen, G. *et al.* Urine- and Blood-Based Molecular Profiling of Human Prostate Cancer. *Front Oncol* **12**, 759791 (2022). https://doi.org/10.3389/fonc.2022.759791

529 Donovan, M. J. *et al.* A molecular signature of PCA3 and ERG exosomal RNA from non-DRE urine is predictive of initial prostate biopsy result. *Prostate Cancer Prostatic Dis* **18**, 370-375 (2015). https://doi.org/10.1038/pcan.2015.40

530 McKiernan, J. *et al.* A Prospective Adaptive Utility Trial to Validate Performance of a Novel Urine Exosome Gene Expression Assay to Predict High-grade Prostate Cancer in Patients with Prostate-specific Antigen 2-10ng/ml at Initial Biopsy. *Eur Urol* **74**, 731-738 (2018). https://doi.org/10.1016/j.eururo.2018.08.019

531 Woo, J. *et al.* Urine Extracellular Vesicle GATA2 mRNA Discriminates Biopsy Result in Men with Suspicion of Prostate Cancer. *J Urol* **204**, 691-700 (2020). https://doi.org/10.1097/ju.0000000000001066

532 Zhao, F. *et al.* A urine-based DNA methylation assay, ProCUrE, to identify clinically significant prostate cancer. *Clin Epigenetics* **10**, 147 (2018). https://doi.org/10.1186/s13148-018-0575-z

533 Barceló, M., Castells, M., Bassas, L., Vigués, F. & Larriba, S. Semen miRNAs Contained in Exosomes as Non-Invasive Biomarkers for Prostate Cancer Diagnosis. *Sci Rep* **9**, 13772 (2019). https://doi.org/10.1038/s41598-019-50172-6

534 De Laere, B. *et al.* TP53 Outperforms Other Androgen Receptor Biomarkers to Predict Abiraterone or Enzalutamide Outcome in Metastatic Castration-Resistant Prostate Cancer. *Clin Cancer Res* **25**, 1766-1773 (2019). https://doi.org/10.1158/1078-0432.Ccr-18-1943

535 De Laere, B. *et al.* Comprehensive Profiling of the Androgen Receptor in Liquid Biopsies from Castration-resistant Prostate Cancer Reveals Novel Intra-AR Structural Variation and Splice Variant Expression Patterns. *Eur Urol* **72**, 192-200 (2017). https://doi.org/10.1016/j.eururo.2017.01.011

536 Kohli, M. *et al.* Prognostic association of plasma cell-free DNA-based androgen receptor amplification and circulating tumor cells in pre-chemotherapy metastatic castration-resistant prostate cancer patients. *Prostate Cancer Prostatic Dis* **21**, 411-418 (2018). https://doi.org/10.1038/s41391-018-0043-z

537 Hofmann, L. *et al.* A Multi-Analyte Approach for Improved Sensitivity of Liquid Biopsies in Prostate Cancer. *Cancers (Basel)* **12**, 2247 (2020). https://doi.org/10.3390/cancers12082247

538 Fettke, H. *et al.* Combined Cell-free DNA and RNA Profiling of the Androgen Receptor: Clinical Utility of a Novel Multianalyte Liquid Biopsy Assay for Metastatic Prostate Cancer. *Eur Urol* **78**, 173-180 (2020). https://doi.org/10.1016/j.eururo.2020.03.044

539 Hodara, E. *et al.* Multiparametric liquid biopsy analysis in metastatic prostate cancer. *JCI Insight* **4**, e125529 (2019). https://doi.org/10.1172/jci.insight.125529

540 Connell, S. P. *et al.* Development of a multivariable risk model integrating urinary cell DNA methylation and cell-free RNA data for the detection of significant prostate cancer. *Prostate* **80**, 547-558 (2020). https://doi.org/10.1002/pros.23968

541 Corrò, C. *et al.* Detecting circulating tumor DNA in renal cancer: An open challenge. *Exp Mol Pathol* **102**, 255-261 (2017). https://doi.org/10.1016/j.yexmp.2017.02.009

542 Yamamoto, Y. *et al.* Clinical significance of the mutational landscape and fragmentation of circulating tumor DNA in renal cell carcinoma. *Cancer Sci* **110**, 617-628 (2019). https://doi.org/10.1111/cas.13906

543 Peneder, P. *et al.* Multimodal analysis of cell-free DNA whole-genome sequencing for pediatric cancers with low mutational burden. *Nat Commun* **12**, 3230 (2021). https://doi.org/10.1038/s41467-021-23445-w

544 Moser, T. & Heitzer, E. Surpassing sensitivity limits in liquid biopsy. *Science* **383**, 260-261 (2024). https://doi.org/10.1126/science.adn1886

545 Kluge, K. *et al.* Comparison of discovery rates and prognostic utility of [(68)Ga]Ga-PSMA-11 PET/CT and circulating tumor DNA in prostate cancer-a cross-sectional study. *Eur J Nucl Med Mol Imaging* **51**, 2833-2842 (2024). https://doi.org/10.1007/s00259-024-06698-7

546 Conteduca, V. *et al.* Changing metastatic patterns associate with dynamics of circulating tumor DNA in metastatic castration-resistant prostate cancer. *Oncologist* **30** (2025). https://doi.org/10.1093/oncolo/oyaf107

547 Fonseca, N. M. *et al.* Prediction of plasma ctDNA fraction and prognostic implications of liquid biopsy in advanced prostate cancer. *Nat Commun* **15**, 1828 (2024). https://doi.org/10.1038/s41467-024-45475-w

548 Tokac, R. H. *et al.* Cross-Sectional Analysis of Metabolic Tumor Burden Detected by F-18 FDG PET/CT and Circulating Tumor DNA in Advanced Breast Cancer. *Cancer Med* **14**, e71049 (2025). https://doi.org/10.1002/cam4.71049

549 Schweizer, M. T. *et al.* Clinical determinants for successful circulating tumor DNA analysis in prostate cancer. *Prostate* **79**, 701-708 (2019). https://doi.org/10.1002/pros.23778

550 Merriel, S. W. D. *et al.* Systematic review and meta-analysis of the diagnostic accuracy of prostate-specific antigen (PSA) for the detection of prostate cancer in symptomatic patients. *BMC Med* **20**, 54 (2022). https://doi.org/10.1186/s12916-021-02230-y

551 Bangma, C. H., Roemeling, S. & Schröder, F. H. Overdiagnosis and overtreatment of early detected prostate cancer. *World J Urol* **25**, 3-9 (2007). https://doi.org/10.1007/s00345-007-0145-z

552 Narayan, P. *et al.* State of the Science and Future Directions for Liquid Biopsies in Drug Development. *Oncologist* **25**, 730-732 (2020). https://doi.org/10.1634/theoncologist.2020-0246

553 Edsjö, A. *et al.* Precision cancer medicine: Concepts, current practice, and future developments. *J Intern Med* **294**, 455-481 (2023). https://doi.org/10.1111/joim.13709

554 Medford, A. J. *et al.* A standing platform for cancer drug development using ctDNA-based evidence of recurrence. *Nat Rev Cancer* **24**, 810-821 (2024). https://doi.org/10.1038/s41568-024-00742-2

7 Appendix

7.1 Appendix A: Tables

Table A1: Results from a questionnaire assessing lifestyle factors and potential PCa risk factors in PCa patients and controls.
Results were compared between PCa patients and controls and across multiple groups, including localized Pca patients, advanced PCa patients, and controls. Continuous variables were analyzed using the Wilcoxon rank-sum test (tumor vs. control) or the Kruskal-Wallis test (multiple-group comparisons), while categorical variables were analyzed using Fisher's exact test. Correction for multiple testing was performed using the Benjamini-Hochberg method, with adjusted p < 0.05 considered significant.

	control median, range	**tumor** median, range	**localized PCa** median, range	**advanced PCa** median, range	**tumor vs. control** p value, adj.p value	**control vs. localized PCa vs. advanced PCa** p value, adj.p value
meat consumption (days/week)	4 [1–7] NA: 3	4 [0–7] NA: 5	4 [0–7] NA: 4	4 [1–6] NA: 1	0.474 [0.744]	0.769 [0.982]
alcohol consumption (days/week)	2 [0–7] NA: 3	2 [0–7] NA: 12	2 [0–7] NA: 8	1.5 [0–7] NA: 4	0.235 [0.744]	0.386 [0.707]
physical examination (days/week)	2.5 [0–10] NA: 2	3 [0–18] NA 11	3 [0–18] NA: 8	3 [0–7] NA: 3	0.439 [0.744]	0.727 [0.982]
smoking (packyears)	0 [0–62] NA: 2	0 [0–20] NA: 8	0 [0–0] NA: 4	0 [0–20] NA: 4	0.0001 [0.001]	0.0004 [0.002]
body mass index	27.1 [18.7–38.9] NA: 2	26.2 [21–42.1] NA: 4	26.5 [21–36.7] NA: 4	25.4 [21.6–42.1] NA: 0	0.314 [0.744]	0.213 [0.468]
PCa family history count (yes/no)	4–30 NA: 2	12–57 NA: 4	6–45 NA: 4	6–12 NA: 0	0.570 [0.784]	0.102 [0.374]
current smoker count (yes/no)	9–25 NA: 2	3–69 NA: 1	0–54 NA: 1	3–15 NA: 0	0.002 [0.009]	0.0001 [0.001]
infertility count (yes/no)	4–30 NA: 2	8–63 NA: 2	7–47 NA: 1	1–16 NA: 1	1	0.843 [0.982]
sterilization count (yes/no)	4–30 NA: 2	8–62 NA: 3	6–47 NA: 2	2–16 NA: 1	1	1 [1]
urinary tract infection (last 6 months) count (yes/no)	3–31 NA: 2	6–65 NA: 2	4–50 NA: 1	2–15 NA: 1	1	0.893 [0.982]
urinary retention (last 6 months) count (yes/no)	1–33 NA: 2	6–65 NA: 2	3–51 NA: 1	3–14 NA: 1	0.424 [0.744]	0.178 [0.468]

Abbreviations: adj. p value = adjusted p value, NA = no data available, PCa = prostate cancer

Table A2: QC metrics for lcWGS and (cf)MeDIP-seq data processing.
Summary of QC parameters for lcWGS and (cf)MeDIP-seq data derived from plasma and urinary cfDNA, as well as from gDNA of PCa tissue and matched buffy coat samples. Reported metrics include duplicate read fractions before and after deduplication (%), deduplication rates, and GC content before and after processing (%).

	duplicates raw, %	**duplicates final**, %	**deduplication rate**, %	**GC content raw**, %	**GC content final**, %
cfDNA, lcWGS	10.36 [6.66–14.96]	1.72 [0.98–12.58]	82.84 [-35.65–91.76]	42 [40–45]	41 [39–43]
plasma, lcWGS	10.31 [6.85–14.96]	1.55 [0.98–2.1]	84.02 [78.08–91.76]	42 [41–45]	41 [41–43]
urine, lcWGS	10.505 [6.66–14.82]	2.49 [1.11–12.58]	74.50 [-35.65–89.70]	42 [40–44]	41 [39–43]
cfDNA, cfMeDIP	23.132 [13.96–45.0]	5.64 [0.06–11.2]	75.29 [47.55–99.62]	46 [41–49]	47 [41–50]
plasma, cfMeDIP	22.86 [15.79–38.23]	6.3 [3.59–9.01]	72.21 [60.17–85.8]	47 [43–49]	47 [43–50]
urine, cfMeDIP	23.58 [13.96–45.0]	4.96 [0.06–11.2]	79.62 [47.55–99.62]	46 [41–48]	47 [41–49]
gDNA, lcWGS	7.0 [6.45–7.58]	7.77 [3.41–15.91]	-14.26 [-131.08–47.74]	41 [40–41]	40 [40–41]
PCa tissue, lcWGS	7.13 [6.94–7.58]	6.16 [5.3–14]	13.07 [-99.29–26.99]	41 [40–41]	40 [40–41]
buffy coat, lcWGS	6.85 [6.45–7.48]	9.63 [3.41–15.91]	-37.04 [-131.08–47.74]	41 [40–41]	40 [40–41]
gDNA, MeDIP	16.20 [13.66–20.38]	4.87 [2.38–6.94]	70.66 [49.98–86.35]	46 [46–47]	47 [47–48]
PCa tissue, MeDIP	17.25 [13.85–20.38]	4.41 [2.38–6.2]	74.69 [60.08–86.35]	46 [46–47]	47 [47–48]
buffy coat, MeDIP	15.92 [13.66–17.32]	5.55 [3.7–6.94]	66.13 [49.98–76.58]	46 [46–47]	47 [47–48]

Abbreviations: cfDNA = cell-free DNA, gDNA = genomic DNA, lcWGS = low-coverage whole-genome sequencing, (cf)MeDIP = (cell-free) methylated DNA immunoprecipitation sequencing, QC = quality control

Table A3: Laboratory-based QC metrics for methylation enrichment.
Laboratory methylation QC metrics for (cf)MeDIP-seq libraries generated from plasma and urinary cfDNA, as well as from gDNA of PCa tissue and matched buffy coat samples. Reported metrics include the difference of Ct values between 5mC- and 5C-spike-ins measured by qPCR in (cf)MeDIP-seq libraries, the recovery rates of 5mC- and 5C-spike-ins (%), and the specificity of the methylation enrichment.

	spike-in difference [median, range]	**recovery 5mC ,%** [median, range]	**recovery 5C,%** [median, range]	**specificity** [median, range]
plasma cfDNA, cfMeDIP	8.47 [6.06–11.47]	78.35 [23.95–140.26]	0.31 [0.09–0.85]	0.996 [0.985–0.999]
control	8.66 [7.21–10.11]	77.54 [23.95–103.39]	0.32 [0.09–0.58]	0.996 [0.99–0.999]
tumor all	8.41 [6.06–11.47]	78.9 [24.28–140.26]	0.30 [0.1–0.85]	0.996 [0.985–0.998]
localized PCa	8.45 [6.52–11.47]	78.9 [24.28–98.49]	0.30 [0.1–0.85]	0.996 [0.985–0.998]
advanced PCa	8.38 [6.06–10.51]	76.48 [53.15–140.26]	0.30 [0.19–0.65]	0.996 [0.99–0.998]
urinary cfDNA, cfMeDIP	8.45 [6.2–11.53]	78.35 [23.46–132.69]	0.31 [0.13–1.21]	0.996 [0.987–0.998]
control	8.35 [7.11–10.14]	81.12 [23.46–132.69]	0.32 [0.13–0.8]	0.996 [0.99–0.997]
tumor all	8.47 [6.2–11.53]	77.27 [26.94–100.56]	0.31 [0.13–1.21]	0.996 [0.987–0.998]
localized PCa	8.52 [6.72–11.53]	77.81 [26.94–93.18]	0.31 [0.13–0.88]	0.996 [0.989–0.998]
advanced PCa	8.47 [6.2–10.54]	72.87 [58.56–100.56]	0.31 [0.2–1.21]	0.996 [0.987–0.997]
gDNA, MeDIP	6.92 [5.57–7.51]	81.69 [68.92–94.48]	1.14 [0.90–1.28]	0.987 [0.983–0.988]
PCa tissue	6.88 [5.57–7.49]	81.81 [68.92–93.18]	1.14 [0.9–1.2]	0.987 [0.983–0.988]
buffy coat	6.98 [5.59–7.51]	81.69 [69.16–94.48]	1.12 [0.99–1.28]	0.987 [0.984–0.988]

Abbreviations: 5C = unmethylated, 5mC = methylated

Table A4: Bioinformatic QC metrics for (cf)MeDIP-seq data.
Bioinformatic methylation QC metrics for lcWGS and (cf)MeDIP-seq data derived from plasma and urinary cfDNA, as well as from gDNA of PCa tissue and matched buffy coat samples. Metrics include the estimated saturation, methylation enrichment score, the proportion of CpG sites covered relative to all CpGs in the reference genome (CpGs covered), the proportion of CpGs covered by more than five sequencing reads (CpGs > 5 reads), and the proportion of reads without a CpG pattern (read w/o CpG).

	saturation median [range]	**enrichment score** median [range]	**CpGs covered** median [range]	**CpGs > 5 reads** median [range]	**reads w/o CpG** median [range]
plasma cfDNA, cfMedIP	0.989 [0.963–0.994]	2.42 [1.88–3.18]	0.786 [0.544–0.885]	0.372 [0.138–0.577]	0.036 [0.022–0.14]
control	0.989 [0.985–0.993]	2.41 [2.14–3.18]	0.786 [0.612–0.843]	0.374 [0.257–0.46]	0.036 [0.022–0.062]
tumor all	0.989 [0.963–0.994]	2.44 [1.88–3.03]	0.786 [0.544–0.885]	0.368 [0.138–0.577]	0.037 [0.024–0.14]
localized PCa	0.989 [0.963–0.994]	2.44 [1.88–3.03]	0.787 [0.544–0.885]	0.37 [0.138–0.577]	0.036 [0.024–0.14]
advanced PCa	0.989 [0.976–0.993]	2.44 [2.01–2.68]	0.782 [0.679–0.83]	0.364 [0.262–0.419]	0.041 [0.029–0.097]
urinary cfDNA, cfMeDIP	0.992 [0.984–0.997]	2.25 [1.82–2.84]	0.801 [0.611–0.931]	0.349 [0.112–0.651]	0.0505 [0.02–0.119]
control	0.992 [0.985–0.996]	2.25 [1.82–2.47]	0.809 [0.675–0.881]	0.354 [0.139–0.5]	0.053 [0.024–0.096]
tumor all	0.992 [0.984–0.997]	2.26 [1.89–2.84]	0.796 [0.611–0.931]	0.347 [0.112–0.651]	0.05 [0.02–0.119]
localized PCa	0.992 [0.984–0.997]	2.28 [1.89–2.84]	0.804 [0.611–0.931]	0.348 [0.149–0.651]	0.049 [0.028–0.119]
advanced PCa	0.991 [0.988–0.996]	2.20 [1.90–2.66]	0.781 [0.669–0.877]	0.342 [0.112–0.396]	0.057 [0.02–0.094]
gDNA, MeDIP	0.993 [0.990– 0.997]	2.37 [2.23–2.62]	0.761 [0.699–0.81]	0.359 [0.247–0.469]	0.020 [0.014–0.031]
PCa tissue	0.993 [0.99–0.997]	2.41 [2.34–2.62]	0.755 [0.699–0.774]	0.347 [0.247–0.377]	0.023 [0.015–0.025]
buffy coat	0.993 [0.991–0.996]	2.31 [2.23–2.58]	0.768 [0.742–0.81]	0.370 [0.347–0.469]	0.017 [0.014–0.031]
plasma cfDNA, lcWGS	0.977 [0.959;0.989]	1.14 [1.07–1.25]	0.895 [0.756–0.949]	0.067 [0.008–0.225]	0.29 [0.228–0.327]
control	0.977 [0.969–0.987]	1.15 [1.07–1.19]	0.903 [0.802–0.945]	0.077 [0.015–0.194]	0.287 [0.228–0.316]
tumor all	0.978 [0.959–0.989]	1.13 [1.08–1.25]	0.890 [0.756–0.949]	0.06 [0.008–0.225]	0.291 [0.255–0.327]
localized PCa	0.978 [0.959–0.989]	1.13 [1.09–1.25]	0.889 [0.756–0.949]	0.058 [0.008–0.225]	0.290 [0.255–0.327]
advanced PCa	0.977 [0.971–0.987]	1.13 [1.08–1.17]	0.903 [0.822–0.933]	0.076 [0.021–0.151]	0.293 [0.269–0.318]
urinary cfDNA, lcWGS	0.984 [0.959;0.992]	1.14 [0.95–1.39]	0.884 [0.677–0.958]	0.057 [0.003–0.324]	0.325 [0.201–0.404]
control	0.982 [0.974–0.992]	1.15 [1.03–1.39]	0.890 [0.707–0.956]	0.065 [0.009–0.324]	0.323 [0.201–0.391]
tumor all	0.984 [0.959–0.990]	1.13 [0.95–1.24]	0.882 [0.677–0.958]	0.052 [0.003–0.272]	0.325 [0.242–0.404]
localized PCa	0.984 [0.972–0.990]	1.14 [0.99–1.24]	0.881 [0.773–0.958]	0.051 [0.012–0.272]	0.325 [0.242–0.4]
advanced PCa	0.985 [0.959–0.990]	1.13 [0.95–1.21]	0.888 [0.677–0.938]	0.055 [0.003–0.151]	0.327 [0.245–0.404]
gDNA, lcWGS	0.977 [0.964;0.990]	1.06 [0.99–1.07]	0.87 [0.828–0.891]	0.036 [0.023–0.051]	0.303 [0.267–0.359]
PCa tissue	0.981 [0.975;0.990]	1.06 [0.99–1.07]	0.86 [0.828–0.891]	0.030 [0.023–0.051]	0.31 [0.285–0.359]

buffy coat	0.974 [0.964;0.986]	1.06 [1.05–1.07]	0.878 [0.862–0.891]	0.040 [0.03–0.051]	0.297 [0.267–0.343]

Abbreviations: w/o = without

Table A5: Maximum relative frequencies and peak positions of plasma cfDNA fragmentation profiles in tumor and control samples.

Assessment of the maximum relative frequencies and the corresponding fragment lengths associated with the first and second fragmentation peaks in plasma cfDNA from PCa patients and controls, including stratification of PCa patients into localized and advanced disease. In addition, ratios of maximum relative frequencies were calculated for tumor versus control samples, as well as for localized and advanced PCa patients relative to controls, respectively. Relative frequencies were calculated relative to the full fragment length range (30–700 bp), and the subranges 30–250 bp, and 250–700 bp.

	control, max freq [bp]	**tumor**, max freq [bp]	**localized PCa**, max freq [bp]	**advanced PCa**, max freq [bp]	**tumor all vs. control**, ratio	**localized PCa vs. control**, ratio	**advanced PCa vs. control**, ratio	**advanced PCa vs. localized PCa**, ratio
Reference fragment length range: 30–700 bp								
first peak	0.0219 [167]	0.0237 [167]	0.0236 [167]	0.0245 [166]	1.082	1.076	1.116	1.037
second peak	0.0018 [332]	0.0016 [323]	0.0016 [330]	0.0016 [320]	0.879	0.885	0.908	1.026
Reference fragment length range: 30–250 bp								
first peak	0.0278 [167]	0.0283 [167]	0.0282 [167]	0.0292 [167]	1.017	1.015	1.052	1.036
Reference fragment length range: 250–700 bp								
second peak	0.0010 [332]	0.0010 [325]	0.0010 [325]	0.0101 [324]	1.002	0.998	1.019	1.021

Abbreviations: bp = base pair, max freq = maximum (relative) frequency

Table A6: Overview of quantitative cfDNA fragmentation characteristics in plasma and urinary cfDNA. Quantitative fragmentation features assessed in plasma and urinary cfDNA from PCa patients and controls, including stratification of PCa patients into localized and advanced disease. The assessed cfDNA fragmentation characteristics comprised descriptive statistical parameters, proportions of fragment length ranges, ratios between fragment length range proportions, and metrics of the 10-bp oscillation pattern. Statistical comparisons between tumor and control samples were performed using the Wilcoxon rank-sum test. Correction for multiple testing was applied using the Benjamini-Hochberg method. Statistically significant results are highlighted in bold.

A: plasma cfDNA					
	control, median [range]	**tumor,** median [range]	**localized PCa,** median [range]	**advanced PCa,** median [range]	**tumor vs. control,** p value [adj. p value]
statistical parameters					
max. rel. frequency	0.022 [0.01–0.029]	0.024 [0.016–0.037]	0.024 [0.016–0.037]	0.024 [0.018–0.029]	0.101 [0.234]
cum. frequency at fragment length of max. rel. frequency	0.348 [0.259–0.453]	0.36 [0.268–0.628]	0.36 [0.268–0.628]	0.358 [0.27–0.47]	0.794 [0.816]
modal fragment length	167 [166–177]	167 [165–176]	167 [165–176]	167 [166–167]	0.599 [0.692]
mean fragment length	205 [180–275]	198 [162–235]	199 [162–235]	196 [181–227]	0.369 [0.525]
rel. frequency at 167 bp	0.022 [0.008–0.029]	0.024 [0.014–0.034]	0.024 [0.014–0.034]	0.024 [0.018–0.028]	0.098 [0.234]
rel. frequency at 334 bp	0.002 [0.001–0.003]	0.002 [0–0.003]	0.002 [0–0.003]	0.001 [0.001–0.003]	0.155 [0.308]
fragment length at cum. frequency of 0.5	176 [168–220]	174 [161–184]	174 [161–184]	174 [168–181]	0.705 [0.768]
proportions of fragment length ranges					
P30–60bp	0.005 [0.002–0.027]	0.004 [0.001–0.017]	0.004 [0.001–0.017]	0.003 [0.001–0.014]	**0.030** [0.170]
P30–100bp	0.018 [0.008–0.063]	0.016 [0.007–0.057]	0.018 [0.007–0.057]	0.014 [0.008–0.043]	0.050 [0.170]
P30–150bp	0.131 [0.082–0.198]	0.121 [0.07–0.3]	0.121 0.07–0.3]	0.119 [0.081–0.22]	0.673 [0.755]
P30–180bp	0.581 [0.287–0.727]	0.625 [0.457–0.891]	0.625 [0.457–0.891]	0.625 [0.487–0.72]	0.473 [0.595]
P90–150bp	0.113 [0.071–0.165]	0.11 [0.064–0.258]	0.11 [0.064–0.258]	0.111 [0.074–0.194]	0.849 [0.849]
P160–180bp	0.376 [0.169–0.471]	0.399 [0.284–0.519]	0.397 [0.284–0.519]	0.399 [0.307–0.466]	0.073 [0.192]
P163–169bp	0.142 [0.053–0.188]	0.154 [0.092–0.233]	0.153 [0.092–0.233]	0.157 [0.117–0.184]	0.117 [0.254]
P180–220bp	0.716 [0.506–0.805]	0.734 [0.577–0.818]	0.734 [0.577–0.818]	0.74 [0.611–0.812]	0.408 [0.539]
P150–300bp	0.206 [0.176–0.283]	0.217 [0.069–0.294]	0.217 [0.069–0.294]	0.208 [0.144–0.233]	**0.041** [0.170]
P324–344bp	0.036 [0.018–0.062]	0.031 [0.007–0.063]	0.031 [0.007–0.063]	0.03 [0.016–0.056]	0.167 [0.309]
P250–320bp	0.053 [0.03–0.083]	0.049 [0.023–0.08]	0.048 [0.023–0.08]	0.052 [0.032–0.074]	0.538 [0.643]
P250–420bp	0.156 [0.079–0.351]	0.136 [0.039–0.283]	0.139 [0.039–0.283]	0.13 [0.08–0.235]	0.240 [0.377]

P420–700bp	0.022 [0.006–0.11]	0.018 [0.003–0.062]	0.018 [0.003–0.062]	0.018 [0.008–0.055]	0.158 [0.308]
ratios of proportions of fragment length ranges					
P30–150bp/P160–180bp	0.328 [0.211–0.558]	0.305 [0.206–0.694]	0.315 [0.206–0.694]	0.276 [0.207–0.583]	0.055 [0.170]
P30–150bp/P163–169bp	0.877 [0.584–1.58]	0.799 [0.54–1.58]	0.811 [0.574–1.58]	0.706 [0.54–1.41]	**0.041** [0.170]
P30–100bp/P160–180bp	0.053 [0.02–0.177]	0.04 [0.015–0.156]	0.045 [0.015–0.156]	0.034 [0.017–0.112]	**0.021** [0.170]
P30–100bp/P163–169bp	0.144 [0.057–0.492]	0.1 [0.037–0.406]	0.109 [0.037–0.406]	0.086 [0.044–0.284]	**0.017** [0.170]
P30–150bp/P150–300bp	0.17 [0.109–0.286]	0.164 [0.097–0.439]	0.166 [0.097–0.439]	0.155 [0.118–0.32]	0.201 [0.338]
P90–150bp/P163–169bp	0.759 [0.542–1.34]	0.714 [0.508–1.26]	0.727 [0.534–1.26]	0.665 [0.508–1.25]	0.060 [0.171]
P160–180bp/P250–420bp	2.53 [0.483–5.82]	3.05 [1.01–11.78]	2.96 [1.01–11.78]	3.12 [1.32–5.45]	0.181 [0.319]
10-bp oscillation pattern (30–150bp fragment length range)					
10-bp oscillation score	0.0021 [0.001–0.01]	0.0016 [-0.01–0.01]	0.0016 [-0.01–0.01]	0.0015 [0.001–0.007]	**0.020** [0.170]
number of local maxima	10 [8–11]	9 [6–11]	9 [6–11]	9 [8–10]	0.009 [0.170]
number of local minima	9 [8–11]	9 [6–11]	9 [6–11]	9 [7–10]	0.036 [0.170]
mean interpeak distance	9.5 [9.1–9.6]	9.3 [7.7–10.5]	9.3 [7.7–9.5]	9.3 [9.1–10.5]	**0.024** [0.170]
mean intervalley distance	9.4 [9.1–9.6]	9.3 [8.7–10.4]	9.3 [8.7–10.4]	9.3 [9.1–9.5]	0.245 [0.377]
10-bp oscillation pattern (150–300bp fragment length range)					
10-bp oscillation score	0.025 [0.002–0.043]	0.025 [0.002–0.04]	0.026 [0.016–0.04]	0.024 [0.002–0.038]	0.482 [0.595]
number of local maxima	4 [1–7]	4 [1–7]	4 [1–7]	3 [1–5]	0.408 [0.539]
number of local minima	4 [1–5]	3 [1–5]	4 [1–5]	3 [1–5]	0.769 [0.813]
mean interpeak distance	27.75 [17.714–48]	28 [17–37.333]	27.25 [17–37.333]	31.458 [24–37.333]	0.259 [0.383]
mean intervalley distance	25.5 [5–34.333]	20.4 [0–52.5]	20.4 [0–37.333]	13.9 [4.5–52.5]	0.054 [0.170]
B: urinary cfDNA					
	control, median [range]	**tumor,** median [range]	**localized PCa,** median [range]	**advanced PCa,** median [range]	**tumor vs. control,** p value [adj. p value]
statistical parameters					
max. rel. frequency	0.007 [0.003–0.018]	0.007 [0.004–0.016]	0.007 [0.004–0.016]	0.007 [0.004–0.011]	0.882 [1.0]
cum. frequency at fragment length of max. rel. frequency	0.228 [0.102–0.546]	0.205 [0.081–0.517]	0.192 [0.092–0.517]	0.325 [0.081–0.501]	0.783 [1.0]

modal fragment length	111 [70–180]	111 [50–179]	111 [50–170]	148 [80–179]	0.911 [1.0]
mean fragment length	194 [145–272]	188 [135–266]	188 [135–266]	184 [151–242]	0.642 [1.0]
rel. frequency at 167 bp	0.004 [0.002–0.01]	0.005 [0.003–0.007]	0.005 [0.003–0.007]	0.005 [0.004–0.007]	**0.036** [0.681]
rel. frequency at 334 bp	0.001 [0–0.002]	0.001 [0–0.002]	0.001 [0–0.002]	0.001 [0.001–0.002]	0.627 [1.0]
fragment length at cum. frequency of 0.5	169 [111–257]	167 [113–236]	167 [113–236]	166 [130–211]	0.775 [1.0]
proportions of fragment length ranges					
P30–60bp	0.031 [0.008–0.124]	0.029 [0.005–0.15]	0.034 [0.008–0.15]	0.026 [0.005–0.073]	1.00 [1.0]
P30–100bp	0.186 [0.053–0.433]	0.176 [0.042–0.408]	0.177 [0.053–0.408]	0.15 [0.042–0.327]	0.994 [1.0]
P30–150bp	0.403 [0.19–0.664]	0.421 [0.173–0.709]	0.421 [0.173–0.709]	0.414 [0.179–0.612]	0.874 [1.0]
P30–180bp	0.548 [0.286–0.755]	0.566 [0.307–0.801]	0.566 [0.307–0.801]	0.575 [0.344–0.75]	0.672 [1.0]
P90–150bp	0.289 [0.148–0.388]	0.286 [0.137–0.433]	0.286 [0.137–0.433]	0.284 [0.152–0.383]	0.494 [1.0]
P160–180bp	0.089 [0.05–0.2]	0.097 [0.059–0.144]	0.094 [0.059–0.144]	0.105 [0.072–0.135]	0.055 [0.681]
P163–169bp	0.030 [0.017–0.07]	0.033 [0.02–0.051]	0.033 [0.02–0.049]	0.035 [0.026–0.051]	**0.044** [0.681]
P180–220bp	0.136 [0.076–0.199]	0.143 [0.082–0.203]	0.139 [0.082–0.203]	0.149 [0.091–0.196]	0.225 [1.0]
P150–300bp	0.429 [0.252–0.56]	0.452 [0.253–0.56]	0.448 [0.253–0.56]	0.472 [0.306–0.56]	0.297 [1.0]
P324–344bp	0.023 [0.009–0.048]	0.022 [0.007–0.049]	0.022 [0.007–0.045]	0.022 [0.012–0.049]	0.624 [1.0]
P250–320bp	0.117 [0.049–0.195]	0.111 [0.05–0.178]	0.111 [0.05–0.178]	0.110 [0.055–0.166]	0.719 [1.0]
P250–420bp	0.202 [0.089–0.38]	0.183 [0.073–0.338]	0.183 [0.073–0.338]	0.177 [0.098–0.316]	0.483 [1.0]
P420–700bp	0.036 [0.011–0.143]	0.036 [0.005–0.129]	0.036 [0.005–0.129]	0.03 [0.016–0.081]	0.472 [1.0]
ratios of proportions of fragment length ranges					
P30–150bp/P160–180bp	4.75 [1.55–13.28]	4.32 [1.47–11.68]	4.57 [1.77–11.68]	3.38 [1.47–8.44]	0.512 [1.0]
P30–150bp/P163–169bp	13.82 [4.59–38.44]	12.65 [4.55–33.75]	13.22 [5.28–33.75]	9.41 [4.55–23.84]	0.438 [1.0]
P30–100bp/P160–180bp	2.16 [0.419–8.66]	1.90 [0.343–6.26]	1.94 [0.542–6.26]	1.32 [0.343–4.22]	0.503 [1.0]
P30–100bp/P163–169bp	6.36 [1.25–25.06]	5.44 [1.06–18.10]	5.59 [1.62–18.10]	3.67 [1.06–11.92]	0.468 [1.0]
P30–150bp/P150–300bp	0.948 [0.396–2.63]	0.931 [0.32–2.80]	0.931 [0.357–2.80]	0.854 [0.32–2]	0.905 [1.0]
P90–150bp/P163–169bp	9.08 [3.74–17.51]	8.48 [3.86–20.64]	8.70 [4.17–20.64]	7.14 [3.86–14.93]	0.410 [1.0]
P160–180bp/P250–420bp	0.487 [0.151–2.02]	0.539 [0.268–1.22]	0.541 [0.268–1.22]	0.535 [0.32–1.03]	0.190 [1.0]

10-bp oscillation pattern (30–150bp fragment length range)					
10-bp oscillation score	0.012 [0.001–0.059]	0.015 [0.001–0.047]	0.016 [0.001–0.047]	0.011 [0.002–0.039]	0.693 [1.0]
number of local maxima	11 [9–12]	11 [8–12]	11 [8–12]	11 [9–11]	0.828 [1.0]
number of local minima	11 [9–11]	11 [6–11]	11 [6–11]	11 [9–11]	0.866 [1.0]
mean interpeak distance	9.4 [9.1–10.2]	9.3 [9–9.8]	9.3 [9–9.8]	9.3 [9.2–9.4]	0.853 [1.0]
mean intervalley distance	9.5 [9.2–9.6]	9.5 [8.7–9.7]	9.5 [8.7–9.7]	9.4 [9–9.5]	0.986 [1.0]
10-bp oscillation pattern (150–300bp fragment length range)					
10-bp oscillation score	0.0025 [-0.005–0.009]	0.0032 [-0.006–0.013]	0.0031 [-0.006–0.012]	0.0048 [-0.002–0.013]	0.132 [0.979]
number of local maxima	13 [4–15]	12 [1–15]	12 [1–15]	11 [1–15]	0.351 [0.975]
number of local minima	13 [3–14]	12 [1–15]	12 [1–14]	10 [3–15]	0.076 [0.699]
mean interpeak distance	9.7 [9.2–31.2]	10.1 [9.2–28.6]	10 [9.3–28.6]	10.8 [9.2–22.7]	0.932 [1.0]
mean intervalley distance	9.7 [5.8–41.7]	9.8 [6–46]	9.7 [6–46]	10.8 [8.6–43.7]	0.223 [1.0]

Abbreviations: cum. frequency = cumulative frequency, (max.) rel. frequency = (maximum) relative frequency

Table A7: Multiple-group comparison of cfDNA fragmentation characteristics in plasma and urinary cfDNA. Results from statistical comparisons of quantitative cfDNA fragmentation features in plasma and urinary cfDNA among localized PCa patients, and advanced PCa patients, and controls. The assessed cfDNA fragmentation characteristics comprised descriptive statistical parameters, proportions of fragment length ranges, ratios between fragment length range proportions, and metrics of the 10-bp oscillation pattern. Multiple-group comparisons were performed using the Kruskal-Wallis test followed by Dunn's post hoc test. Correction for multiple testing was applied using the Benjamini–Hochberg method. Statistically significant results are highlighted in bold.

	plasma cfDNA		urinary cfDNA	
	p value	**adj. p value**	**p value**	**adj. p value**
P30–60bp				
Kruskal-wallis test	0.063		0.649	
localized PCa vs. controls	0.076	0.114	0.758	0.758
advanced PCa vs. controls	**0.030**	0.089	0.516	0.758
advanced PCa vs. localized PCa	0.362	0.362	0.352	0.758
P30–100bp				
Kruskal-wallis test	0.108		0.516	
localized PCa vs. controls	0.103	0.155	0.693	0.693
advanced PCa vs. controls	0.054	0.155	0.429	0.643
advanced PCa vs. localized PCa	0.446	0.446	0.250	0.643
P30–150bp				
Kruskal-wallis test	0.888		0.588	
localized PCa vs. controls	0.742	0.812	0.622	0.622
advanced PCa vs. controls	0.639	0.812	0.555	0.622
advanced PCa vs. localized PCa	0.812	0.812	0.309	0.622
P30–180bp				
Kruskal-wallis test	0.618		0.763	
localized PCa vs. controls	0.633	0.633	0.547	0.824
advanced PCa vs. controls	0.327	0.633	0.913	0.913
advanced PCa vs. localized PCa	0.506	0.633	0.549	0.824
P90–150bp				
Kruskal-wallis test	0.980		0.602	
localized PCa vs. controls	0.839	0.949	0.371	0.691
advanced PCa vs. controls	0.927	0.949	0.985	0.985
advanced PCa vs. localized PCa	0.949	0.949	0.461	0.691
P160–180bp				
Kruskal-wallis test	0.157		**0.045**	
localized PCa vs. controls	0.136	0.203	0.201	0.201
advanced PCa vs. controls	0.079	0.203	**0.013**	**0.039**
advanced PCa vs. localized PCa	0.491	0.491	0.112	0.167
P163–169bp				
Kruskal-wallis test	0.185		**0.041**	
localized PCa vs. controls	0.232	0.341	0.164	0.164
advanced PCa vs. controls	0.074	0.223	**0.012**	**0.036**
advanced PCa vs. localized PCa	0.341	0.341	0.127	0.164
P150–300bp				
Kruskal-wallis test	0.112		0.329	
localized PCa vs. controls	0.072	0.119	0.528	0.528
advanced PCa vs. controls	0.079	0.119	0.136	0.408
advanced PCa vs. localized PCa	0.656	0.656	0.288	0.432
P180–220bp				
Kruskal-wallis test	0.479		0.334	
localized PCa vs. controls	0.287	0.566	0.388	0.398
advanced PCa vs. controls	0.968	0.968	0.144	0.398
advanced PCa vs. localized PCa	0.377	0.566	0.398	0.398
P324–344bp				
Kruskal-wallis test	0.353		0.577	
localized PCa vs. controls	0.234	0.351	0.439	0.658

advanced PCa vs. controls	0.206	0.351	0.773	0.773
advanced PCa vs. localized PCa	0.686	0.686	0.355	0.658
P250–320bp				
Kruskal-wallis test	0.526		0.725	
localized PCa vs. controls	0.376	0.565	0.561	0.813
advanced PCa vs. controls	0.812	0.812	0.813	0.813
advanced PCa vs. localized PCa	0.342	0.565	0.474	0.813
P250–420bp				
Kruskal-wallis test	0.484		0.583	
localized PCa vs. controls	0.297	0.470	0.358	0.669
advanced PCa vs. controls	0.314	0.470	0.982	0.982
advanced PCa vs. localized PCa	0.804	0.804	0.446	0.669
P420–700bp				
Kruskal-wallis test	0.361		0.704	
localized PCa vs. controls	0.199	0.387	0.411	0.822
advanced PCa vs. controls	0.258	0.387	0.822	0.822
advanced PCa vs. localized PCa	0.851	0.851	0.671	0.822
P30–150bp/P160–180bp				
Kruskal-wallis test	0.127		0.207	
localized PCa vs. controls	0.106	0.159	0.944	0.944
advanced PCa vs. controls	0.068	0.159	0.105	0.158
advanced PCa vs. localized PCa	0.505	0.505	0.099	0.158
P30–150bp/P163–169bp				
Kruskal-wallis test	0.074		0.169	
localized PCa vs. controls	0.102	0.154	0.873	0.873
advanced PCa vs. controls	**0.031**	0.093	0.080	0.129
advanced PCa vs. localized PCa	0.317	0.317	0.086	0.129
P30–100bp/P160–180bp				
Kruskal-wallis test	**0.045**		0.251	
localized PCa vs. controls	0.056	0.084	0.900	0.900
advanced PCa vs. controls	**0.023**	0.068	0.123	0.192
advanced PCa vs. localized PCa	0.360	0.360	0.128	0.192
P30–100bp/P163–169bp				
Kruskal-wallis test	**0.035**		0.231	
localized PCa vs. controls	0.052	0.077	0.865	0.865
advanced PCa vs. controls	**0.016**	**0.049**	0.110	0.182
advanced PCa vs. localized PCa	0.308	0.308	0.122	0.182
P30–150bp/P150–300bp				
kruskal wallis test	0.396		0.541	
localized PCa vs. controls	0.282	0.423	0.801	0.801
advanced PCa vs. controls	0.220	0.423	0.393	0.589
advanced PCa vs. localized PCa	0.649	0.649	0.270	0.589
P90–150bp/P163–169bp				
Kruskal-wallis test	0.121		0.170	
localized PCa vs. controls	0.125	0.187	0.830	0.830
advanced PCa vs. controls	0.056	0.169	0.077	0.137
advanced PCa vs. localized PCa	0.415	0.415	0.091	0.137
P160–180bp/P250–420bp				
Kruskal-wallis test	0.382		0.421	
localized PCa vs. controls	0.246	0.368	0.218	0.507
advanced PCa vs. controls	0.230	0.368	0.338	0.507
advanced PCa vs. localized PCa	0.720	0.720	0.986	0.986
10-bp oscillation score (30–150bp fragment length range)				
Kruskal-wallis test	**0.049**		0.275	
localized PCa vs. controls	**0.048**	0.072	0.370	0.426
advanced PCa vs. controls	**0.028**	0.072	0.426	0.426

advanced PCa vs. localized PCa	0.441	0.441	0.119	0.358
10-bp oscillation score (150–300bp fragment length range)				
Kruskal-wallis test	0.213		0.320	
localized PCa vs. controls	0.259	0.388	0.160	0.406
advanced PCa vs. controls	0.484	0.484	0.271	0.406
advanced PCa vs. localized PCa	0.107	0.322	0.971	0.971

Table A8: Overview of ctDNA-positive plasma and urine samples based on cfDNA fragmentation features. CtDNA positivity was assessed using the 10-bp oscillation score in plasma cfDNA and the proportion of fragments within the range 163–169 bp (P163–169 bp) in urinary cfDNA. CtDNA detection thresholds were defined based on the 5th-percentile (plasma) and 95th-percentile (urine) of values obtained from control samples. The maximum values for P163–169 bp of the entire tumor and control cohort were identified in two control samples. Grey shading indicates tumor samples that exceeded alternative threshold definitions influenced by extreme values in the control cohort: light grey denotes tumor samples classified as positive when one extreme control value was not considered, whereas dark grey denotes tumor samples classified as positive when both extreme control values were not considered.

plasma cfDNA

patient ID	10-bp oscillation score	tumor stage	PSA, ng/ml	TNM	Gleason score
T-21_024D	0.0010	localized PCa (intermediate risk)	11.6	cN0 cM0	7a
T-21_186D	0.0010	localized PCa (intermediate risk)	11.5	pT2c pNx (cN0 cM0) L0 V0 Pn0	7a
T-22_298D	0.0010	advanced PCa (M1)	18.	cT2a cN1 cM1b	7b
T-22_069D	0.0010	localized PCa (intermediate risk)	6.2	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a
T-21_030D	0.0009	advanced PCa (M1)	249.0	cT4 cN1 cM1	8
T-21_268D	0.0009	localized PCa (intermediate risk)	8.6	cN0 cMx	7b
T-21_038E	0.0009	localized PCa (intermediate risk)	3.2	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a
T-22_002E	0.0009	localized PCa (intermediate risk)	7.7	pT3a pNx (cN0 cM0) L0 V0 Pn1	7a
T-21_158D	0.0009	localized PCa (intermediate risk)	4.8	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a
T-21_089D	0.0009	advanced PCa (M1)	30.9	cT4 cN1 cM1b	8
T-22_200D	0.0009	advanced PCa (M1)	11	pT3b pN1 (2/21) cM1b L0 V0 Pn1	7a
T-21_209E	-0.0097	localized PCa (high risk)	9.5	pT3b pN0 (0/22) cM0 L0 V0 Pn1	9

urinary cfDNA

patient ID	P163–169bp	tumor stage	PSA, ng/ml	TNM	Gleason score
T-22_018E	0.051	advanced PCa (N1M0)	19.3	pT3b pN1 (2/22) cM0 L1 V0 Pn1	9
T-21_052D	0.049	localized PCa (high risk)	9.6	cT2c cN0 cM0	9
T-21_218E	0.049	localized PCa (intermediate risk)	7.8	pT2c pNx (cN0 cM0) L0 V0 Pn0	7a
T-21_069_1D	0.048	localized PCa (intermediate risk)	7.4	pT2c pN0 (0/13) cM0 L1 V0 Pn1	7b
T-22_052E	0.047	advanced PCa (M1)	7.1	cT4 cN1 cM1	10
T-21_069D	0.044	localized PCa (low risk)	6.4	cT1c cN0 cM0	6
T-22_015E	0.044	advanced PCa (M1)	90	cT2a cN0 cM1	9
T-21_060E	0.043	localized PCa (intermediate risk)	8.1	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a

T-22_200D	0.043	advanced PCa (M1)	11	pT3b pN1 (2/21) cM1b L0 V0 Pn1	7a
T-21_143D	0.041	localized PCa (intermediate risk)	5.1	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a
T-21_259E	0.041	advanced PCa (N1M0)	7.5	pT3b pN1 (1/23) cM0 L1 V0 Pn1	9

Abbreviations: cM = clinical assessment of distant metastases, cN/pN = clinical/pathological assessment of lymph node metastases, c/pT = clinical/pathological assessment of the primary tumor's extent, ctDNA = circulating tumor-derived DNA, L0/L1 = absence/presence of lymphatic invasion, M0/M1 = absence/presence of distant metastases, N0/N1 = absence/presence of lymph node metastases, Nx = unknown lymph node status, Pn0/Pn1 = absence/presence of perineural invasion, PSA = prostate-specific antigen, V0/V1 = absence/presence of venous invasion

Table A9: Estimated TFx values based on CNV analyses without and with *in silico* size selection in plasma and urinary cfDNA.
Summary of TFx values derived from CNV analyses of plasma and urinary cfDNA from PCa patients and controls, including stratification of PCa patients into localized and advanced disease. Results are reported as median TFx with minimum and maximum values and interquartile range. TFx values are shown for CNV analyses without and with *in silico* size selection (90–150 bp) in plasma cfDNA, as well as for CNV analysis without size selection in urinary cfDNA. For plasma cfDNA, the absolute increases in TFx values following *in silico* size selection are reported, together with the corresponding statistical comparisons between TFx values obtained from the analyses with and without size selection, performed using Wilcoxon signed-rank test.

plasma cfDNA				
	TFx, median [min, max]	**TFx, IQR**	**absolute TFx increase** median (TFx, size selection – TFx, w/o size selection)	**TFx, size selection vs. TFx, w/o size selection** p value
all samples				
without size selection	0.0 [0.0, 0.464]	0.0, 0.010	0.007	3.41e-10
with size selection (90–150bp)	0.012 [0.0, 0.577]	0.006, 0.017		
control				
without size selection	0.0 [0.0, 0.040]	0.0, 0.010	0.005	1.41e-04
with size selection (90–150bp)	0.012 [0.0, 0.056]	0.005, 0.016		
tumor all				
without size selection	0.0 [0.0, 0.464]	0.0, 0.011	0.009	2.07e-07
with size selection (90–150bp)	0.012 [0.0, 0.577]	0.008, 0.017		
localized PCa				
without size selection	0.0 [0.0, 0.024]	0.0, 0.001	0.008	5.49e-06
with size selection (90–150bp)	0.012 [0.0, 0.059]	0.009, 0.017		
advanced PCa				
without size selection	0.0 [0.0, 0.464]	0.0, 0.011	0.010	1.05e-02
with size selection (90–150bp)	0.012 [0.0, 0.577]	0.002, 0.016		

urinary cfDNA				
	TFx, median [min, max]	**TFx, IQR**	**X**	**X**
all samples				
without size selection	0.014 [0.0, 0.678]	0.006, 0.027	X	X
control				
without size selection	0.013 [0.0, 0.064]	0.0, 0.020	X	X
tumor all				
without size selection	0.017 [0.0, 0.678]	0.008, 0.035	X	X
localized PCa				
without size selection	0.014 [0.0, 0.127]	0.006, 0.028	X	X

advanced PCa				
without size selection	0.026 [0.003, 0.678]	0.013, 0.039	X	X

Abbreviations: IQR = interquartile range, TFx = tumor fraction

Table A10: Overview of CNV analyses with *in silico* size selection across 15 fragment length ranges in urinary cfDNA.

Summary of CNV analysis results for urinary cfDNA from PCa patients and controls following *in silico* size selection across 15 different fragment length ranges. Reported metrics include the median TFx values and the statistical comparison between the TFx values derived from the CNV analyses with and without size selection, performed using Wilcoxon signed-rank test. Correction for multiple testing was applied using the Benjamini-Hochberg method. In addition, three comparative metrics are reported: (i) the median of per-sample TFx ratios calculated as $TFx_{with\ size\ selection}/TFx_{without\ size\ selection}$; (ii) the median of per-sample absolute TFx differences calculated as $TFx_{with\ size\ selection} - TFx_{without\ size\ selection}$; and (iii) the ratio of the absolute TFx difference in tumor samples relative to the absolute TFx difference in control samples.

size selection	median TFx	comparison to analysis w/o size selection p value, adj. p value	median (TFx_{with} + 0.001 / $TFx_{w/o}$ + 0.001) ratio	median ($TFx_{with} - TFx_{w/o}$) abs. TFx diff.	abs. TFx diff.$_{tumor}$ / abs. TFx diff.$_{control}$ ratio
control samples (n = 35)					
without size selection	0.013	X	X	X	X
20–100 bp	0.015	0.681 (0.730)	1.00	0.00003	X
20–110 bp	0.012	0.342 (0.631)	1.00	0.0	X
20–120 bp	0.012	0.371 (0.631)	1.00	0.0	X
20–130 bp	0.011	0.340 (0.631)	1.00	0.0	X
20–140 bp	0.012	0.257 (0.631)	0.91	-0.00124	X
20–150 bp	0.012	0.277 (0.631)	0.96	-0.00055	X
20–160 bp	0.010	0.620 (0.715)	1.00	0.0	X
20–170 bp	0.013	0.527 (0.659)	1.03	0.00085	X
90–150 bp	0.013	0.888 (0.888)	1.06	0.00110	X
110–170 bp	0.014	0.453 (0.631)	1.07	0.00084	X
140–170 bp	0.015	0.125 (0.631)	1.01	0.00055	X
140–180 bp	0.014	0.385 (0.631)	1.19	0.00119	X
140–190 bp	0.014	0.462 (0.631)	1.16	0.00243	X
140–200 bp	0.014	0.148 (0.631)	1.22	0.00373	X
140–210 bp	0.014	0.181 (0.631)	1.17	0.00292	X
tumor samples (n = 67)					
without size selection	0.017	X	X	X	X
20–100 bp	0.016	0.506 (0.759)	0.98	-0.00031	-10.33
20–110 bp	0.016	0.039 (0.146)	0.97	-0.00055	x
20–120 bp	0.016	0.394 (0.656)	1.00	0.0	x
20–130 bp	0.016	0.151 (0.376)	0.97	-0.00056	x
20–140 bp	0.016	0.229 (0.490)	0.95	-0.00106	0.86

20–150 bp	0.015	0.034 (0.146)	0.95	-0.00148	2.69
20–160 bp	0.016	0.117 (0.350)	0.91	-0.00230	-2300
20–170 bp	0.017	0.035 (0.146)	0.89	-0.00301	-3.54
90–150 bp	0.015	0.007 (0.098)	0.88	-0.00418	-3.80
110–170 bp	0.016	0.651 (0.836)	1.00	0.0	0.0
140–170 bp	0.018	0.978 (0.978)	1.00	0.0	0.0
140–180 bp	0.019	0.731 (0.844)	1.00	0.0	0.0
140–190 bp	0.017	0.877 (0.940)	1.00	0.0	0.0
140–200 bp	0.017	0.261 (0.490)	1.07	0.00170	0.46
140–210 bp	0.017	0.669 (0.836)	1.01	0.00022	0.08

Abbreviations: abs. TFx diff. = absolute tumor fraction difference

Table A11: Overview of ctDNA-positive plasma and urine samples based on CNV analysis.
CtDNA positivity was assessed using TFx values derived from the CNV analyses without and with *in silico* size selection for plasma cfDNA and from the CNV analysis without size selection for urinary cfDNA. CtDNA detection thresholds were defined based on the 95th-percentile of TFx values from controls. PCa patients with matched ctDNA-positive plasma and urine are highlighted in bold.

plasma cfDNA

patient ID	TFx	tumor stage	PSA, ng/ml	TNM	Gleason score
without *in silico* size selection					
T-22_015E	0.464	advanced PCa (M1)	90	cT2a cN0 cM1	9
T-21_089D	0.160	advanced PCa (M1)	30.9	cT4 cN1 cM1b	8
T-22_003D	0.024	localized PCa (high risk)	15	cN0 cMx	9
T-22_172E	0.022	localized PCa (intermediate risk)	4.31	pT2c pN0 (0/12) cM0 L0 V0 Pn1	7a
with *in silico* size selection (90–150 bp)					
T-22_015E	0.577	advanced PCa (M1)	90	cT2a cN0 cM1	9
T-21_089D	0.211	advanced PCa (M1)	30.9	cT4 cN1 cM1b	8
T-21_060E	0.059	localized Pca (intermediate risk)	8.1	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a
T-21_120D	0.048	advanced PCa (N1M0)	4.3	cT4 cN1 cM0	8
T-22_008D	0.036	localized PCa (intermediate risk)	11	cT2c cN0	7b
T-21_045E	0.033	localized PCa (intermediate risk)	4.3	pT3a pN0 (0/19) cM0 L0 V0 Pn1	7a
T-22_002E	0.028	localized PCa (intermediate risk)	7.7	pT3a pNx (cN0 cM0) L0 V0 Pn1	7a
T-21_259E	0.026	advanced PCa (N1M0)	7.5	pT3b pN1 (1/23) cM0 L1 V0 Pn1	9
T-21_234D	0.025	localized PCa (low risk)	9.1	cN0 cMx	6

urinary cfDNA

patient ID	TFx	tumor stage	PSA, ng/ml	TNM	Gleason score
without *in silico* size selection					
T-22_200D	0.678	advanced PCa (M1)	11	pT3b pN1 (2/21) cM1b L0 V0 Pn1	7a
T-21_120D	0.182	advanced PCa (N1M0)	4.3	cT4 cN1 cM0	8
T-21_228D	0.127	localized PCa (intermediate risk)	5.5	cT1c cN0 cMx	7a
T-21_131D	0.126	localized PCa (intermediate risk)	8.1	cT1c cN0 cM0	7a
T-21_259E	0.126	advanced PCa (N1M0)	7.5	pT3b pN1 (1/23) cM0 L1 V0 Pn1	9
T-22_074E	0.116	localized PCa (intermediate risk)	5.6	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a
T-21_045E	0.108	localized PCa (intermediate risk)	4.3	pT3a pN0 (0/19) cM0 L0 V0 Pn1	7a
T-21_060E	0.050	localized PCa (intermediate risk)	8.1	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a

Table A12: Overview of ctDNA-positive plasma and urine samples of PCa patients based on the CIA score. CtDNA positivity was assessed using CIA scores derived from the assessment of chromosomal instability without and with *in silico* size selection (90–150 bp) for plasma cfDNA and the assessment without size selection for urinary cfDNA. CtDNA detection thresholds were defined based on the respective 95th-percentile of CIA scores obtained from control samples. PCa patients with matched ctDNA-positive plasma and urine samples are highlighted in bold.

plasma cfDNA

patient ID	CIA score	tumor stage	PSA, ng/ml	TNM	Gleason score
without *in silico* size selection					
T-22_015E	2162.3	advanced PCa (M1)	90	cT2a cN0 cM1	9
T-21_089D	772.7	advanced PCa (M1)	30.9	cT4 cN1 cM1b	8
T-22_016D	305.9	localized PCa (high risk)	9.6	cN0 cMx	8
T-21_060E	291.6	localized PCa (intermediate risk)	8.1	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a
T-22_008D	279.0	localized PCa (intermediate risk)	11	cT2c cN0	7b
with *in silico* size selection (90–150 bp)					
T-22_015E	1057.3	advanced PCa (M1)	90	cT2a cN0 cM1	9
T-21_089D	431.5	advanced PCa (M1)	30.9	cT4 cN1 cM1b	8
T-22_016D	272.9	localized PCa (high risk)	9.6	cN0 cMx	8
T-21_060E	250.9	localized PCa (intermediate risk)	8.1	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a

urinary cfDNA

patient ID	CIA score	tumor stage	PSA, ng/ml	TNM	Gleason score
without *in silico* size selection					
T-22_200D	2788.4	advanced PCa (M1)	11	pT3b pN1 (2/21) cM1b L0 V0 Pn1	7a
T-21_120D	1664.9	advanced PCa (N1M0)	4.3	cT4 cN1 cM0	8
T-21_259E	538.8	advanced PCa (N1M0)	7.5	pT3b pN1 (1/23) cM0 L1 V0 Pn1	9
T-21_131D	515.7	localized PCa (intermediate risk)	8.1	cT1c cN0 cM0	7a
T-21_030D	512.0	advanced PCa (M1)	249	cT4 cN1 cM1	8
T-21_208D	418.5	localized PCa (high risk)	26.4	pT3b pN0 (0/14) cM0	9
T-21_083D	398.1	localized PCa (intermediate risk)	2.7	pT2a pN0 (0/16) cM0 L0 V0 Pn0	7a
T-22_022D	385.7	localized PCa (intermediate risk)	5.1	pT3a pNx (cN0 cM0) L0 V0 Pn1	7a
T-21_045E	348.5	localized PCa (intermediate risk)	4.3	pT3a pN0 (0/19) cM0 L0 V0 Pn1	7a
T-22_018E	310.4	advanced PCa (N1 M0)	19.3	pT3b pN1 (2/22) cM0 L1 V0 Pn1	9
T-22_082D	304.2	localized PCa (intermediate risk)	8.93	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a
T-22_052E	296.4	advanced PCa (M1)	7.1	cT4 cN1 cM1	10

T-21_089D	293.3	advanced PCa (M1)	30.9	cT4 cN1 cM1b	8
T-21_228D	285.2	localized PCa (intermediate risk)	5.5	cT1c cN0 cMx	7a

Abbreviations: CIA = chromosomal instability analysis

Table A13: Overview of differential methylation analysis results in plasma and urinary cfDNA across PCa patient groups and controls.
Summary of results from differential methylation analyses performed on plasma and urinary cfDNA from PCa patients and controls, as well as across PCa subgroups. DMRs were identified based on absolute differences in methylation levels (β-values) calculated in 300-bp genomic windows between the compared cohorts, including only windows with > 2 NRPKMs in at least one sample. Correction for multiple testing was applied using the Benjamini-Hochberg method. Adapted from Riediger *et al.* [410]

comparison	**# samples** total, cohort 1 vs. cohort 2	**PLASMA** # significant windows (DMRs)	**URINE** # significant windows (DMRs)
PCa (all) vs. all controls	plasma: 109 (73 vs. 36) urine: 102 (67 vs. 35)	0 / 1,168,518	0 / 1,195,872
advanced PCa (M1) vs. all controls	plasma: 45 (9 vs. 36) urine: 44 (9 vs. 35)	712 / 980,927	48 / 992,774
advanced PCa (all) vs. all controls	plasma: 54 (18 vs. 36) urine: 53 (18 vs. 35)	0 / 1,008,208	2 / 1,030,695
localized PCa high risk vs. all controls	plasma: 45 (9 vs. 36) urine: 44 (9 vs. 35)	0 / 945,422	0 / 980,210
localized PCa intermediate risk vs. all controls	plasma: 78 (42 vs. 36) urine: 72 (37 vs. 35)	0 / 1,112,508	0 / 1,127,546
localized PCa (all) vs. all controls	plasma: 91 (55 vs. 36) urine: 84 (49 vs. 35)	0 / 1,134,888	0 / 1,152,290
localized PCa (all) + advanced PCa (N1 M0) vs. all controls	plasma: 100 (64 vs. 36) urine: 93 (58 vs. 35)	0 / 1,147,368	0 / 1,171,278
advanced PCa (M1) vs. advanced PCa (N1 M0)	plasma: 18 (9 vs. 9) urine: 18 (9 vs. 9)	2 / 838,576	2 / 821,533
advanced PCa (all) vs. localized PCa (all)	plasma: 73 (18 vs. 55) urine: 67 (18 vs. 49)	0 / 1,115,932	0 / 1,112,691
advanced PCa (M1) vs. localized PCa (all) + advanced PCa (N1 M0)	plasma: 73 (9 vs. 64) urine: 67 (9 vs. 58)	890 / 1,115,932	64 / 1,112,691
advanced PCa (M1) vs. localized PCa intermediate risk	plasma: 51 (9 vs. 42) urine: 46 (9 vs. 37)	309 / 1,064,258	18 / 1,048,296
advanced PCa (all) vs. localized PCa intermediate risk	plasma: 60 (18 vs. 37) urine: 55 (18 vs. 37)	0 / 1,086,590	0 / 1,081,990
advanced PCa (M1) vs. localized PCa high risk	plasma: 18 (9 vs. 9) urine: 18 (9 vs. 9)	29 / 830,239	0 / 806,021
advanced PCa (all) vs. localized PCa high risk	plasma: 27 (18 vs. 9) urine: 27 (18 vs. 9)	0 / 895,667	0 / 889,649

Abbreviations: # = number of, DMR = differentially methylated region, NRPKM = normalized reads per kilobase million mapped reads

Table A14: Repetition of the differential methylation analysis in plasma and urinary cfDNA following exclusion of individual tumor samples.
Summary of detectable DMRs identified in plasma and urinary cfDNA when comparing advanced PCa patients with distant metastases (M1) to controls or to PCa patients without distant metastases (N0/N1 M0), respectively, based on repeated DMR analyses in which one to three tumor samples exhibiting pronounced deviations in methylation patterns in prior cfDNA methylation profiling were temporarily excluded. DMRs were identified based on absolute differences in methylation levels (β-values) calculated in 300-bp genomic windows, including only windows with > 2 NRPKMs in at least one sample. Correction for multiple testing was performed using the Benjamini-Hochberg method.

plasma cfDNA		
excluded tumor sample(s)	**advanced PCa (M1)** vs. **all controls** # significant windows (DMRs)	**advanced PCa (M1)** vs. **localized PCa (all) + advanced PCa (N1M0)** # significant windows (DMRs)
none	**712** / 980,927	**890** / 1,115,932
"T-22_015E", ”T-21_089D", "T-21_041D"	1 / 935,929	0 / 1,097,243
"T-22_015E", "T-21_089D"	1 / 963,087	0 / 1,105,459
"T-21_089D", "T-21_041D"	148 / 946,979	274 / 1,104,839
"T-22_015E", "T-21_041D"	1 / 948,204	0 / 1,101,038
"T-22_015E"	47 / 971,119	23 / 1,108,727
"T-21_089D"	136 / 973,339	322 / 1,112,882
"T-21_041D"	407 / 958,534	543 / 1,108,386

urinary cfDNA		
excluded tumor sample(s)	**advanced PCa (M1)** vs. **all controls** # significant windows (DMRs)	**advanced PCa (M1)** vs. **localized PCa (all) + advanced PCa (N1M0)** # significant windows (DMRs)
none	**48** / 992,774	**72** / 1,112,691
"T-22_200D", "T-21_030D"	12 / 983,635	1 / 1,105,517
"T-22_200D"	3 / 985,471	12 / 1,106,671
"T-21_030D"	25 / 991,014	19 / 1,111,583

Table A15: Shared hypermethylated DMRs between plasma and urinary cfDNA in metastatic PCa patients compared with controls.
Overview of eight hypermethylated DMRs identified in both plasma and urinary cfDNA when comparing metastatic PCa patients to controls. Genomic annotation (location within 3' or 5'UTR, distal intergenic region, downstream region, exon or intron, promoter region) and CpG-associated landscape distributions (CpG island, shore, shelf, and open sea), as well as associated genes are shown for the overlapping DMRs. Regions highlighted in gray indicate hypermethylated DMRs that were also identified as overlapping DMRs in both plasma and urinary cfDNA when comparing metastatic PCa to PCa patients without distant metastases.

DMR, chr:start-end	**plasma,** logFC	**plasma,** adj. p value	**urine,** logFC	**urine,** adj. p value	**CpG landscape**	**genomic annotation**	**associated gene**	**gene,** chr:start-end (length)	**gene ID**	**distance to TSS**
chr1: 247850401-247850700	2.27	1.84e-02	2.69	2.98e-04	Open Sea	Distal Intergenic	**OR13G1**: olfactory receptor family 13 subfamily G member 1	chr1: 247835420-247836343 (924)	441933	-14058
chr2: 70779901-70780200	1.91	1.56e-02	2.40	2.98e-04	Island	Intron	**TGFA**: transforming growth factor alpha	chr2: 70674412-70780770 (106359)	7039	570
chr3: 181441201-181441500	1.84	8.69e-04	1.78	3.32e-02	Shelf	Intron	**SOX2-OT**: SOX2 overlapping transcript	chr3: 180774468-181460013 (685546)	347689	666733
chr6: 28367101-28367400	2.72	6.80e-06	2.46	2.98e-04	Island	Promoter	**ZSCAN12**: zinc finger and SCAN domain containing 12	chr6: 28346598-28367544 (20947)	9753	144
chr7: 157361401-157361700	1.72	3.24e-02	2.14	1.76e-03	Island	Exon	**PTPRN2**: protein tyrosine phosphatase receptor type N2	chr7: 157331750-158334468 (1002719)	5799	972768
chr7: 157484101-157484400	1.68	1.24e-02	2.07	2.39e-02	Island	Intron	**PTPRN2**: protein tyrosine phosphatase receptor type N2	chr7: 157331750-158334468 (1002719)	5799	850068
chr19: 31846501-31846800	1.43	4.77e-02	1.96	2.39e-02	Shore	Distal Intergenic	**TSHZ3**: teashirt zinc finger homeobox 3	chr19: 31765851-31840190 (74340)	57616	-6311
chr20: 62461201-62461500	1.93	7.61e-03	2.14	9.33e-03	Island	Intron	**ZBTB46**: zinc finger and BTB domain containing 46	chr20: 62375875-62462597 (86723)	140685	1097

Abbreviation: logFC = log fold change, TSS = transcription starting site

Table A16: Shared hypermethylated DMRs between plasma and urinary cfDNA in metastatic PCa patients compared with PCa patients without distant metastases.
Overview of seven hypermethylated DMRs identified in **both plasma and urinary cfDNA** when comparing metastatic PCa patients to PCa patients without distant metastases. Genomic annotation (location within 3' or 5'UTR, distal intergenic region, downstream region, exon or intron, promoter region) and CpG-associated landscape distributions (CpG island, shore, shelf, and open sea), as well as associated genes are shown for the overlapping DMRs. Regions highlighted in grey indicate hypermethylated DMRs that were also identified as overlapping DMRs in both plasma and urinary cfDNA when comparing metastatic PCa to controls.

DMR, chr:start-end	**plasma,** logFC	**plasma,** adj. p value	**urine,** logFC	**urine,** adj. p value	**CpG landscape**	**genomic annotation**	**associated gene**	**gene,** chr:start-end (length)	**gene ID**	**distance to TSS**
chr2: 70779901-70780200	1.92	4.50e-04	2.04	3.07e-02	Island	Intron	**TGFA:** transforming growth factor alpha	chr2: 70674412-70780770 (106359)	7039	570
chr3: 181441201-181441500	1.50	5.49e-03	1.63	3.07e-02	Shelf	Intron	**SOX2-OT:** SOX2 overlapping transcript	chr3: 180774468-181460013 (685546)	347689	666733
chr3: 181441501-181441800	1.59	1.84e-03	1.62	4.45e-02	Shelf	Intron	**SOX2-OT:** SOX2 overlapping transcript	chr3: 180774468-181460013 (685546)	347689	667033
chr7: 157361401-157361700	1.87	1.49e-04	1.81	4.65e-02	Island	Exon	**PTPRN2:** protein tyrosine phosphatase receptor type N2	chr7: 157331750-158334468 (1002719–2)	5799	972768
chr10: 72200701-72201000	1.74	1.94e-03	1.70	4.65e-02	Island	Promoter	**NODAL:** nodal growth differentiation factor	chr10: 72191692-72201465 (9774–2)	4838	465
chr14: 65008801-65009100	1.49	5.27e-04	2.29	5.73e-04	Island	Exon	**HSPA2:** heat shock protein family A (Hsp70) member 2	chr14: 65007186-65009954 (2769–1)	3306	1615
chr20: 62461201-62461500	1.84	8.66e-04	1.68	4.65e-02	Island	Intron	**ZBTB46:** zinc finger and BTB domain containing 46	chr20: 62375875-62462597 (86723–2)	140685	1097

Table A17: Mean absolute methylation levels in PCa tissue-derived marker regions across plasma and urinary cfDNA.

Mean absolute methylation levels (β-values) across localized PCa patients, advanced PCa patients, and controls are shown for 67 PCa tissue–derived methylation marker regions in plasma and urinary cfDNA, respectively. The table lists all 67 genomic 300-bp regions together with the corresponding mean β-values for each cohort in plasma or urine. The last row reports the median methylation scores across the three cohorts in plasma and urinary cfDNA. Methylation scores were calculated for each sample as the median of β-values across all 67 marker regions.

	plasma cfDNA			**urinary cfDNA**		
genomic position	**controls,** mean β-value	**localized PCa,** mean β-value	**advanced PCa,** mean β-value	**controls,** mean β-value	**localized PCa,** mean β-value	**advanced PCa,** mean β-value
chr1_15480901_15481200	0.027	0.028	0.044	0.031	0.033	0.060
chr1_24648601_24648900	0.023	0.023	0.049	0.028	0.036	0.055
chr1_119526901_119527200	0.034	0.039	0.062	0.041	0.052	0.076
chr1_119527201_119527500	0.027	0.032	0.062	0.047	0.049	0.096
chr1_119528401_119528700	0.078	0.076	0.153	0.123	0.136	0.185
chr1_119528701_119529000	0.058	0.067	0.111	0.091	0.101	0.153
chr1_119543101_119543400	0.046	0.043	0.120	0.065	0.067	0.093
chr1_146551801_146552100	0.042	0.039	0.065	0.044	0.054	0.095
chr1_197889301_197889600	0.042	0.040	0.072	0.061	0.065	0.099
chr1_203598301_203598600	0.045	0.046	0.090	0.050	0.059	0.080
chr2_220117501_220117800	0.045	0.046	0.077	0.044	0.041	0.080
chr2_237079501_237079800	0.065	0.056	0.101	0.200	0.228	0.275
chr3_68980801_68981100	0.030	0.041	0.075	0.036	0.041	0.084
chr3_154145401_154145700	0.062	0.070	0.125	0.087	0.088	0.123
chr3_154145701_154146000	0.033	0.032	0.087	0.035	0.042	0.090
chr3_170746201_170746500	0.031	0.027	0.057	0.032	0.033	0.073
chr4_11429401_11429700	0.053	0.057	0.100	0.055	0.070	0.095
chr4_16085401_16085700	0.028	0.030	0.045	0.035	0.030	0.044
chr4_16085701_16086000	0.057	0.070	0.094	0.075	0.091	0.141
chr4_41868301_41868600	0.077	0.065	0.094	0.067	0.073	0.136
chr4_41882401_41882700	0.048	0.051	0.067	0.063	0.064	0.109
chr4_85402201_85402500	0.040	0.042	0.093	0.070	0.065	0.126
chr4_85402501_85402800	0.034	0.047	0.054	0.054	0.058	0.064
chr4_85414501_85414800	0.028	0.028	0.062	0.038	0.039	0.063
chr4_85414801_85415100	0.035	0.035	0.066	0.045	0.045	0.104
chr4_185937001_185937300	0.058	0.065	0.090	0.134	0.158	0.158
chr6_28367101_28367400	0.028	0.035	0.097	0.034	0.041	0.134
chr6_28367401_28367700	0.025	0.025	0.064	0.034	0.035	0.082
chr6_28367701_28368000	0.042	0.037	0.099	0.058	0.058	0.124
chr6_29974801_29975100	0.024	0.030	0.045	0.036	0.032	0.049
chr6_56818501_56818800	0.041	0.042	0.067	0.066	0.055	0.105
chr6_56818801_56819100	0.031	0.032	0.058	0.044	0.052	0.105
chr6_137809501_137809800	0.030	0.032	0.042	0.037	0.041	0.073
chr7_19146001_19146300	0.036	0.042	0.071	0.047	0.061	0.068
chr7_19146301_19146600	0.033	0.033	0.065	0.045	0.044	0.068
chr7_29185501_29185800	0.038	0.036	0.066	0.037	0.040	0.091
chr7_45613501_45613800	0.043	0.045	0.064	0.044	0.040	0.067
chr7_96632101_96632400	0.099	0.102	0.121	0.112	0.115	0.193
chr7_97360801_97361100	0.042	0.050	0.081	0.083	0.083	0.104
chr7_97361101_97361400	0.029	0.031	0.066	0.033	0.038	0.069
chr7_129422101_129422400	0.034	0.039	0.075	0.040	0.044	0.090
chr7_157478701_157479000	0.031	0.027	0.040	0.033	0.034	0.044
chr7_157481401_157481700	0.037	0.040	0.055	0.036	0.038	0.061
chr7_157484101_157484400	0.031	0.033	0.068	0.035	0.040	0.111
chr8_70947301_70947600	0.037	0.037	0.056	0.049	0.068	0.109
chr9_112810501_112810800	0.051	0.051	0.081	0.042	0.045	0.075

chr9_126777901_126778200	0.046	0.048	0.071	0.051	0.052	0.085
chr9_135462001_135462300	0.032	0.037	0.057	0.042	0.044	0.070
chr9_135620401_135620700	0.063	0.072	0.103	0.231	0.281	0.272
chr9_135620701_135621000	0.057	0.066	0.079	0.183	0.206	0.169
chr10_102895801_102896100	0.051	0.054	0.085	0.054	0.059	0.087
chr10_102905701_102906000	0.042	0.041	0.054	0.051	0.042	0.065
chr11_62691001_62691300	0.040	0.049	0.063	0.043	0.046	0.062
chr11_62691301_62691600	0.039	0.044	0.066	0.033	0.040	0.053
chr12_54441301_54441600	0.037	0.040	0.049	0.142	0.169	0.157
chr12_54447601_54447900	0.032	0.041	0.055	0.118	0.144	0.126
chr12_65220301_65220600	0.040	0.042	0.084	0.053	0.053	0.108
chr12_104852101_104852400	0.024	0.025	0.043	0.033	0.033	0.070
chr12_122016901_122017200	0.040	0.036	0.054	0.042	0.047	0.060
chr13_100641301_100641600	0.028	0.037	0.070	0.041	0.046	0.081
chr13_100641601_100641900	0.045	0.051	0.083	0.072	0.082	0.090
chr17_41363701_41364000	0.033	0.036	0.060	0.037	0.043	0.076
chr18_56939401_56939700	0.036	0.033	0.049	0.035	0.040	0.066
chr20_20345701_20346000	0.069	0.052	0.102	0.056	0.070	0.094
chr20_37356301_37356600	0.049	0.057	0.089	0.067	0.066	0.125
chr20_50721001_50721300	0.029	0.032	0.049	0.034	0.036	0.056
chr20_50721301_50721600	0.026	0.030	0.046	0.037	0.034	0.071

	plasma cfDNA			**urinary cfDNA**		
	controls, median, range	**localized PCa,** median, range	**advanced PCa**, median, range	**controls,** median, range	**localized PCa,** median, range	**advanced PCa**, median, range
methylation score	0.032 (0.015–0.054)	0.033 (0.019–0.084)	0.036 (0.025–0.363)	0.039 (0.020–0.085)	0.041 (0.022–0.143)	0.059 (0.025–0.318)

Table A18: Overview of ctDNA-positive plasma and urine samples based on the methylation analysis.
CtDNA positivity was assessed using methylation scores, calculated as the median of absolute methylation levels (β-values) in 67 PCa tissue-derived marker regions, across plasma and urine samples from localized and advanced PCa patients. CtDNA detection thresholds were defined based on the respective 95th-percentile of methylation scores obtained from control samples. PCa patients with matched ctDNA-positive plasma and urine samples are highlighted in bold.

plasma cfDNA

patient ID	methylation score	tumor stage	PSA, ng/ml	TNM	Gleason score
T-22_015E	0.363	advanced PCa (M1)	90	cT2a cN0 cM1	9
T-21_089D	0.144	advanced PCa (M1)	30.9	cT4 cN1 cM1b	8
T-21_024D	0.084	localized PCa (intermediate risk)	11.6	cN0 cM0	7a
T-21_266E	0.068	advanced PCa (N1M0)	4.1	pT3a pN1 (1/12) cM0 L0 V0 Pn1	7b
T-21_209E	0.059	localized PCa (high risk)	9.5	pT3b pN0 (0/22) cM0 L0 V0 Pn1	9
T-22_082D	0.058	localized PCa (intermediate risk)	8.93	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a
T-22_069D	0.058	localized PCa (intermediate risk)	6.2	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a
T-22_137D	0.057	localized PCa (intermediate risk)	6.3	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a
T-22_200D	0.054	advanced PCa (M1)	11	pT3b pN1 (2/21) cM1b L0 V0 Pn1	7a
T-21_069_1D	0.054	localized PCa (intermediate risk)	7.4	pT2c pN0 (0/13) cM0 L1 V0 Pn1	7b
T-21_175D	0.048	localized PCa (high risk)	21.4	pT3a pN0 (0/22) cM0 L0 V0 pN1	9
T-22_298D	0.048	advanced PCa (M1)	18.3	cT2a cN1 cM1b	7b

urinary cfDNA

patient ID	methylation score	tumor stage	PSA, ng/ml	TNM	Gleason score
T-22_200D	0.318	advanced PCa (M1)	11	pT3b pN1 (2/21) cM1b L0 V0 Pn1	7a
T-21_030D	0.148	advanced PCa (M1)	249	cT4 cN1 cM1	8
T-21_131D	0.143	localized PCa (intermediate risk)	8.1	cT1c cN0 cM0	7a
T-21_259E	0.128	advanced PCa (N1M0)	7.5	pT3b pN1 (1/23) cM0 L1 V0 Pn1	9
T-21_266E	0.088	advanced PCa (N1M0)	4.1	pT3a pN1 (1/12) cM0 L0 V0 Pn1	7b
T-22_298D	0.086	advanced PCa (M1)	18.3	cT2a cN1 cM1b	7b
T-21_209E	0.085	localized PCa (high risk)	9.5	pT3b pN0 (0/22) cM0 L0 V0 Pn1	9
T-22_052E	0.083	advanced PCa (M1)	7.1	cT4 cN1 cM1	10
T-22_018E	0.077	advanced PCa (N1M0)	19.3	pT3b pN1 (2/22) cM0 L1 V0 Pn1	9
T-21_120D	0.073	advanced PCa (N1M0)	4.3	cT4 cN1 cM0	8

T-22_082D	0.069	localized PCa (intermediate risk)	8.93	pT2c pNx (cN0 cM0) L0 V0 Pn1	7a
T-22_016D	0.068	localized PCa (high risk)	9.6	cN0 cMx	8

Table A19: CtDNA positivity based on complementary genomic and epigenomic cfDNA features in plasma and urinary cfDNA.
CtDNA positivity was determined based on four complementary analyses of (epi)genomic features assessed in plasma and urinary cfDNA: (1) estimated TFx, (2) CIA score, (3) methylation score, and (4) cfDNA fragmentation features (10-bp oscillation score in plasma cfDNA and the proportion of fragments within the 163–169 bp range in urinary cfDNA). The table summarizes the number of localized and advanced PCa patients with positive signals in one to three out of four analyses, stratified by detection in plasma only, urine only, in either plasma or urine, or in both biofluids. Sample availability: localized PCa (plasma, n = 55; urine, n = 50) and advanced PCa (plasma and urine, n = 18). Adapted from Riediger *et al.* [410]

	tumor status	plasma	urine	plasma or urine	matched plasma and urine
1 out of 4 analyses (1/4)	*localized PCa*	11 / 55 (20%)	7 / 50 (14%)	14 / 55 (25%)	**localized PCa:** plasma, 1/4 + urine, 1/4 (n = 1) plasma, 2/4 + urine, 1/4 (n = 2) **advanced PCa:** plasma, 1/4 + urine, 1/4 (n = 1) plasma, 1/4 + urine, 3/4 (n = 2) plasma, 2/4 + urine, 1/4 (n = 1) plasma, 2/4 + urine 3/4 (n = 1) plasma, 4/4 + urine, 1/4 (n = 1)
	advanced PCa	4 / 18 (22%)	3 / 18 (17%)	1 / 18 (6%)	
2 out of 4 analyses (2/4)	*localized PCa*	6 / 55 (11%)	0 / 50	4 / 55 (7%)	
	advanced PCa	2 / 18 (11%)	0 / 18	0 / 18	
3 out of 4 analyses (3/4)	*localized PCa*	0 / 55	1 / 50 (2%)	1 / 55 (2%)	
	advanced PCa	1 / 18 (6%)	3 / 18 (17%)	1 / 18 (6%)	
4 out of 4 analyses (4/4)	*localized PCa*	0 / 55	0 / 50	0 / 55	
	advanced PCa	1 / 18 (6%)	0 / 18	0 / 18	

Abbreviations: n = number

Table A20: Clinical and tumor characteristics of PCa patients with ctDNA-positivity based on multimodal liquid biopsy analyses.
The table summarizes initial PSA levels, Gleason score, and TNM staging of localized and advanced PCa patients with detectable ctDNA, as identified by multimodal liquid biopsy analyses. CtDNA positivity was determined based on complementary genomic and epigenomic cfDNA features assessed in plasma and urinary cfDNA.

Patient ID	Gleason Score	PSA level (ng/ml)	TNM stage
ctDNA-positive localized PCa patients			
21_024D	7a	11.6	cT1c cN0 cM0
21_038E	7a	3.2	pT2c cN0 cM0 L0 V0 Pn1
21_045E	7a	4.3	pT3a pN0 (0/19) cM0 L0 V0 Pn1
21_060E	7a	8.1	pT2c cN0 cM0 L0 V0 Pn1
21_069_1D	7b	7.4	pT2c pN0 (0/13) cM0 L1 V0 Pn1
21_083D	7a	2.7	pT2a pN0 (0/16) cM0 L0 V0 Pn0
21_131D	7a	8.1	cT1c cN0 cM0
21_158D	7a	4.8	pT2c cN0 cM0 L0 V0 Pn1
21_175D	9	21.4	pT3a pN0 (0/22) cM0 L0 V0 pN1
21_186D	7a	11.5	pT2c cN0 cM0 L0 V0 Pn0
21_208D	9	26.4	pT3b pN0 (0/14) cM0
21_209E	9	9.5	pT3b pN0 (0/22) cM0 L0 V0 Pn1
21_228D	7a	5.5	cT1c cN0 cMx
21_234D	6	9.1	cT1c cN0 cM0
21_268D	7b	8.6	cT1c cN0 cM0
22_002E	7a	7.7	pT3a cN0 cM0 L0 V0 Pn1
22_008D	7b	11.0	cT2c cN0 cM0
22_016D	8	9.6	cT1c cN0 cM0
22_022D	7a	5.1	pT3a cN0 cM0 L0 V0 Pn1
22_069D	7a	6.2	pT2c cN0 cM0 L0 V0 Pn1
22_074E	7a	5.6	pT2c cN0 cM0 L0 V0 Pn1
22_082D	7a	8.93	pT2c cN0 cM0 L0 V0 Pn1
22_137D	7a	6.3	pT2c cN0 cM0 L0 V0 Pn1
all	G6, n = 1 G7a, n = 15 G7b, n = 3 G8, n = 1 G9, n = 1	median: 8.1	locally-restricted (T1/T2), n = 17 locally-advanced (T3/T4), n = 6
ctDNA-positive advanced PCa patients (N1 M0)			
21_120D	8	4.3	cT4 cN1 cM0
21_259E	9	7.5	pT3b pN1 (1/23) cM0 L1 V0 Pn1
21_266E	7b	4.1	pT3a pN1 (1/12) cM0 L0 V0 Pn1
22_018E	9	19.3	pT3b pN1 (2/22) cM0 L1 V0 Pn1
all	G7b, n = 1 G8, n = 1 G9, n = 2	median: 5.9	locally-advanced (T3/T4), n = 4
ctDNA-positive advanced PCa patients (M1)			
21_030D	8	249.0	cT4 cN1 cM1
21_089D	8	30.9	cT4 cN1 cM1b
22_015E	9	90.0	cT2a cN0 cM1
22_052E	10	7.1	cT4 cN1 cM1

22_200D	7a	11	pT3b pN1 (2/21) cM1b L0 V0 Pn1
22_298D	7b	18.3	cT2a cN1 cM1b
all	G7a, n = 1 G7b, n = 1 G8, n = 2 G9, n = 1 G10, n = 1	median: 24.6	locally-restricted (T1/T2), n = 2 locally-advanced (T3/T4), n = 4

7.2 Appendix B: Figures

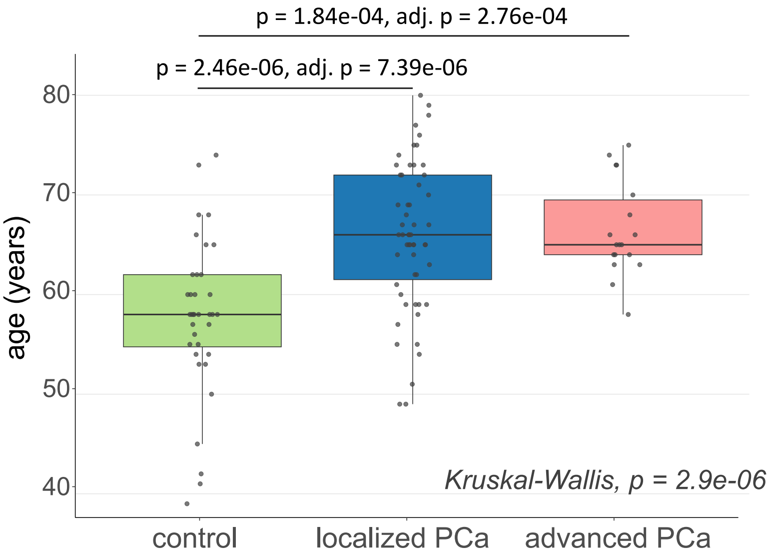


Figure A1: Age distribution in PCa patients and cancer-free controls.
Age at the initial timepoint of diagnosis for PCa patients and at the time of PCa screening or sample collection for controls. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one individual. Differences between the three cohorts were assessed using Kruskal-Wallis testing, followed by Dunn's post hoc test. Correction for multiple testing was performed with Benjamini-Hochberg method. Only significant differences are shown. Abbreviations: adj. p = adjusted p value, PCa = prostate cancer.

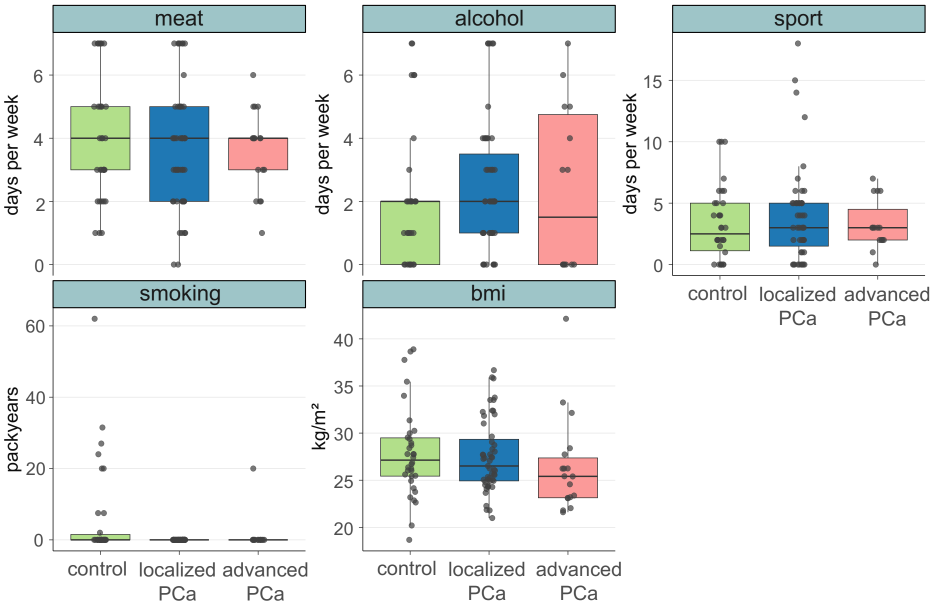


Figure A2: Lifestyle factors and potential PCa risk factors assessed by questionnaire.
Box plots showing questionnaire-based assessments of alcohol consumption (days per week), meat consumption (days per week), smoking exposure (pack-years), physical activity (sports), and body mass index (bmi, kg/m^2) for localized PCa patients, advanced PCa patients, and cancer-free controls. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one individual. Abbreviations: bmi = body mass index.

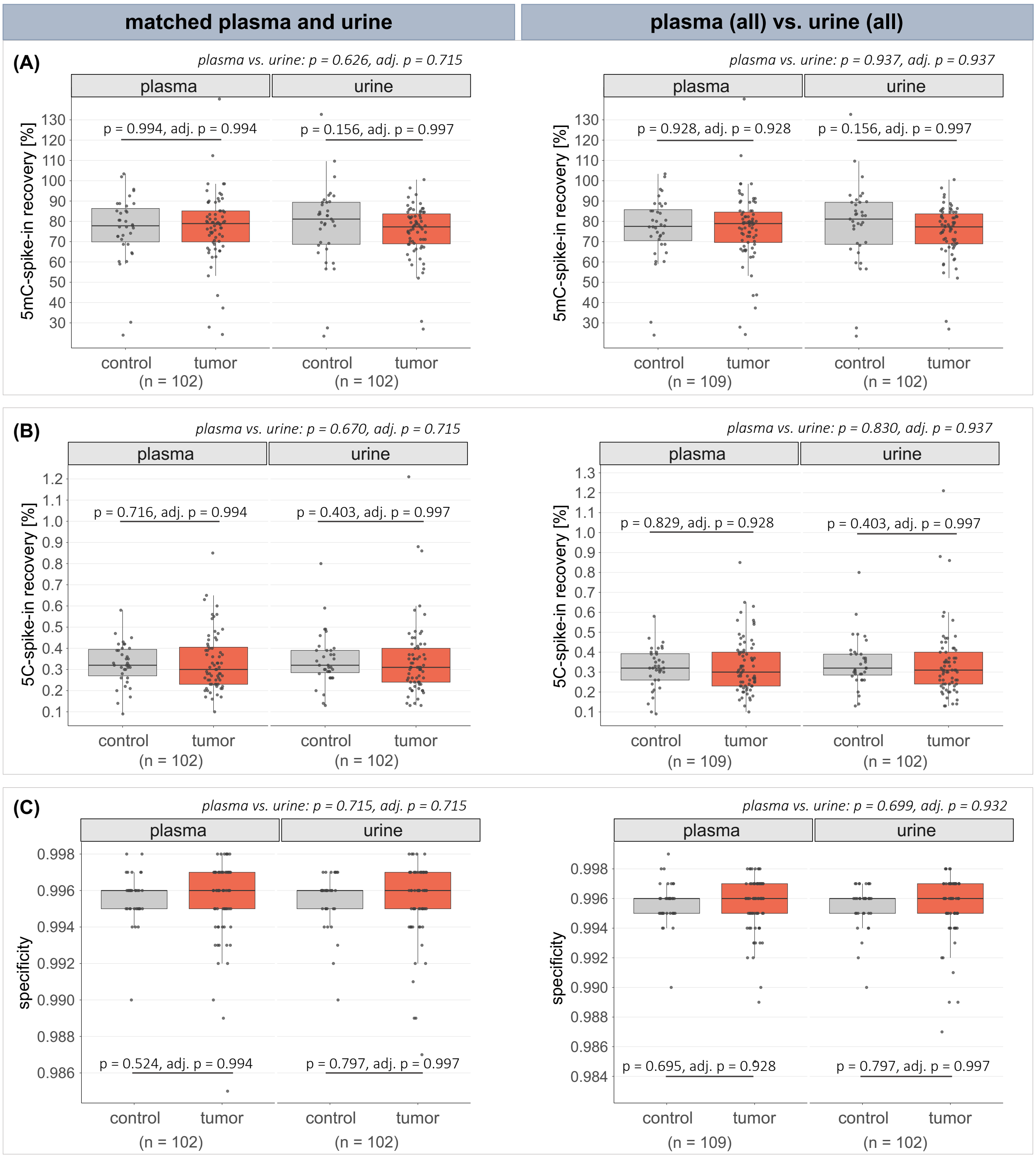


Figure A3: Comparison of laboratory-derived methylation QC metrics in plasma and urinary cfDNA samples. Box plots showing laboratory-based methylation QC metrics in cfMeDIP-seq libraries derived from plasma and urinary cfDNA from PCa patients and controls: **(A)** recovery of the methylated (5mC-) spike-in, **(B)** recovery of the unmethylated (5C-) spike-in, **(C)** specificity of the methylation enrichment. **(A–C)** Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one sample. Left panels include only matched plasma and urine samples (n = 102), whereas right panels include all available plasma (n = 109) and urine (n = 102) samples. Statistical comparisons were performed using Wilcoxon rank-sum test or Wilcoxon signed-rank test, as appropriate. Correction for multiple testing was performed with Benjamini-Hochberg method. Abbreviations: 5C = unmethylated, 5mC = methylated, cfDNA = cell-free DNA, cfMeDIP-seq = cell-free methylated DNA immunoprecipitation sequencing, n = number, QC = quality control.

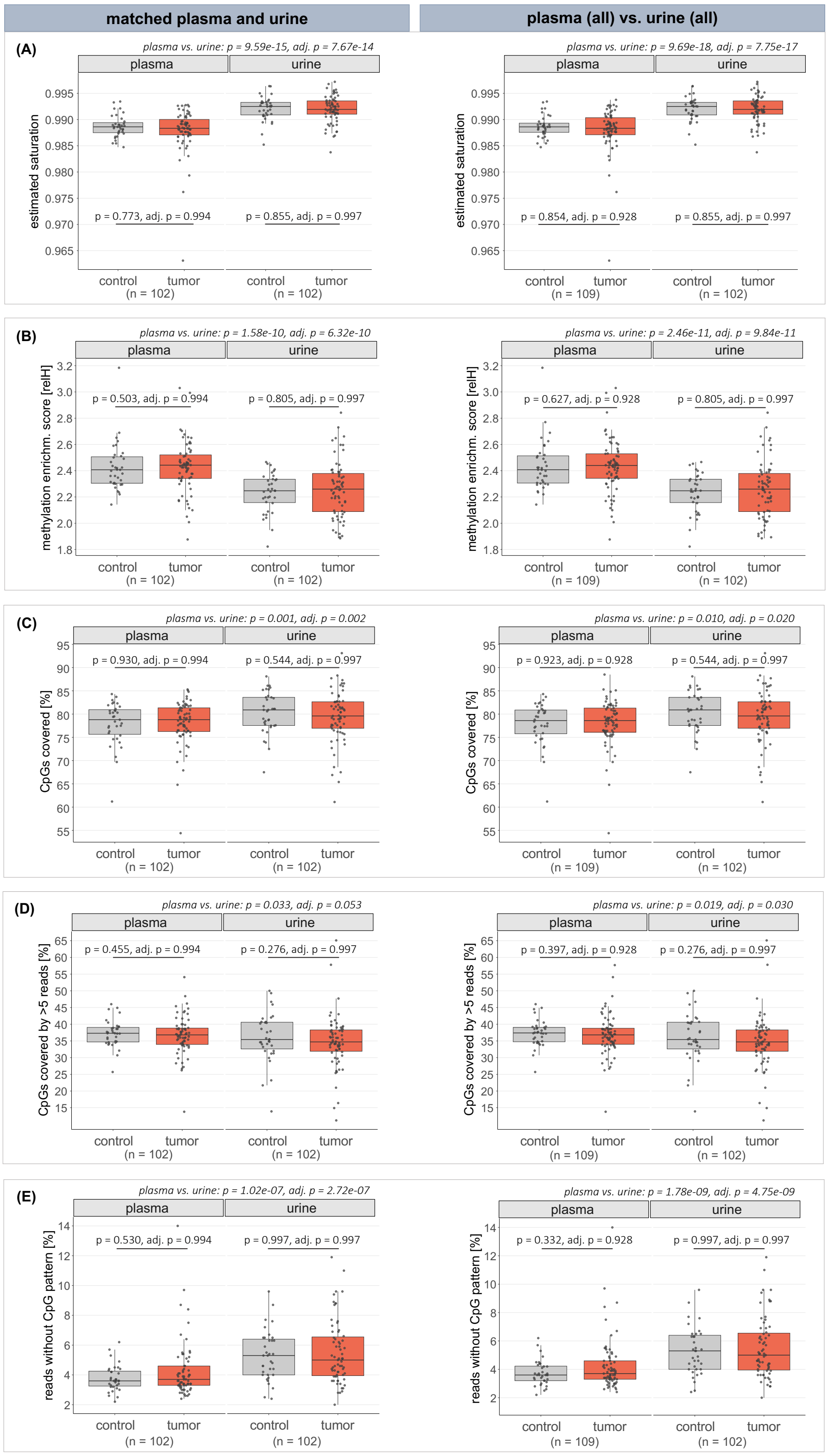


Figure A4: Comparison of bioinformatic methylation QC metrics in plasma and urinary cfDNA samples.
Box plots showing bioinformatic methylation QC metrics in cfMeDIP-seq data derived from plasma and urinary cfDNA from PCa patients and controls: **(A)** estimated saturation, derived from the Pearson correlation coefficient obtained in the saturation analysis, **(B)** relative methylation enrichment score (relH), **(C)** proportion of covered CpGs relative to all CpGs present in the human genome (CpGs covered), **(D)** proportion of CpGs covered by more than five reads, **(E)** proportion of reads lacking a CpG pattern. **(A–E)** Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Each dot represents one sample. Left panels include only matched plasma and urine samples (n = 102), whereas right panels include all available plasma (n = 109) and urine (n = 102) samples. Statistical comparisons were performed using Wilcoxon rank-sum test or Wilcoxon signed-rank test, as appropriate. Correction for multiple testing was performed with Benjamini-Hochberg method.

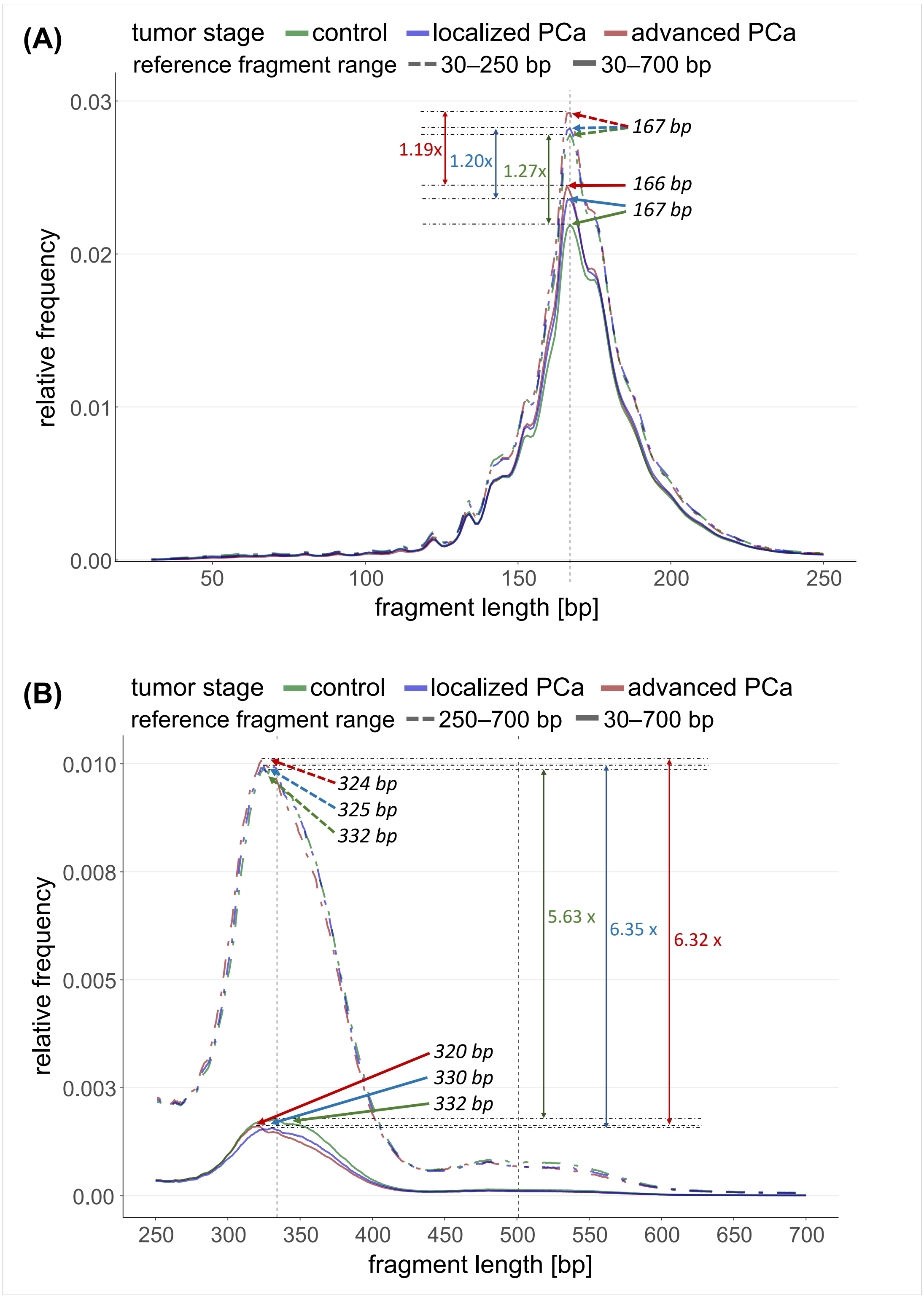


Figure A5: Comparison of plasma cfDNA fragmentation profiles calculated using different reference fragment length ranges.
Group-wise median relative frequency distributions of plasma cfDNA fragment lengths derived from lcWGS data are shown for localized PCa patients, advanced PCa patients, and controls. Fragmentation profiles with relative frequencies calculated relative to the entire fragment length range (30–700 bp; solid lines) are compared with profiles calculated relative to the subranges **(A)** 30–250 bp and **(B)** 250–700 bp (dashed lines). Vertical dotted grey lines indicate 167 bp and its multiples, 334 bp (2x 167 bp) and 501 bp (3x 167 bp). Abbreviations: bp = base pair, lcWGS = low-coverage whole-genome sequencing.

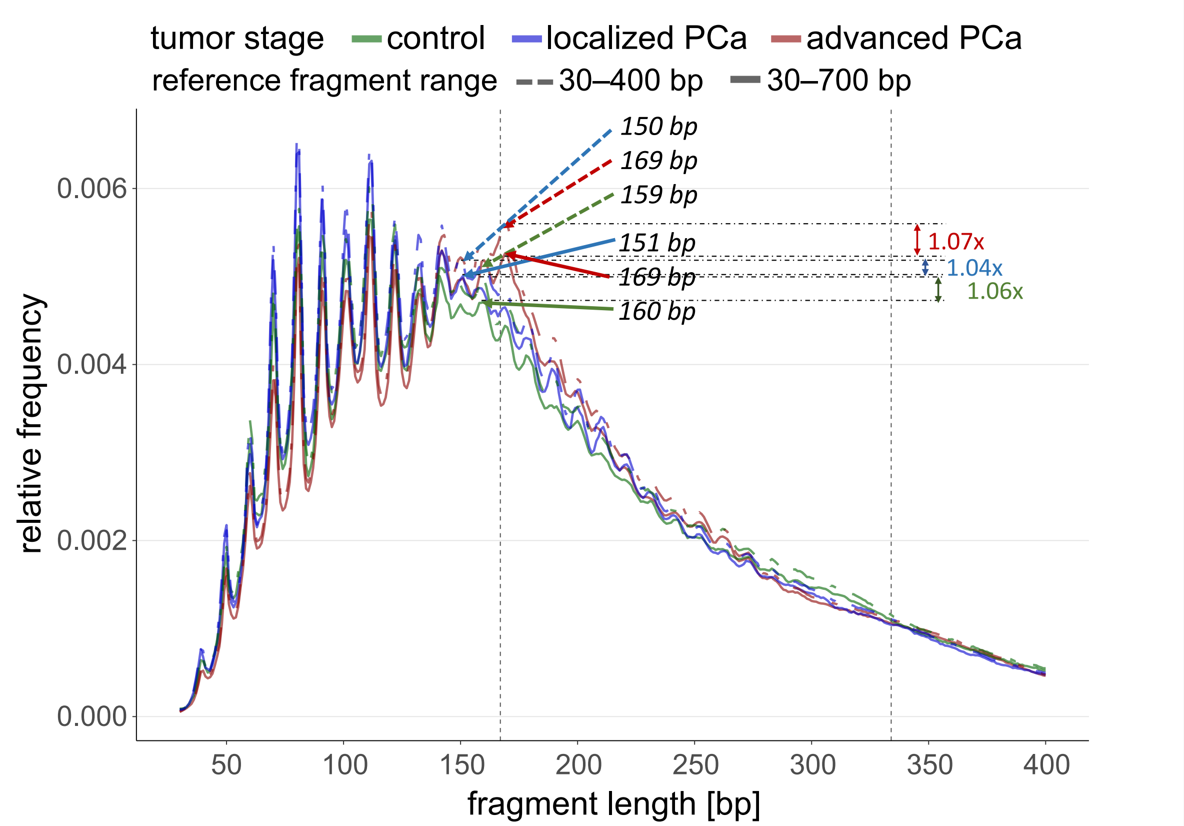


Figure A6: Comparison of urinary cfDNA fragmentation profiles calculated using different reference fragment length ranges.
Group-wise median relative frequency distributions of urinary cfDNA fragment lengths derived from lcWGS data are shown for localized PCa patients, advanced PCa patients, and controls. Fragmentation profiles with relative frequencies calculated relative to the entire fragment length range (30–700 bp; solid lines) are compared with profiles calculated relative to the subrange 30–400 bp (dashed lines). Vertical dotted grey lines indicate 167 bp and its multiple, 334 bp (2x 167 bp).

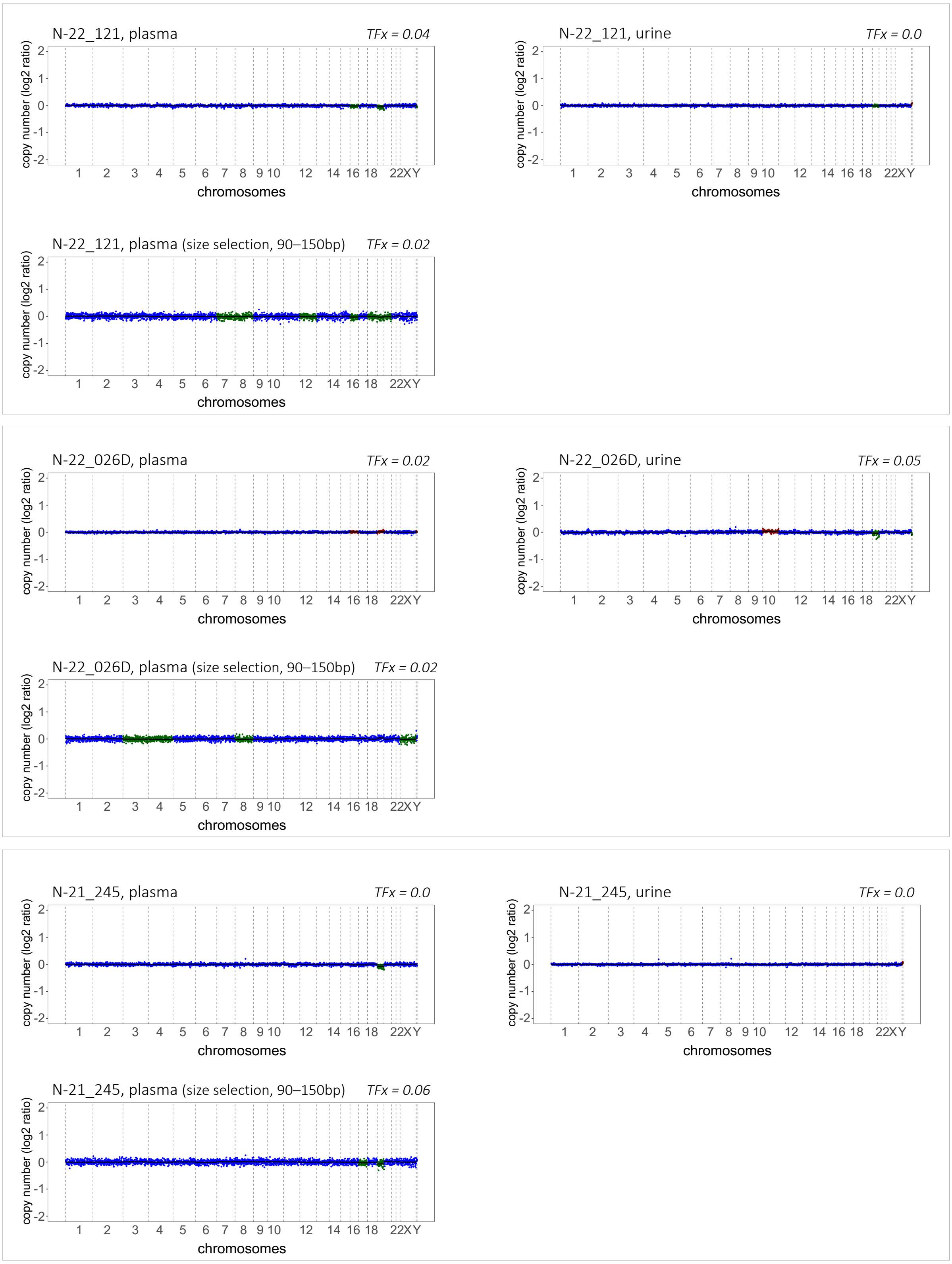


Figure A7: CNV profiles of plasma and urinary cfDNA from cancer-free controls with measurable TFx values. CNV profiles from three cancer-free controls are shown based on CNV analyses performed without size selection and with *in silico* size selection (90–150 bp) in plasma cfDNA, as well as CNV analysis without size selection in urinary cfDNA. These control samples exhibited measurable TFx values within the control cohort. CNV profiles display genome-wide chromosomal gains (red) and losses (green) relative to the copy-number-neutral state (blue) at chromosomal positions shown on the x-axis. Abbreviations: CNV = copy number variation, TFx = tumor fraction.

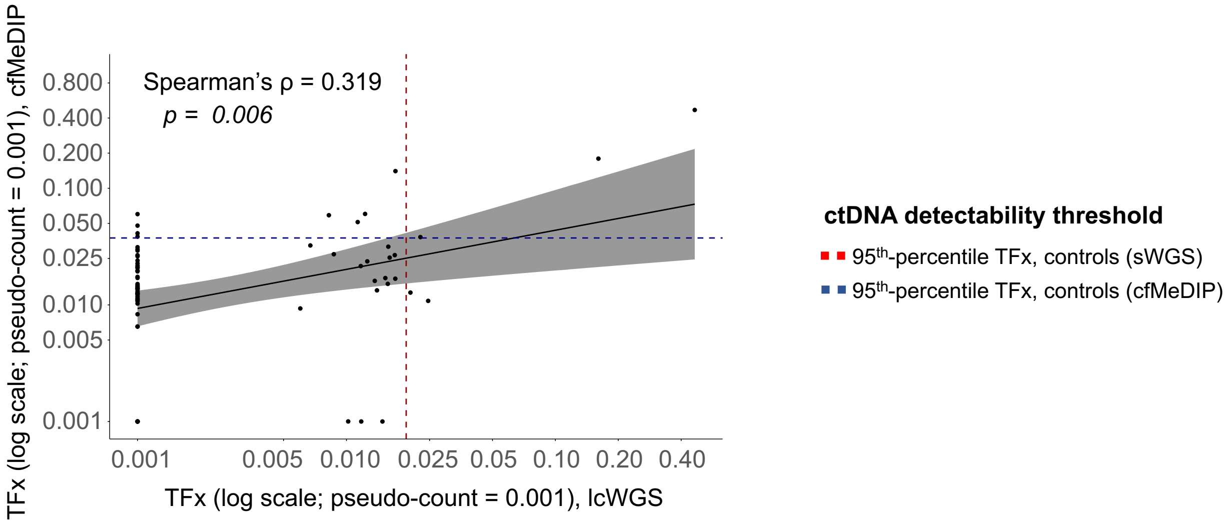
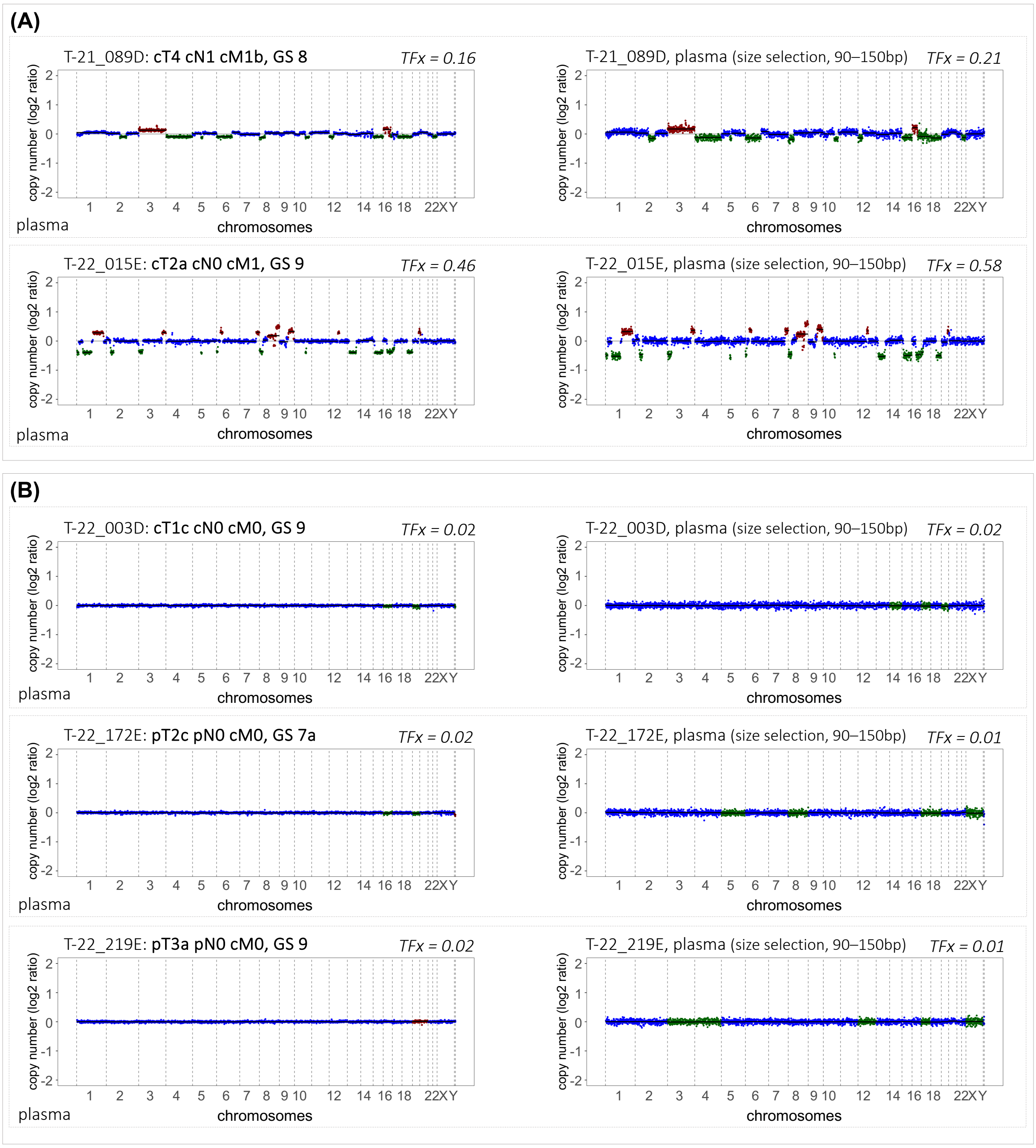


Figure A8: Correlation between TFx values derived from lcWGS and cfMeDIP-seq data in plasma cfDNA from PCa patients.
Scatter plot showing the correlation between matched TFx values derived from lcWGS data (x-axis) and cfMeDIP-seq data (y-axis) in plasma samples from PCa patients. TFx values are shown on log-transformed x- and y-axes, and a pseudo-count of 0.001 was added for visualization. Each dot represents one plasma sample. The association was assessed using Spearman rank correlation, with the correlation coefficient (Spearman's ρ) and corresponding p value reported. The solid line indicates the monotonic trend between TFx values.


(A)
T-21_089D: cT4 cN1 cM1b, GS 8
TFx = 0.16
T-21_089D, plasma (size selection, 90–150bp)
TFx = 0.21
T-22_015E: cT2a cN0 cM1, GS 9
TFx = 0.46
T-22_015E, plasma (size selection, 90–150bp)
TFx = 0.58
(B)
T-22_003D: cT1c cN0 cM0, GS 9
TFx = 0.02
T-22_003D, plasma (size selection, 90–150bp)
TFx = 0.02
T-22_172E: pT2c pN0 cM0, GS 7a
TFx = 0.02
T-22_172E, plasma (size selection, 90–150bp)
TFx = 0.01
T-22_219E: pT3a pN0 cM0, GS 9
TFx = 0.02
T-22_219E, plasma (size selection, 90–150bp)
TFx = 0.01
copy number (log2 ratio)
2
1
0
-1
-2
1 2 3 4 5 6 7 8 9 10 12 14 16 18 22XY
chromosomes
plasma

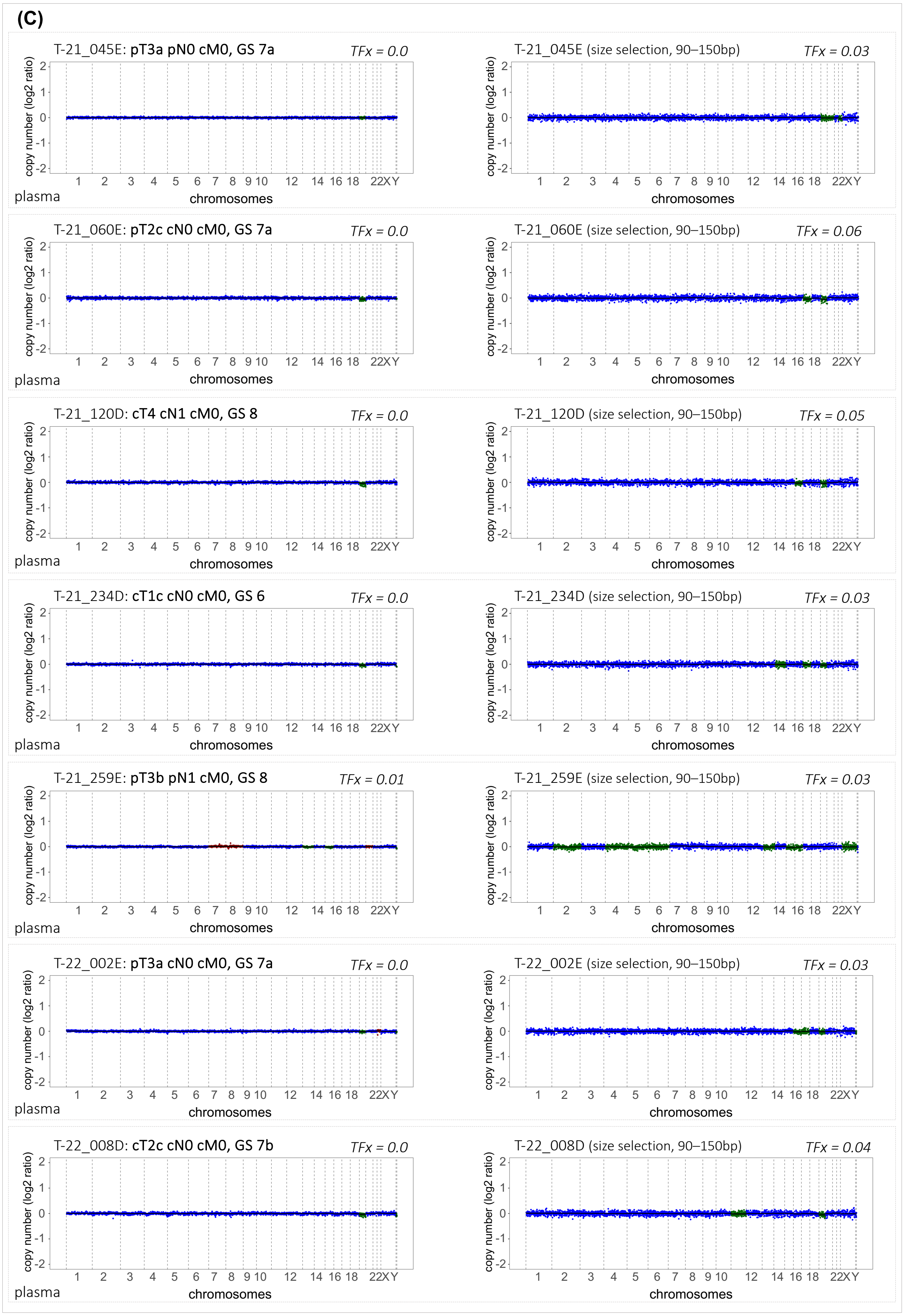


Figure A9: CNV profiles of plasma cfDNA from PCa patients derived from CNV analyses with and without *in silico* size selection.

Genome-wide CNV profiles from 12 PCa patients are shown based on CNV analyses performed on plasma cfDNA without size selection (left side) and with in silico size selection (90–150 bp; right side). CNV profiles were derived from lcWGS data and depict chromosomal gains (red) and losses (green) relative to the copy-number-neutral state (blue) across chromosomal positions shown on the x-axis. **(A)** Two PCa patients with ctDNA-positive plasma samples in both CNV analyses with and without size selection, exhibiting increased TFx values following size selection. **(B)** Three PCa patients with ctDNA-positive plasma samples in the CNV analysis without size selection, but ctDNA-negative results after size selection due to decreased TFx values. **(C)** Seven PCa patients with ctDNA-negative plasma samples in the CNV analysis without size selection, but ctDNA-positive results after size selection due to increased TFx values. Abbreviations: ctDNA = circulating tumor-derived DNA.

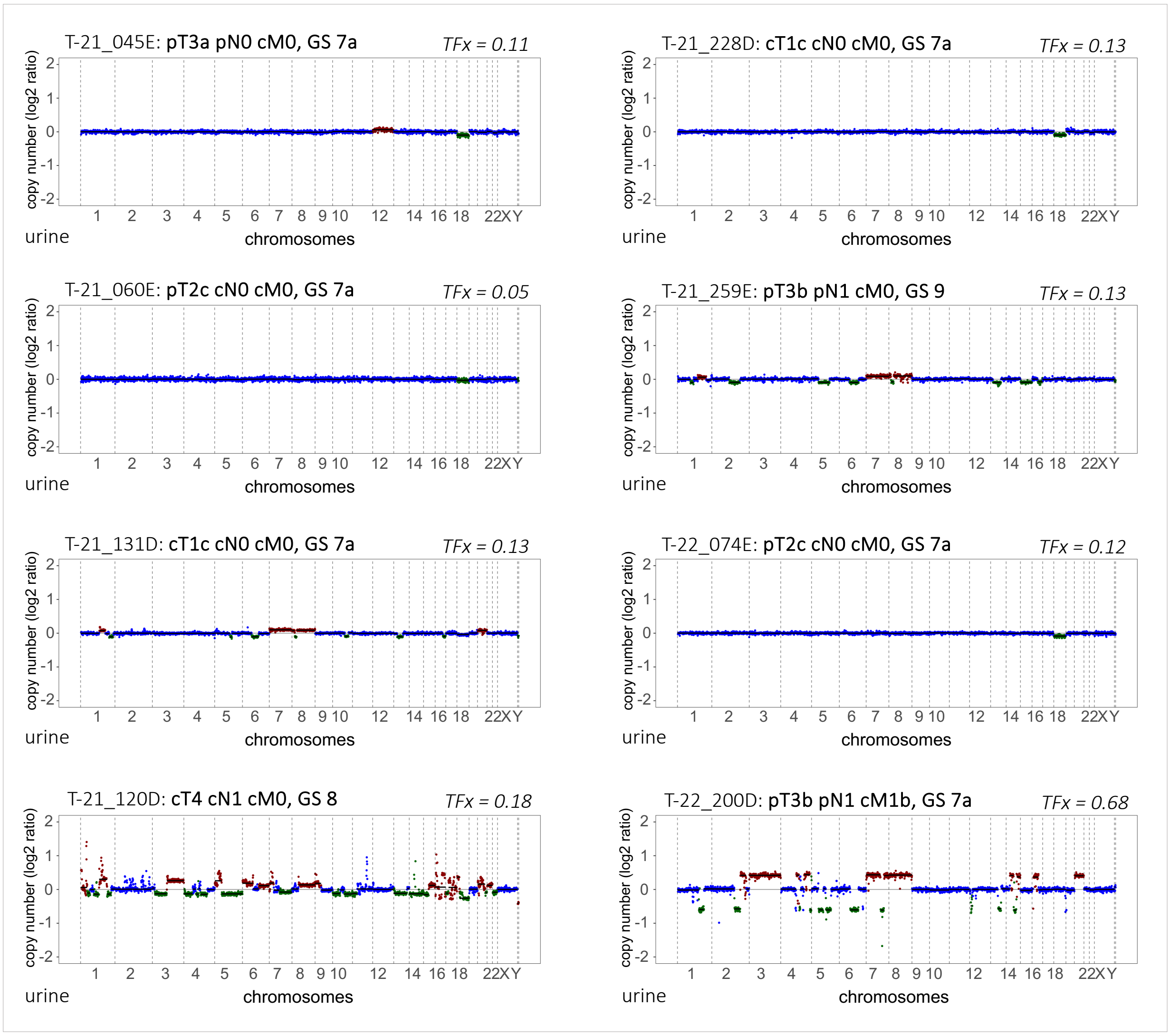


Figure A10: CNV profiles of urinary cfDNA from PCa patients derived from CNV analyses without size selection. Genome-wide CNV profiles from eight PCa patients are shown based on CNV analyses performed on urinary cfDNA without size selection. CNV profiles were derived from lcWGS data and depict chromosomal gains (red) and losses (green) relative to the copy-number-neutral state (blue) across chromosomal positions shown on the x-axis. All eight urine samples were ctDNA-positive, exhibiting TFx values above the ctDNA detection threshold defined as 95th-percentile of values derived from control samples.

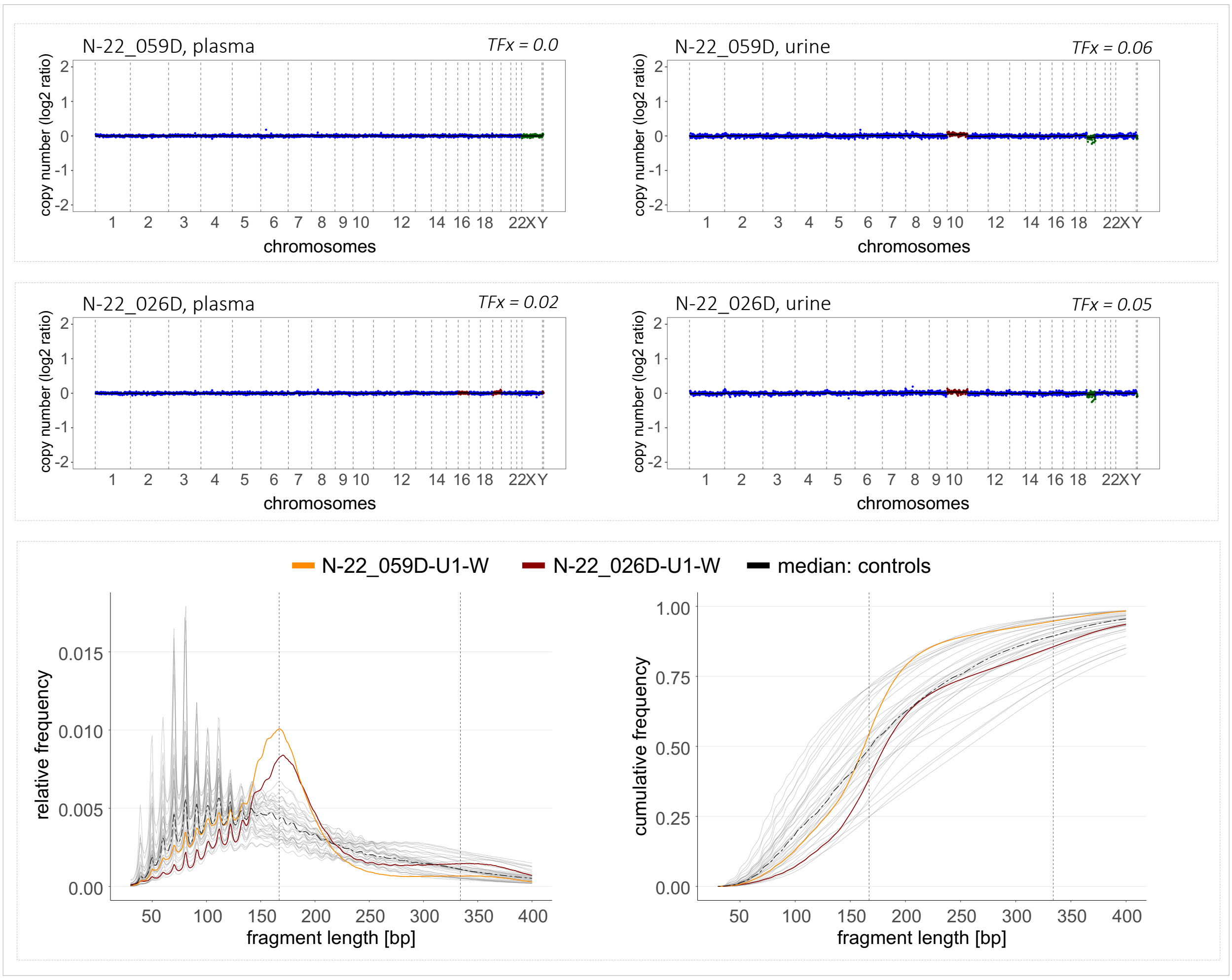


Figure A11: CNV and fragmentation profiles of urinary cfDNA from control samples with elevated TFx values. Top panels: Genome-wide CNV profiles of two control samples with the highest and second-highest TFx values among controls based on the CNV analysis without size selection in urinary cfDNA. CNV profiles derived from lcWGS data are shown for matched plasma (left) and urine (right) samples and depict chromosomal gains (red) and losses (green) relative to the copy-number-neutral state (blue) across chromosomal positions on the x-axis. Bottom panels: CfDNA fragmentation profiles derived from lcWGS data of the same two urine samples from cancer-free controls, showing relative (left) and cumulative (right) fragment length frequency distributions within the 30–400 bp range, calculated relative to the entire fragment length range (30–700 bp). The control sample with the highest TFx value among controls is shown in orange, and the sample with the second-highest TFx value in red. Dashed black lines indicate the median frequency distribution across all control samples, and grey lines represent individual control samples. Vertical dotted grey lines indicate 167 bp and its multiple, 334 bp (2x 167 bp).

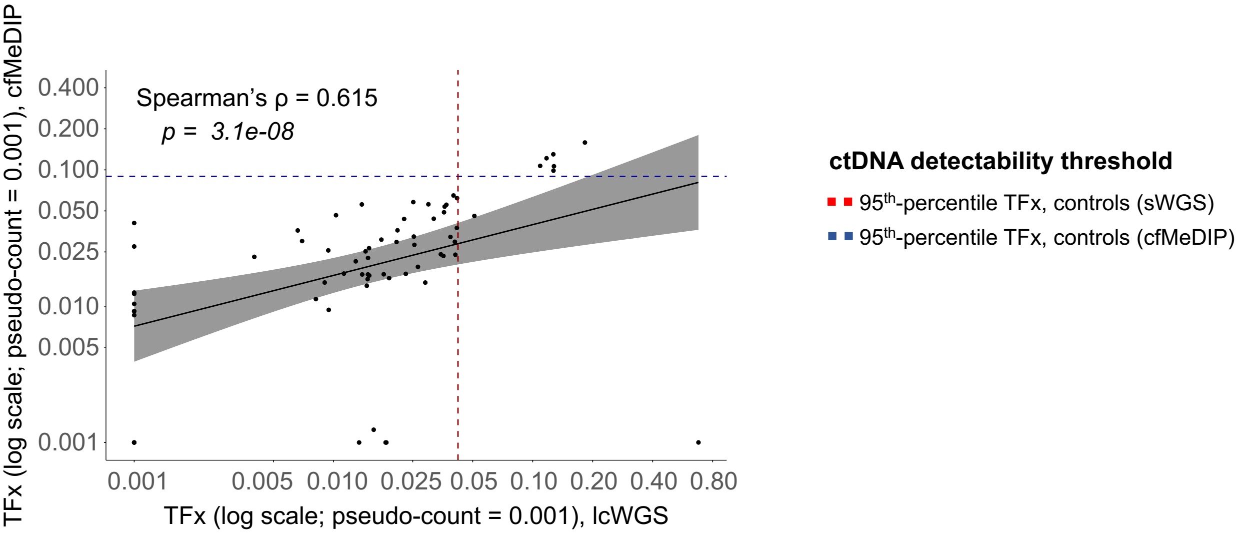


Figure A12: Correlation between TFx values derived from lcWGS and cfMeDIP-seq data in urinary cfDNA from PCa patients.
Scatter plot showing the correlation between matched TFx values derived from lcWGS data (x-axis) and cfMeDIP-seq data (y-axis) in urine samples from PCa patients. TFx values are shown on log-transformed x- and y-axes, and a pseudo-count of 0.001 was added for visualization. Each dot represents one urine sample. Associations were assessed using Spearman rank correlation, with the correlation coefficient (Spearman's ρ) and corresponding p value reported. The solid line indicates the monotonic trend between TFx values.

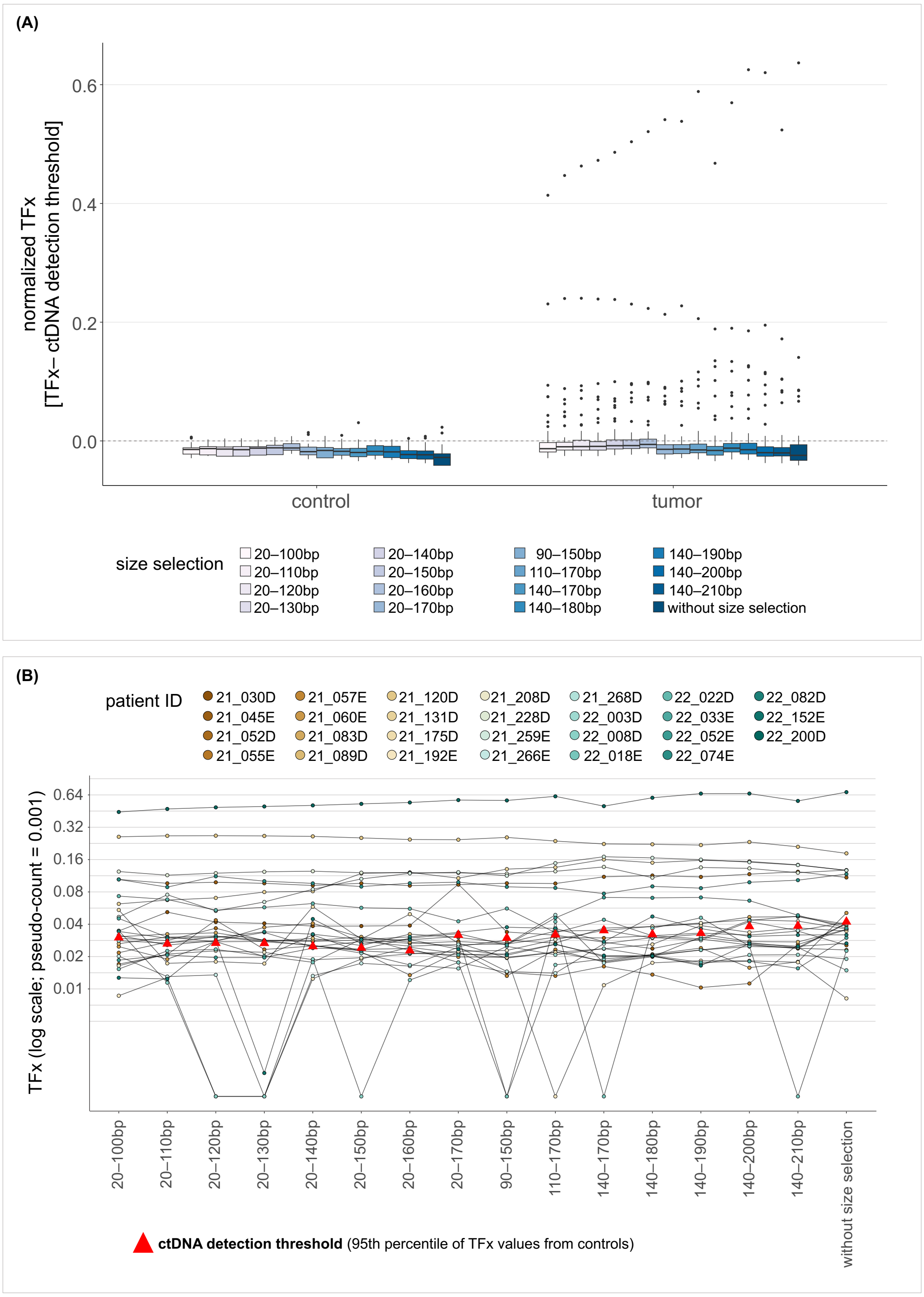
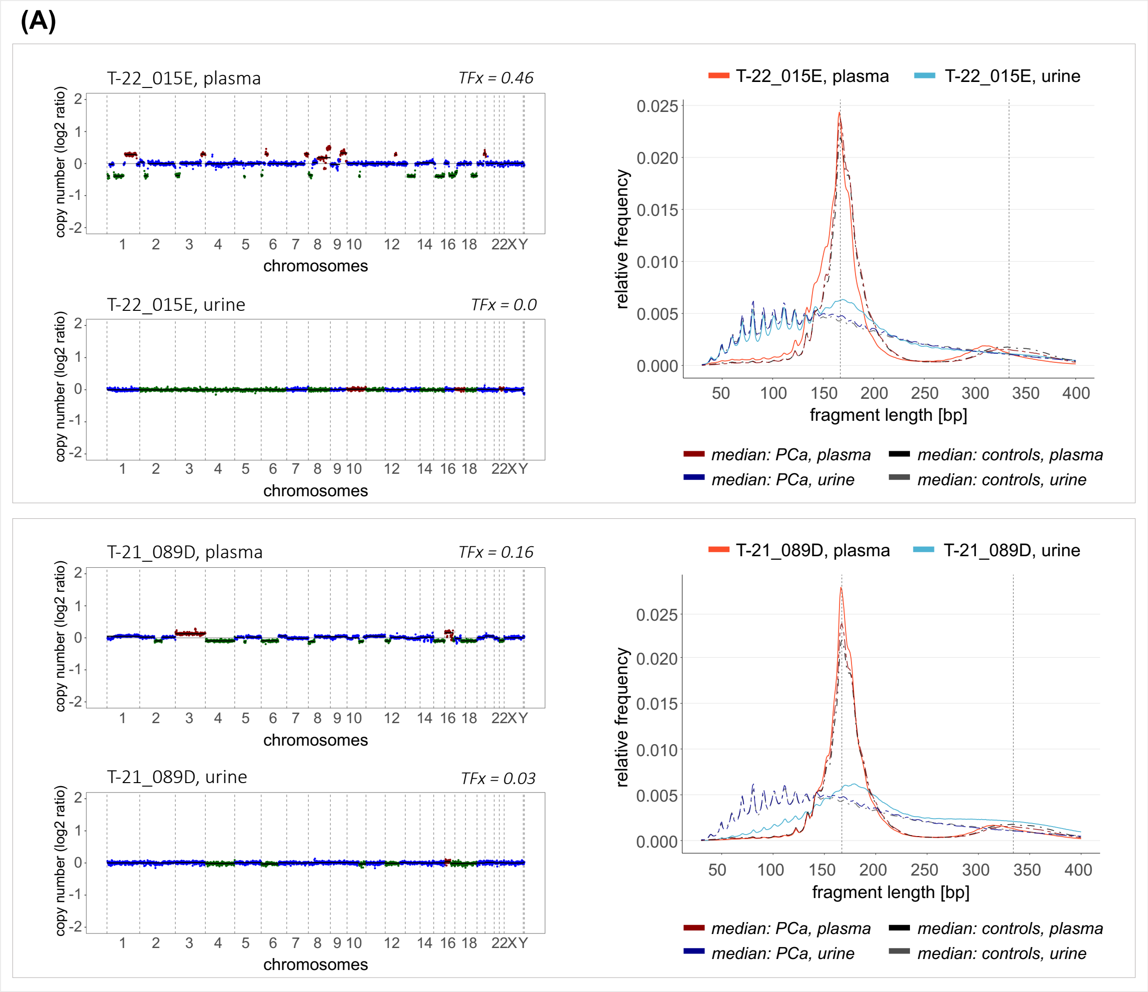
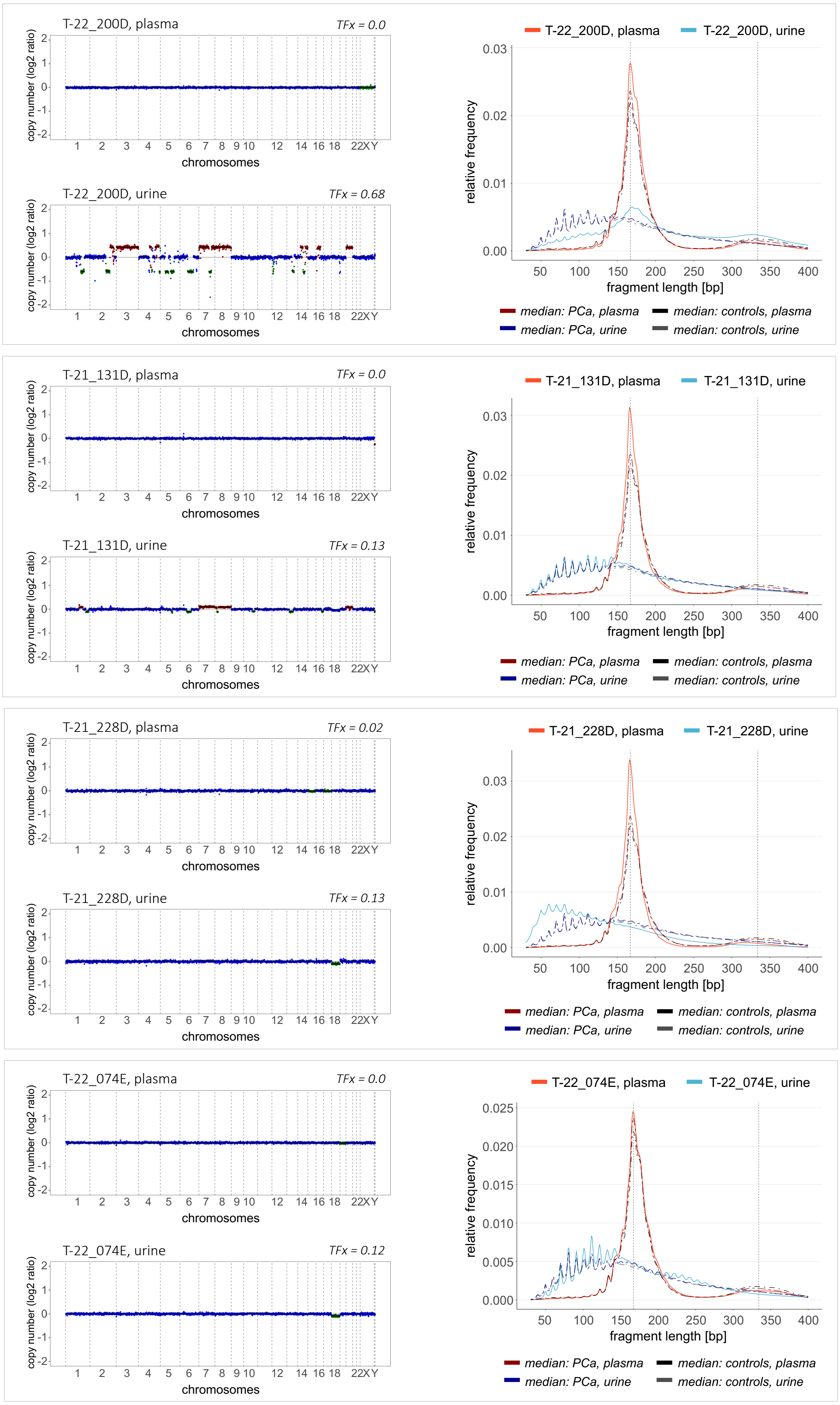


Figure A13: Changes in TFx values derived from CNV analyses with *in silico* size selection across different fragment length ranges in urinary cfDNA.
(A) Distribution of the absolute differences between estimated TFx values and the corresponding ctDNA detection threshold (95th-percentile of TFx values derived from control samples) across urine samples from PCa patients and controls. Each data point represents one urine sample, with values calculated as TFx minus the ctDNA detection threshold. TFx values were estimated from genome-wide CNV profiles derived from lcWGS data following *in silico* size selection for 15 different fragment length ranges. Box plot center lines indicate the median, and boxes illustrate the interquartile range with Tukey whiskers. Dots represent outlying samples. **(B)** Changes in estimated TFx values derived from CNV analyses following *in silico* size selection across 15 different fragment length ranges in urine samples from PCa patients with at least one ctDNA-positive result. Each dot represents one sample. For each patient, TFx values obtained from all CNV analyses without and with size selection are shown, and values from the same patient are connected by black lines. TFx values are displayed on a log-transformed y-axis, and a pseudo-count of 0.001 was added for visualization.


(A)
T-22_015E, plasma
TFx = 0.46
copy number (log2 ratio)
2
1
0
-1
-2
1 2 3 4 5 6 7 8 9 10 12 14 16 18 22XY
chromosomes
T-22_015E, urine
TFx = 0.0
T-22_015E, plasma
T-22_015E, urine
relative frequency
0.025
0.020
0.015
0.010
0.005
0.000
50 100 150 200 250 300 350 400
fragment length [bp]
median: PCa, plasma
median: controls, plasma
median: PCa, urine
median: controls, urine
T-21_089D, plasma
TFx = 0.16
T-21_089D, urine
TFx = 0.03
T-21_089D, plasma
T-21_089D, urine

(B)


T-22_200D, plasma
TFx = 0.0
T-22_200D, urine
TFx = 0.68
copy number (log2 ratio)
chromosomes
1 2 3 4 5 6 7 8 9 10 12 14 16 18 22XY
T-22_200D, plasma
T-22_200D, urine
relative frequency
fragment length [bp]
median: PCa, plasma
median: controls, plasma
median: PCa, urine
median: controls, urine
T-21_131D, plasma
TFx = 0.0
T-21_131D, urine
TFx = 0.13
T-21_228D, plasma
TFx = 0.02
T-21_228D, urine
TFx = 0.13
T-22_074E, plasma
TFx = 0.0
T-22_074E, urine
TFx = 0.12

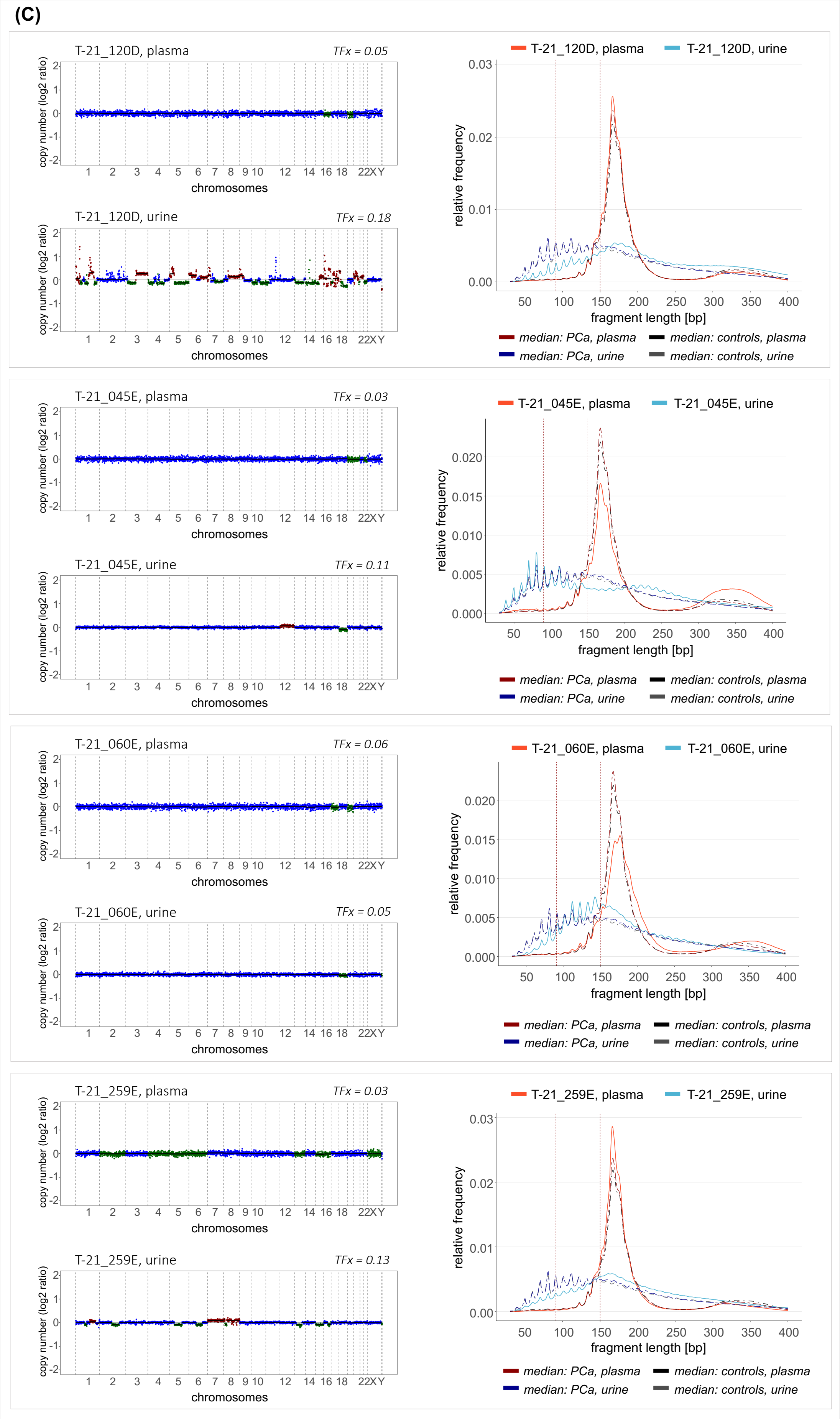

(C)
T-21_120D, plasma
TFx = 0.05
T-21_120D, urine
TFx = 0.18
T-21_120D, plasma
T-21_120D, urine
T-21_045E, plasma
TFx = 0.03
T-21_045E, urine
TFx = 0.11
T-21_045E, plasma
T-21_045E, urine
T-21_060E, plasma
TFx = 0.06
T-21_060E, urine
TFx = 0.05
T-21_060E, plasma
T-21_060E, urine
T-21_259E, plasma
TFx = 0.03
T-21_259E, urine
TFx = 0.13
T-21_259E, plasma
T-21_259E, urine
copy number (log2 ratio)
2
1
0
-1
-2
1 2 3 4 5 6 7 8 9 10 12 14 16 18 22XY
chromosomes
relative frequency
0.03
0.02
0.01
0.00
0.020
0.015
0.010
0.005
0.000
50 100 150 200 250 300 350 400
fragment length [bp]
median: PCa, plasma
median: controls, plasma
median: PCa, urine
median: controls, urine

Figure A14: CNV and fragmentation profiles of matched plasma and urinary cfDNA in PCa patients with concordant and discordant ctDNA detection.
Genome-wide CNV profiles and cfDNA fragmentation profiles of matched plasma and urine samples from **(A)** two PCa patients with ctDNA-positive plasma samples based on the CNV analysis with *in silico* size selection (90–150 bp), but ctDNA-negative urine samples based on the CNV analysis without size selection, **(B)** four PCa patients with ctDNA-negative plasma samples based on the CNV analysis with *in silico* size selection (90–150 bp), but ctDNA-positive urine samples based on the CNV analysis without size selection, and **(C)** four PCa patients with concordant ctDNA-positive plasma samples based on the CNV analysis with *in silico* size selection (90–150 bp) and ctDNA-positive urine samples based on the CNV analysis without size selection. **(A–C)** For each patient, CNV profiles of the matched plasma and urine samples are shown as two separate plots on the left. CNV profiles were derived from lcWGS data and depict chromosomal gains (red) and losses (green) relative to the copy-number-neutral state (blue) across chromosomal positions on the x-axis. Plasma and urinary cfDNA fragmentation profiles are displayed on the right as relative frequency distributions of fragment lengths in a combined plot. Light red and light blue lines indicate the plasma and urine sample of the respective patient, respectively. Dark red and dark blue lines represent the median fragmentation profiles of all plasma and urine samples from PCa patients, respectively. Black and gray lines denote the median fragmentation profiles of all plasma and urine samples from controls, respectively. Adapted from Riediger *et al.* [410]

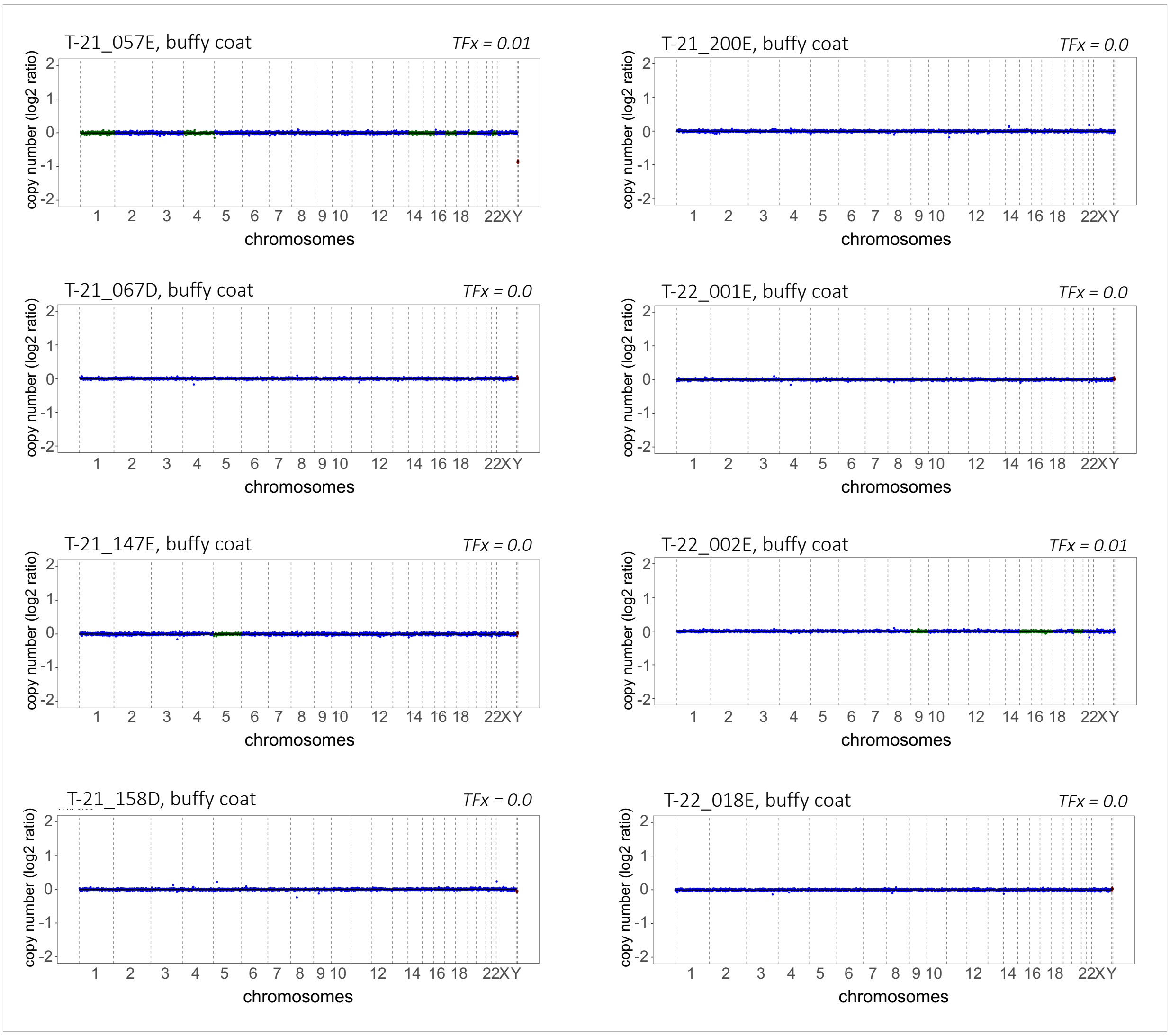


Figure A15: Genome-wide CNV profiling of gDNA from buffy coat samples.
Individual CNV profiles of eight buffy coat samples derived from lcWGS data. CNV profiles display chromosomal gains (red) and losses (green) relative to the copy-number-neutral state (blue) across chromosomal positions shown on the x-axis. No obvious genomic alterations were observed in buffy coat samples.

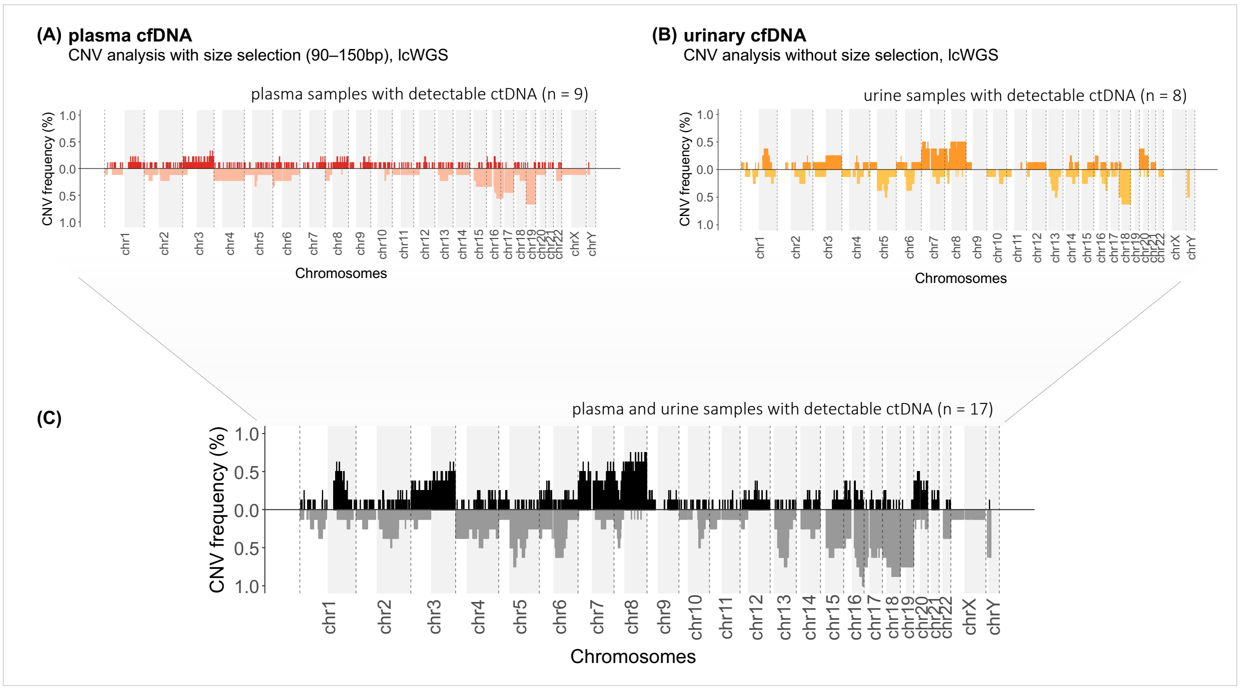


Figure A16: Recurrent CNVs in plasma and urinary cfDNA from PCa patients based on all liquid biopsy samples with detectable ctDNA.
Summary of recurrent amplifications and deletions across all plasma (**A**, n = 9) and urine (**B**, n = 8) samples with detectable CNVs and TFx values above the ctDNA detection threshold. CNV analyses were performed on the lcWGS data applying *in silico* size selection for 90–150 bp fragments in plasma cfDNA and no size selection in urinary cfDNA. **(C)** Recurrent CNVs across plasma and urine samples combined. **(A–C)** CNV profiles display chromosomal gains (dark color) and losses (light color) relative to the copy-number-neutral state. The y-axis indicates the proportion of PCa tissue samples with a detected copy number state (frequency) at each chromosomal position shown on the x-axis. Panels **(A)** and **(B)** show frequencies across nine plasma and eight urine samples, respectively, whereas panel **(C)** depicts frequencies across all 17 liquid biopsy samples combined. Areas shaded in gray represent the q-arm of the respective chromosome.

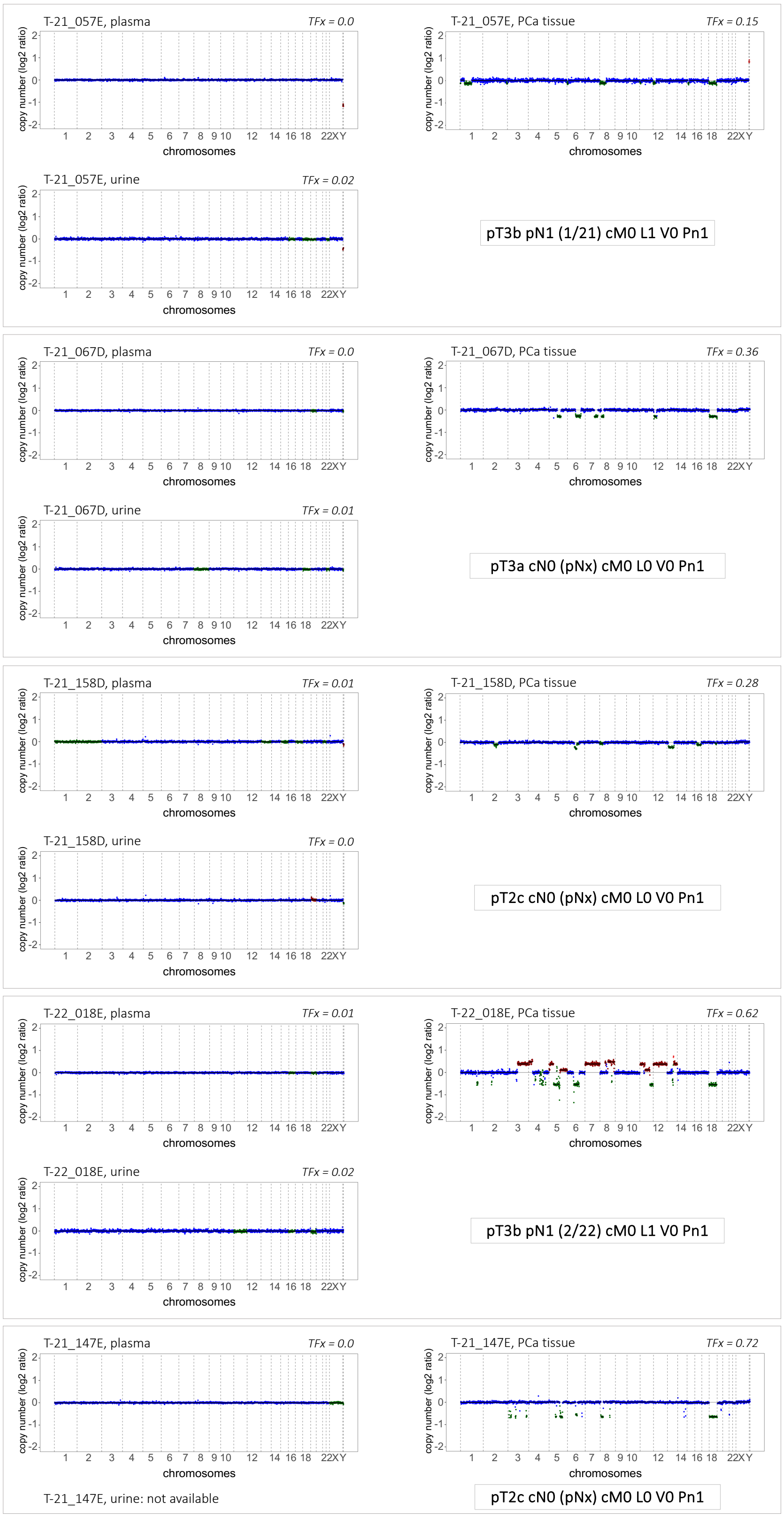
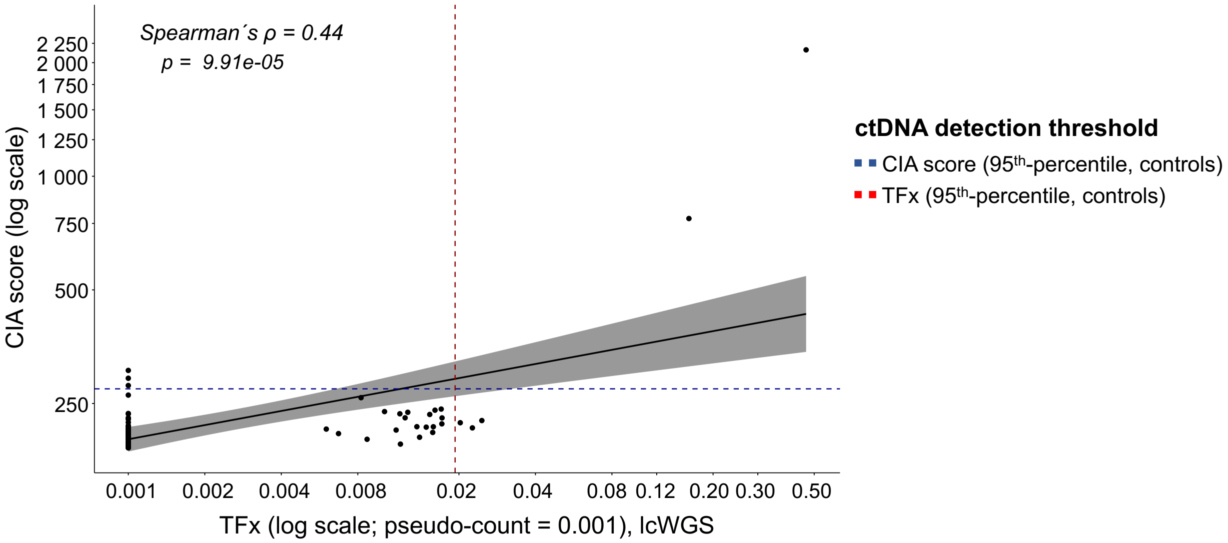


Figure A17: Genome-wide CNV profiling of cfDNA and gDNA from matched plasma, urine, and PCa tissue samples.
Individual CNV profiles of cfDNA from matched plasma and urine samples are shown for five PCa patients with detectable genomic alterations identified by CNV profiling of gDNA from PCa tissue. The corresponding liquid biopsy samples showed no obvious genomic alterations and TFx values were below the ctDNA detection threshold. CNV profiles were derived from lcWGS data and depict chromosomal gains (red) and losses (green) relative to the copy-number-neutral state (blue) across chromosomal positions shown on the x-axis. One PCa patient had no urine sample available.

(A) plasma cfDNA



(B) urinary cfDNA

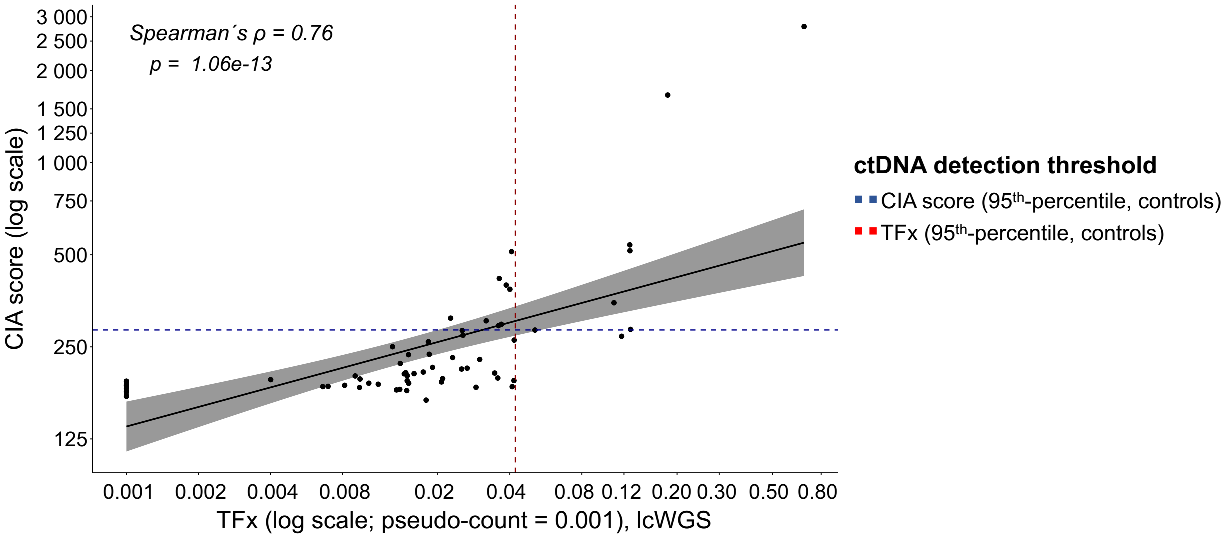


Figure A18: Correlation between matched TFx values and CIA scores in plasma and urinary cfDNA from PCa patients.
Scatter plots showing the correlation between matched TFx values derived from the CNV analysis (x-axis) and CIA scores (y-axis) in **(A)** plasma and **(B)** urine samples from PCa patients. TFx values were derived from the CNV analysis of lcWGS data following *in silico* size selection (90–150 bp) in plasma samples, and the CNV analysis without size selection in urine samples. CIA scores were derived from the analysis of chromosomal instability based on lcWGS data in both plasma and urine samples. TFx values are displayed on a log-transformed x-axis, and a pseudo-count of 0.001 was added for visualization. CIA scores are shown on a log-transformed y-axis. Each dot represents one sample. Associations were assessed using Spearman rank correlation, with the correlation coefficients (Spearman's ρ) and corresponding p values reported. The solid lines indicate the monotonic trend between TFx values and CIA scores. Abbreviations: CIA = chromosomal instability analysis.

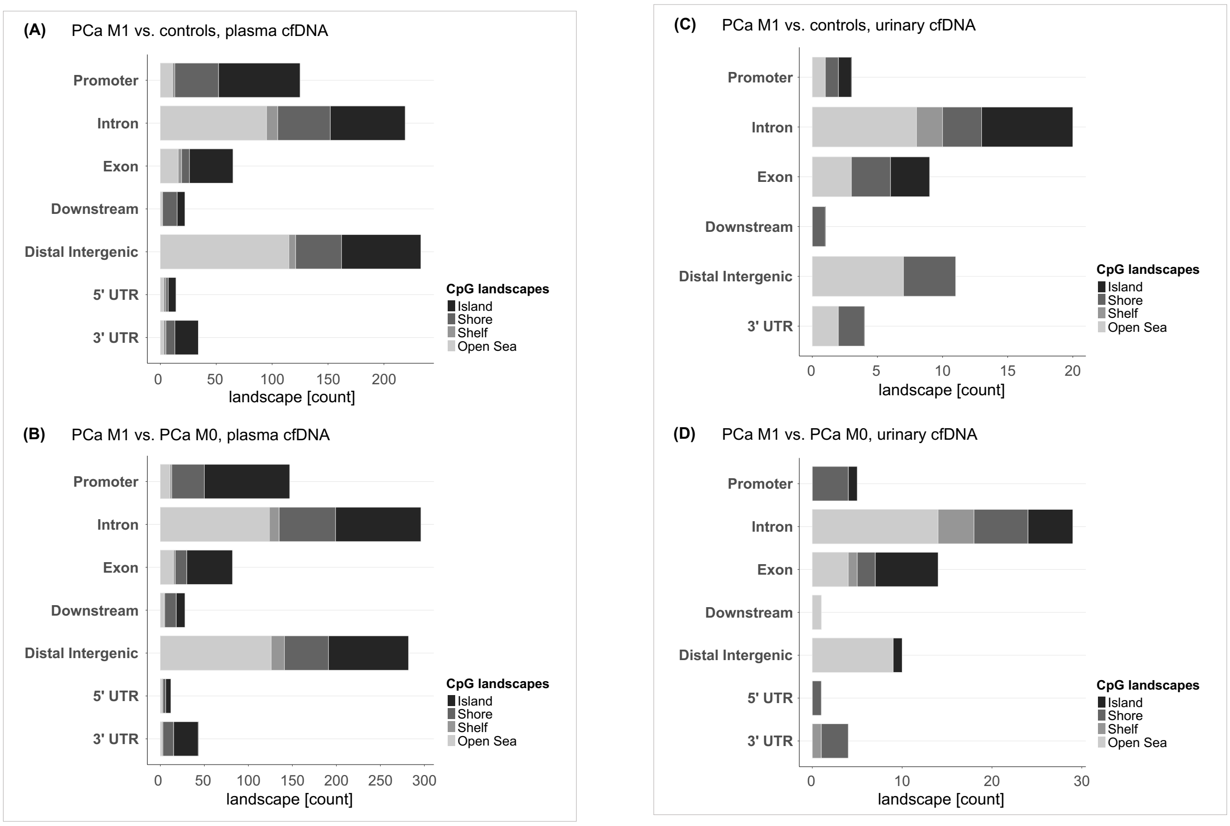


Figure A19: Genomic annotation of DMRs identified between metastatic PCa patients and controls or PCa patients without distant metastases in plasma and urinary cfDNA.

Bar plots showing the genomic annotation and CpG context distribution of **(A)** 712 DMRs identified between metastatic PCa patients and controls in plasma cfDNA, **(B)** 890 DMRs identified between metastatic PCa patients and PCa patients without distant metastases in plasma cfDNA, **(C)** 48 DMRs identified between metastatic PCa patients and controls in urinary cfDNA, and **(D)** 64 DMRs identified between metastatic PCa patients and PCa patients without distant metastases in urinary cfDNA. **(A–D)** Distributions are shown across genomic features (3' or 5'UTR, distal intergenic region, downstream region, exon, intron, or promoter region) and CpG-associated landscapes (CpG island, shore, shelf, or open sea). Abbreviations: DMR = differentially methylated region, M0/M1 = absence/presence of distant metastases, UTR = untranslated region.

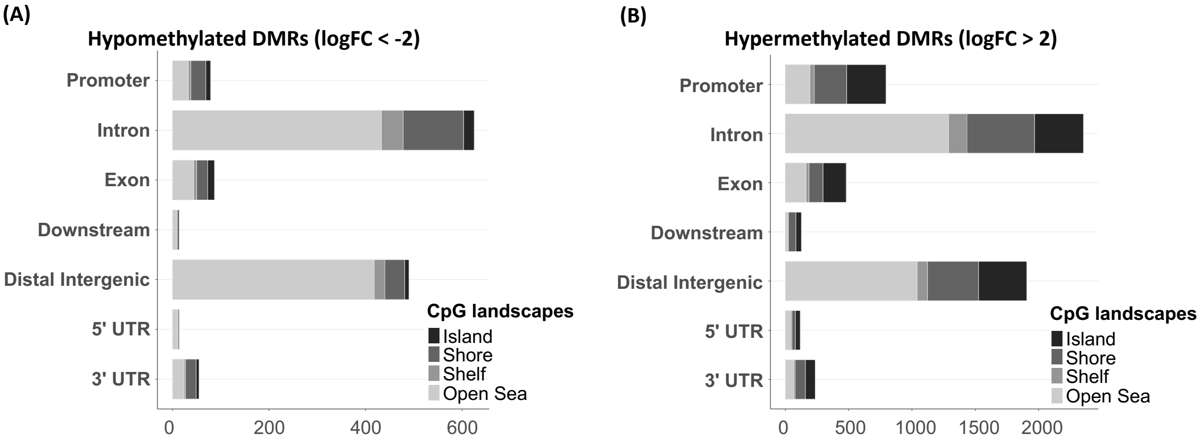


Figure A20: Genomic annotation of DMRs identified in gDNA from PCa tissue compared with matched buffy coat samples.

Bar plots showing the genomic annotation and CpG context distribution of **(A)** 1364 significant hypomethylated DMRs (logFC < -2), and **(B)** 6015 significant hypermethylated DMRs (logFC > 2) identified in gDNA from PCa tissue samples compared to matched buffy coat samples. **(A–B)** Distributions are shown across genomic features (3' or 5'UTR, distal intergenic region, downstream region, exon, intron, or promoter region) and CpG-associated landscapes (CpG island, shore, shelf, or open sea). Abbreviations: gDNA = genomic DNA, logFC = log fold change. Adapted from Riediger *et al.*[410]

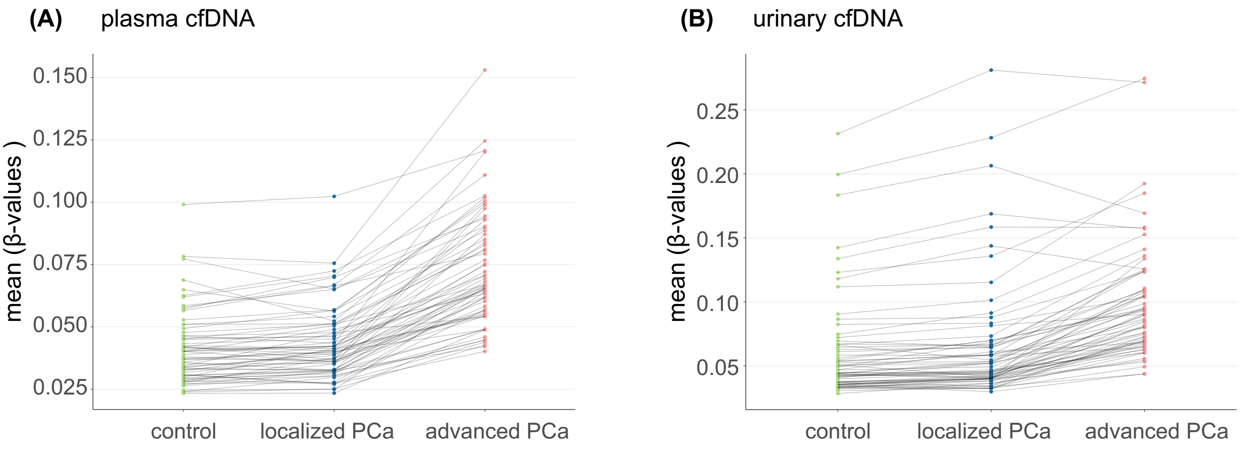


Figure A21: Distribution of mean beta-values across 67 PCa tissue-derived methylation marker regions in plasma and urinary cfDNA from PCa patients and controls.
Dot plots showing the distribution of mean β-values across 67 PCa tissue-derived methylation marker regions in **(A)** plasma and **(B)** urinary cfDNA from localized PCa patients, advanced PCa patients, and cancer-free controls. Each data point represents one region and corresponds to the mean β-value across all plasma or urine samples, respectively, within each cohort for that region. Gray lines connect the mean β-values of the three cohorts within the same region.

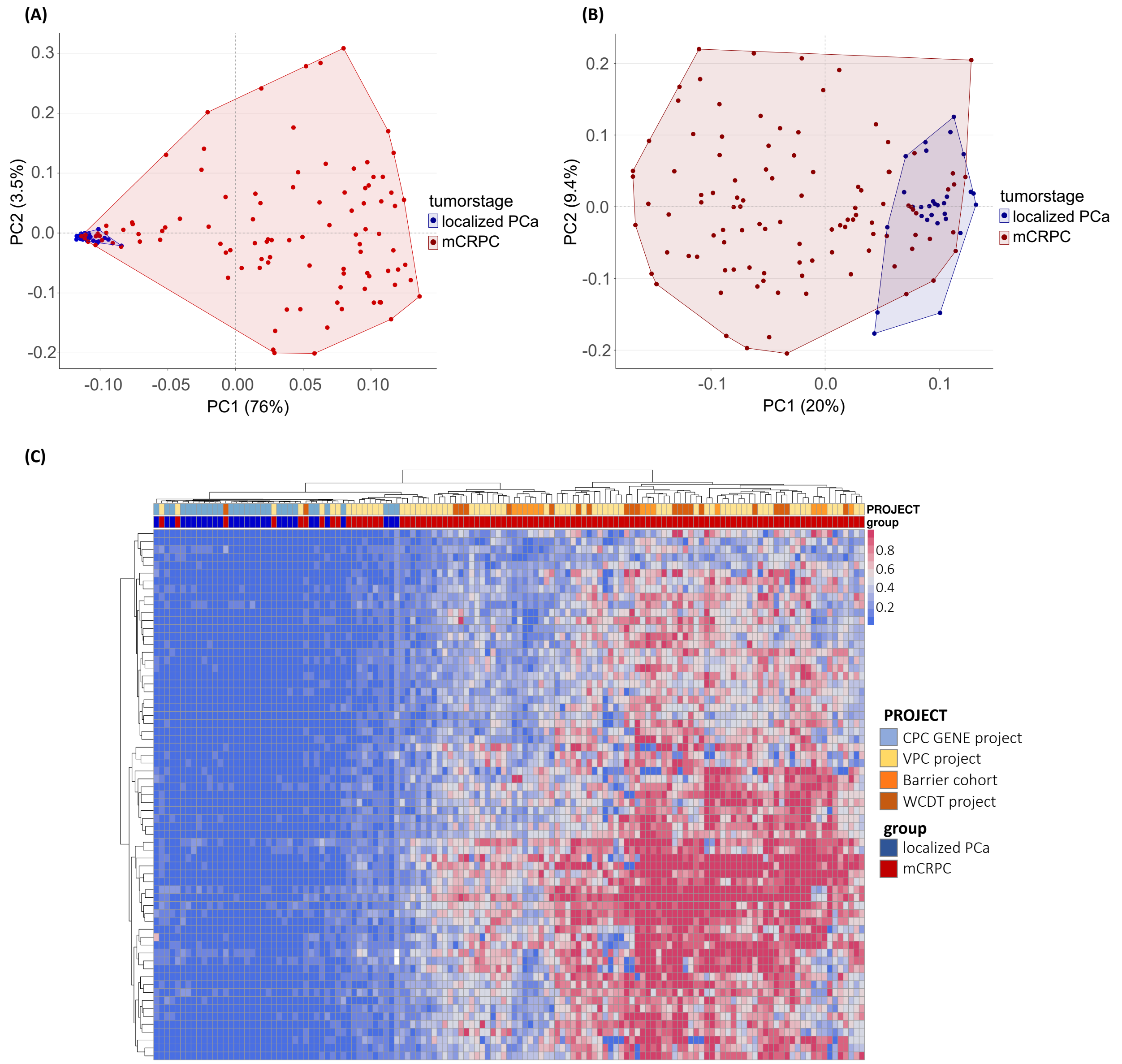


Figure A22: Unsupervised clustering based on β-values in 67 PCa tissue-derived methylation marker regions across plasma samples from an external PCa cohort.
Unsupervised clustering using PCA and hierarchical clustering based on absolute methylation levels in selected genomic regions across plasma samples from patients with localized PCa or mCRPC from an external PCa cohort (Chen et al. [354]). **(A)** PCA based on β-values in the 67 PCa tissue-derived methylation marker regions. **(B)** PCA based on β-values in 67 randomly selected, genome-wide 300-bp regions with similar CpG density (CpG density > 6) to the 67 PCa tissue-derived methylation marker regions. **(A–B)** The x- and y-axes of the PCA plots represent the first and second principal component, explaining the largest and second largest proportions (%) of the variance, respectively. **(C)** Heatmap showing hierarchical clustering based on β-values in the 67 PCa tissue-derived methylation marker regions. Additional annotations indicate the tumor status and the project/cohort name from which each sample of the external PCa cohort was derived. The project names were obtained from the original publication[354]. Abbreviations: mCRPC = metastatic castration-resistant prostate cancer, PC1/2 = principal component 1/2, PCA = principal component analysis. Adapted from Riediger *et al.* [410]

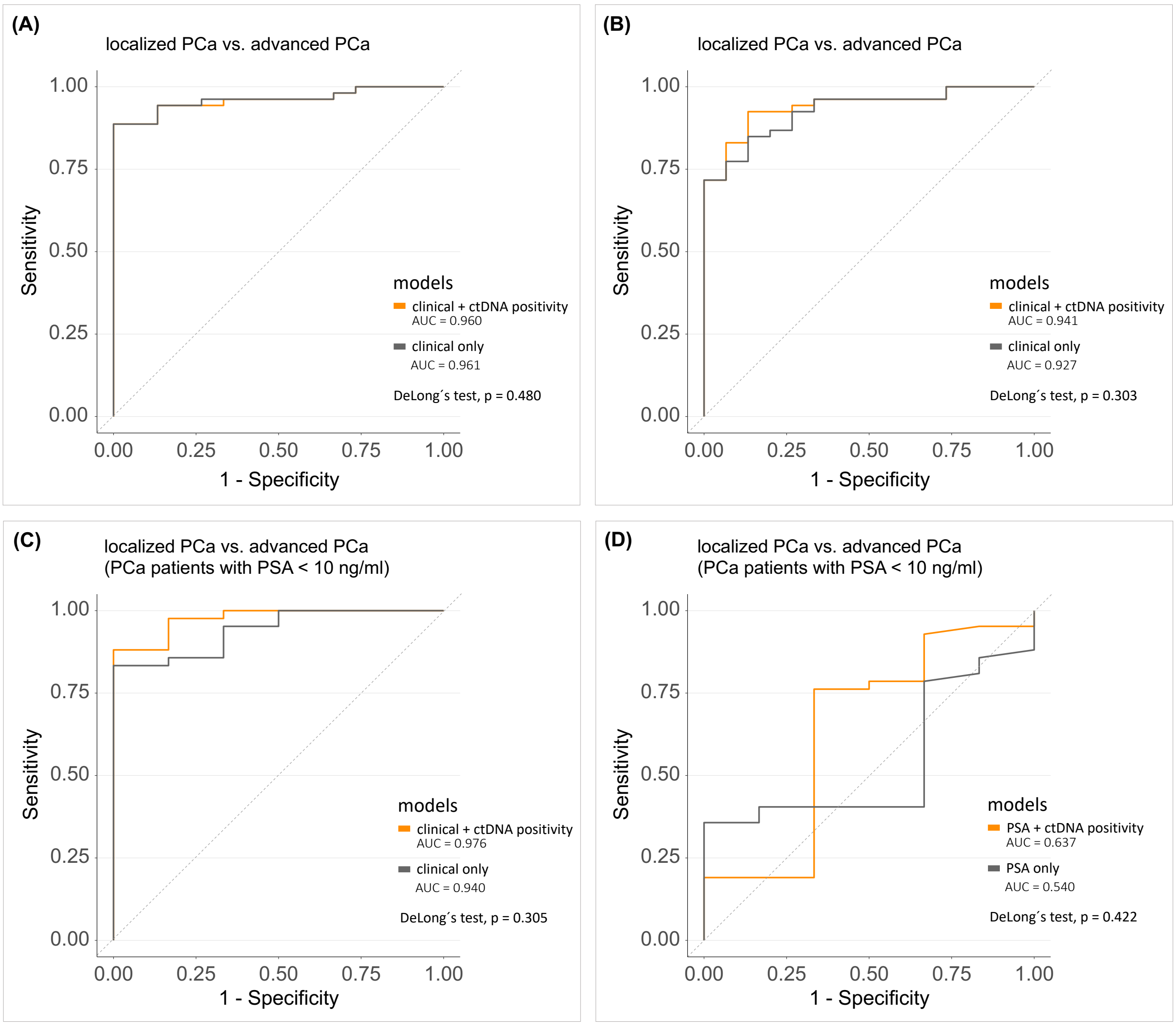


Figure A23: ROC analysis of GLMs combining clinical parameters with ctDNA positivity.
ROC curves for GLMs based on clinical parameters alone and clinical parameters combined with binary ctDNA positivity status (yes/no). A patient was considered ctDNA-positive, if either plasma or urinary cfDNA showed a positive result in any of the genomic or epigenomic analyses. ctDNA positivity was included as a binary covariate in the models. **(A–B)** Models distinguishing localized PCa from advanced PCa among all PCa patients. **(C–D)** Models distinguishing localized PCa from advanced PCa among PCa patients with PSA levels < 10 ng/ml. **(A)** Models including clinical parameters (PSA level, clinical T stage, and Gleason score) alone versus clinical parameters combined with ctDNA positivity. **(B–C)** Models including clinical parameters (PSA level and clinical T stage) alone versus clinical parameters combined with ctDNA positivity. **(D)** Models including clinical parameters (PSA level) alone versus clinical parameters combined with ctDNA positivity. **(A–D)** The diagonal reference line indicates random classification performance. Abbreviations: AUC = area under the curve, GLM = generalized linear model, PSA = prostate-specific antigen, ROC = receiver operating characteristic.

8 Acknowledgment

This work would not have been possible without the many people who supported me throughout my PhD, and I am sincerely grateful to all of them.

First and foremost, I would like to thank my supervisors PD Dr. Magdalena Görtz and Prof. Dr. Holger Sültmann for their continuous support throughout my PhD. Dr. Magdalena Görtz welcomed me as her first group member and I am deeply grateful for her trust, the freedom to develop and pursue my own scientific ideas, and her consistent support. Prof. Dr. Holger Sültmann has supervised me for almost eleven years and I am deeply thankful for his long-standing support, which has been essential for my scientific development.

My sincere thanks go to the research group of Dr. Magdalena Görtz for their scientific support, as well as for the social interactions within the group. In particular, I thank Daniela Janscho for her great work as study nurse, which was essential for the success of this project. I also thank Dr. Samaneh Eickelschulte for close scientific collaboration and shared data generation, Dr. Olga Lazareva for bioinformatic and statistical support, and Isabella Schindler for organizational assistance.

I sincerely thank the research group of Prof. Dr. Holger Sültmann for the enriching scientific and social exchange. My deepest thanks go to Dr. Florian Janke for the close scientific collaboration and daily exchange. I also thank Simon Ogrodnik and Sabrina Gerhardt for excellent experimental support, and Dr. Kate Glennon, Dr. Isabell Berneburg, Dr. Astrid Laut, and Dr. Arlou Angeles for engaging in scientific discussions during regular meetings. I sincerely thank Dr. Ann-Kathrin Daum, who has become both a colleague and a close friend, supporting me scientifically and personally in many ways.

I thank Prof. Dr. Oliver Stegle, Dr. Dr. Daniel Hübschmann, and Prof. Dr. Stefan Duensing for their valuable feedback and constructive discussions as members of the thesis advisory committee. I am also grateful to Sabine Blum for her excellent organizational support in coordinating the scientific meetings. Furthermore, I thank Prof. Dr. Jan Lohmann and Dr. Tim Waterboer for accepting my request to serve as examiners for my PhD defense.

I am grateful to the Medical Faculty Heidelberg for including me in their MD-PhD program, which enabled me to pursue this PhD. I also thank the Heidelberg Biosciences International Graduate School for providing scientific training, career development opportunities, and administrative support.

I also thank the German Cancer Research Center (DKFZ) core facilities for their support, in particular the NGS Core Facility for sequencing analyses and the DKFZ Omics IT and Data Management Core Facility for data management and processing. Furthermore, I thank the Tissue Bank of the National Center for Tumor Diseases (NCT) Heidelberg for providing the tissue samples.

I sincerely acknowledge the patients who participated in this study, as well as the medical staff at the Department of Urology at Heidelberg University Hospital who were involved in their care.

Finally, I would like to thank my family and friends for their consistent support of my personal and professional development and for standing by my side throughout this journey. My deepest gratitude goes to my mother. Her unconditional support, trust, patience and love have carried me through all phases of my life. Thank you for being so proud of me and for always being there for me.