Elise Sandsmark

Multi-Level Molecular Characterisation of Prostate Cancer

Elise Sandsmark

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Medicine and Health Sciences
Department of Circulation and Medical Imaging

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

NTNU

Elise Sandsmark

Multi-Level Molecular Characterisation of Prostate Cancer

Thesis for the Degree of Philosophiae Doctor

Trondheim, June 2017

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Circulation and Medical Imaging

NTNU
Norwegian University of
Science and Technology

Karakterisering av prostatakreft på molekylært nivå

Prostatakreft er den hyppigste kreftformen blant norske menn, og utgjør en betydelig helse-byrde. Sykdommen har varierende prognose, fra svært saktevoksende og snill, til aggressiv og dødelig. En av de største kliniske utfordringene er å skille mellom lav og høy aggressivitet ved diagnosetidspunktet, og dette fører til overbehandling av pasienter med snille kreftformer. Økt forståelse av de molekylære forskjellene mellom aggressiv og ikke-aggressiv kreft kan bidra til bedre risikoinndeling og behandlingsstrategier hos pasienter med prostatakreft.

Denne avhandlingen består av tre studier, hvor det overordnede målet var å øke kunnskapen om molekylære forskjeller innad i prostatakreft, samt identifisere potensielle biologiske markører for aggressivitet. Studiene inkluderte to pasientkohorter med vevsprøver som var samlet inn etter *prostatektomi* (kirurgisk behandling) fra samtykkende prostatakreftpasienter. For å få et mer helhetlig inntrykk av de molekylære prosessene ble hver vevsprøve undersøkt på flere måter. *Histologisk undersøkelse* med lysmikroskop ble gjort for å påvise kreft og gradere den etter det kliniske *Gleason score* systemet. Kreft med høy Gleason score avviker mest fra normalt vev og betraktes som mer aggressiv. *Magnetisk resonans (MR) spektroskopi* ble brukt til å studere metabolismen (stoffskiftet) i alle vevsprøvene. I prostatavev kan denne metoden måle nivået til ca. 25 metabolitter (mellomstadier eller produkter i metabolismen). En slik undersøkelse er relevant fordi endringer i metabolismen er et av kjennetegnene til kreft, og dette skjer fordi kreft-celler har økt behov for energi og byggeklosser sammenliknet med normale celler. *Genuttrykke*t representerer oppskriften til hvilke proteiner som kan produseres i cellene, og gir en indikasjon på hvilke molekylære prosesser som foregår i celler og vev. I denne avhandlingen ble det brukt *genuttrykk-analyse* i én kohort for å måle aktiviteten til de ulike genene. I den andre kohorten ble det gjort spesifikke analyser for å støtte funnene fra genuttrykk-analysen. Dette inkluderte undersøkelse av endringer på kromosomnivå med *fluorescensmikroskopi* (studie I), og analyse av spesifikke proteiner med en metode kalt *immunohistokjemi* (studie II og III). Pasientene ble fulgt opp i minst fem år, og *biokjemisk tilbakefall* av kreftsykdommen (stigning av prostataspesifikt antigen (PSA) i blodprøve) ble brukt som et mål på aggressivitet. For å validere funnene og øke styrken av funnene, ble det i studie II og III inkludert flere allment tilgjengelige kohorter med

genuttrykk og oppfølgningsdata fra prostatakreftpasienter. En oversikt over kohortene, metodene, og de tre studiene er vist i figur 1.
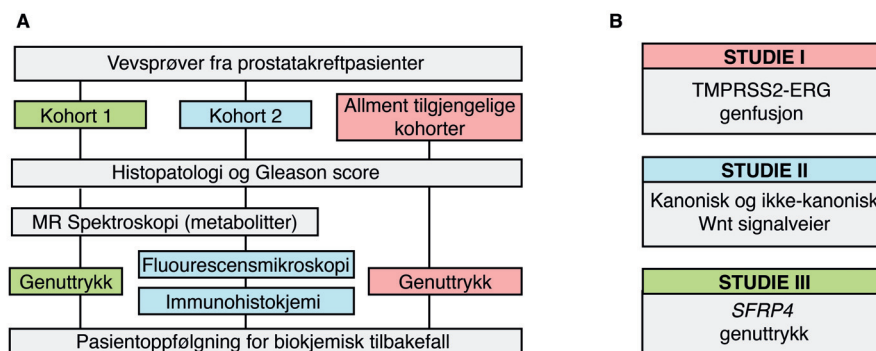
Figur 1 A. Oversikt over kohortene og metodene som ble brukt i avhandlingen. **B.** Oversikt over hva som ble undersøkt i de tre studiene.

I den første studien var målet å undersøke endringer i metabolismen ved tilstedeværelsen av en kjent genfusjon som finnes i ca. halvparten av prostatakreftsvulster. Denne genfusjonen heter TMPRSS2-ERG og fører til endringer i genuttrykket, men det er usikkert om dette gir en mer aggressiv krefttype. Ved hjelp av en etablert genuttrykkssignatur og fluorescensmikroskopi ble det funnet at vevsprøver med genfusjonen hadde endret metabolisme, hvor redusert konsentrasjon av metabolittene citrat og spermin var spesielt fremtredende. Prostatakjertelen produserer sædvæske som inneholder høye nivåer av både citrat og spermin. Lav konsentrasjon av nettopp disse metabolittene kan derfor tyde på at kreftcellene har mistet deler av normalfunksjonen til prostataceller, og har også tidligere blitt funnet i prostatakreft med høy Gleason score. Til sammen tyder dette på at prostatakreft med TMPRSS2-ERG genfusjon har en metabolisme som samsvarer med metabolismen til aggressiv prostatakreft.

Kreft kan oppnå aggressive egenskaper ved å aktivere ulike signalveier i cellene. I studie II ble genuttrykket til komponentene i en slik gruppe signalveier, kalt Wnt, studert. Generelt fører aktivering av Wnt til aggressive egenskaper som økt invadering av nabovev og metastasering (spredning) til andre organer i kreftceller. Signalveiene i Wnt blir ofte delt i to grupper: kanonisk og ikke-kanonisk Wnt. Begge undergruppene har vist relevans for kreft, men kanonisk Wnt er mest studert. I studie II ble det ikke funnet aktivering av kanonisk Wnt i prostatakreft, men det ble funnet økt genuttrykk av ikke-kanoniske Wnt komponenter i en undergruppe av vevsprøvene. En ny genuttrykkssignatur bestående av femten gener ble laget for å måle aktiveringen av denne signalveien. Signaturen var assosiert med høyere Gleason score, biokjemisk tilbakefall

av kreftsykdommen etter kirurgi og redusert konsentrasjon av metabolittene citrat og spermin. Disse funnene tyder på at prostatakreft med økt aktivering av ikke-kanonisk Wnt er mer aggressiv.

Et av genene i genuttrykkssignaturen fra studie II var *SFRP4*. For kreft generelt har *SFRP4* en hemmende effekt på Wnt signalveier, og undertrykker vekst og utvikling av kreftsvulster. Resultatene fra studie II tydet imidlertid på at dette ikke er tilfellet for prostatakreft. Målet i studie III var derfor å undersøke assosiasjonen mellom genuttrykket av *SFRP4* og aggressivitet av prostatakreft. For å øke styrken på studien ble vevsprøver fra åtte unike pasientkohorter, med totalt 1884 pasienter, undersøkt. Resultatene viste at økt genuttrykk av *SFRP4* var assosiert med mer aggressiv kreft med lavere nivåer av citrat og spermin, høyere Gleason score og biokjemisk tilbakefall og metastaser etter kirurgisk behandling. *SFRP4* er derfor en potensiell klinisk markør for aggressivitet av prostatakreft og bør studeres videre.

Oppsummert viser funnene i denne avhandlingen at TMPRSS2-ERG genfusjon, signaturen for ikke-kanonisk Wnt signalvei og genuttrykk av *SFRP4* er assosiert med endringer i metabolismen til prostatakreft, med redusert konsentrasjon av metabolittene citrat og spermin. Dette er metabolitter som også kan måles i pasienter med en vanlig MR-skanner, og har dermed potensiale som ikke-invasive biologiske markører. Videre var signaturen for ikke-kanonisk Wnt signalvei og genuttrykket av *SFRP4* assosiert med aggressiv prostatakreft med hyppigere tilbakefall etter behandling. Signaturen og *SFRP4* kan derfor være mulige markører for å skille mellom aggressiv og ikke-aggressiv prostatakreft på diagnosetidspunktet, og dette potensialet bør undersøkes i fremtidige studier.

Kandidatens navn:	Elise Sandsmark
Institutt:	Institutt for sirkulasjon og bildediagnostikk
Veiledere:	May-Britt Tessem (hovedveileder)
	Kirsten M. Selnæs og Tone F. Bathen (biveiledere)
Finansieringskilde:	Norges teknisk-naturvitenskapelige universitet (NTNU)

Ovennevnte avhandling er funnet verdig til å forsvares
offentlig for graden Philosophiae Doctor (PhD) i medisin.
Disputas finner sted i auditoriet på MTFS, NTNU,
Onsdag 21. juni 2017 kl. 12.15.

Acknowledgements

immunohistochemistry, and to Ailin F. Hansen for your good and invaluable collaboration.

I am grateful for all the good colleagues, and the great working and social environment at the MR Centre and in the MR Cancer Group. I consider myself extremely lucky to be surrounded by such a good group of people every day. I especially want to thank my office mates over the years: Trygve, Ailin, Saurabh, Tonje, Marie, Leslie, Maria K, Torfinn, and Liv, for always being helpful, for many good scientific and non-scientific discussions, afternoon orange breaks, late night tea breaks, and never ending puns... Torill also deserves a special thanks for taking care of me, everyone, and everything, and for always knowing who to contact if you cannot fix it yourself. A special thanks also goes to Debbie for always being there when I needed to talk, and for invaluable English and scientific feedback and discussions. Furthermore, I would like to especially thank Anne Line for always being there for me for everything from scientific and medical school questions and listening to my complaints, to dinner breaks and weight lifting. I truly feel lucky for having you as a colleague and, even more, as a friend, thank you.

Finally, this thesis could not have been completed without the love and support from my friends and family. A special thanks to everyone involved in NTNUI Rowing, a truly amazing student club which has given me friends and memories for life, as well as good training sessions to keep me sane throughout the most intense periods of my PhD work. To my parents, Elin and Erik, thank you for always believing in me, for encouraging me to be curious and creative, and for being hard-working and inspiring role models throughout my life. To my grandparents for endless support, love, and encouragement. To my sisters, Kristine, Helene, and Margrete, for always being there for me, but also for making me tougher by some healthy competition and arguments growing up. And lastly, to my nephew, Kristian, for brightening up my days and reminding me there is more to life.

Tusen hjertelig takk,
Trondheim, June 2017

Elise Sandsmark

Summary

Prostate cancer is the most common malignancy in Norwegian men, and represents a substantial health burden. The disease is heterogeneous, ranging from slow growing and indolent, to very aggressive and lethal. One of the major unsolved clinical challenges is to accurately separate indolent from harmful disease at an early time point. This causes substantial overtreatment of patients with harmless cancers, as well as undertreatment of patients with aggressive cancers. To enable improved treatment selection, an increased understanding of the molecular characteristics of prostate cancer progression is needed. In this thesis, multi-level molecular analyses of gene expression and metabolism were performed in an integrated fashion on prostate tissue samples. The aim was to obtain more comprehensive knowledge of prostate cancer aggressiveness, and to identify candidate biomarkers for improved risk stratification of prostate cancer patients.

Gene expression analysis is a method that detects active genes; it can indicate which molecular processes occur in cells and tissue. The expression of genes is the instruction for which proteins are produced in the cells. Proteins are components of cellular signalling pathways, where the pathway activity can be altered to favour cancer survival. Activation of the Wnt signalling pathway may increase the cells' motility, and can therefore be exploited by cancer cells to gain invasive and metastatic properties. The work in this thesis showed increased activation of a subgroup of the Wnt pathway, called the non-canonical Wnt pathway. By using a set of genes representing the non-canonical Wnt pathway (NCWP), combined with markers of increased cell mobility (epithelial-mesenchymal transition (EMT)), a gene signature coined NCWP-EMT was developed. An increased signature score suggests increased activation of this pathway. High signature score, representing increased activation of the pathway, was associated with aggressive cancer, where more patients experienced recurrent and metastatic disease after surgery. The signature may therefore have clinical potential to improve the discrimination of aggressive from indolent prostate cancer at an early time point.

One of the signature members, secreted frizzled-related protein 4 (*SFRP4*), was further investigated on its own. The expression level of *SFRP4* was shown to be a predictor for aggressive,

recurrent and metastatic disease, and this was validated in several independent patient cohorts, and in a total of 1884 patients. *SFRP4* alone, may therefore have potential as a biomarker for prediction of prostate cancer outcome.

Changes in the genome can alter gene expression, and an example of this is a fusion of two genes, called TMPRSS2-ERG. This gene fusion is found in approximately half of malignant prostate tumours, however, little is known about its relation to other molecular processes, such as cancer cell metabolism. In this thesis, a distinctive metabolic profile was seen in cancer tissue possessing TMPRSS2-ERG, and this profile was similar to metabolic alterations previously observed in aggressive prostate cancer.

Metabolism in tissues and cells can be studied by magnetic resonance (MR) spectroscopy. Cancer cell metabolism differ from healthy cells, as cancer often prioritise growth which require increased energy production and synthesis of new building blocks. Reprogramming of metabolism is therefore regarded as one of the hallmarks of cancer cells. The normal prostate cells produce and excrete high levels of the metabolite citrate for the prostatic fluid. Previously, a reduced levels of citrate have been detected in prostate cancer compared with healthy tissue, and this is likely due to citrate being used for energy and fatty acid production, rather than production and excretion. Furthermore, alterations to polyamine metabolism, and in particular to spermine, are important in prostate cancer, where decreased spermine concentration has been associated with the disease. In this thesis, reduced concentrations of both citrate and spermine were detected in cancer tissue samples containing the TMPRSS2-ERG gene fusion, samples with a high score of the non-canonical Wnt pathway signature, and samples with a high expression level of *SFRP4*. This suggests that citrate and spermine have great potential as tissue biomarkers of prostate cancer. Importantly, these metabolic alterations were also detected by non-invasive patient MR examination, which is therefore a candidate as a prognostic tool in prostate cancer diagnosis.

To summarise, the work presented in this thesis shows that the TMPRSS2-ERG gene fusion, the non-canonical Wnt pathway, and *SFRP4* expression are all associated with reprogramming of prostate cancer metabolism. Additionally, activation of the non-canonical Wnt pathway and the expression level of *SFRP4* were associated with recurrent and metastatic disease after surgery. Further investigation of these aggressive molecular characteristics may lead to clinical biomarkers for improved early risk stratification in prostate cancer patients.

List of papers

Paper I

Presence of TMPRSS2-ERG is associated with alterations of the metabolic profile in human prostate cancer.

Hansen AF, <u>Sandsmark E</u>, Rye MB, Wright AJ, Bertilsson H, Richardsen E, Viset T, Bofin AM, Angelsen A, Selnæs KM, Bathen TF, Tessem MB.

Oncotarget. 2016 Jul 5;7(27):42071-42085.

Paper II

A novel non-canonical Wnt signature for prostate cancer aggressiveness.

<u>Sandsmark E</u>, Hansen AF, Selnæs KM, Bertilsson H, Bofin AB, Wright AJ, Viset T, Richardsen E, Drabløs F, Bathen TF, Tessem MB, Rye MB.

Oncotarget. 2017 Feb 7;8(6):9572-9586.

Paper III

***SFRP4* gene expression is increased in aggressive prostate cancer**

<u>Sandsmark E</u>, Andersen MK, Bofin AM, Bertilsson H, Drabløs F, Bathen TF, Rye MB, Tessem MB.

Manuscript.

Contents

Contents

List of Figures

List of Tables

Definitions and Abbreviations

Gene symbols	Symbols for genes are in uppercase and italicised (e.g. *SFRP4*)
Protein symbols	Symbols for proteins are in uppercase and not italicised (e.g. SFRP4)

^{1}H	Proton
ACACA	Acetyl-CoA carboxylase alpha
ACLY	Adenosine triphosphate citrate lyase
ACO1/2	Aconitase 1/2
ACON	Aconitase
APC	Adenomatous polyposis coli
AR	Androgen receptor
ATP	Adenosine triphosphate
B_0	External magnetic field
BPH	Benign prostate hyperplasia
BRCA1-2	Breast cancer 1-2
Ca^{2+}	Calcium ions
Cadherins	Calcium-dependent adhesion proteins
cDNA	Complementary DNA
Cohen's *d*	Standardised effect size of the mean difference between two groups
CPMG	Carr-Purcell-Meiboom-Gill
cRNA	Complementary RNA
D_2O	Deuterium oxide
DHT	Dihydrotestosterone
DNA	Deoxyribonucleic acid – contains the genetic information
E-cadherin	Epithelial cadherins, marker for epithelial cells
EMT	Epithelial-mesenchymal transition
ERG	ETS-related gene
ETS	Erythroblast transformation-specific transcription factor
Ex vivo	Out of the living (*Latin*)

Definitions and Abbreviations

FASN	Fatty acid synthase
FID	Free induction decay
FISH	Fluorescence *in situ* hybridisation
FZD2	Frizzled2
GSEA	Gene set enrichment analysis
HE	Haematoxylin and Eosin
HES	Haematoxylin Eosin Saffron
HOXB13	Homeobox B13
HR-MAS MRS	High resolution magic angle spinning magnetic resonance spectroscopy
IHC	Immunohistochemistry
In situ	In the original place *(Latin)*
In vivo	Within the living (*Latin*)
LCModel	Linear combination of Model spectra
M stage	Distant metastasis, part of TNM staging
MALDI MS	Matrix-assisted laser desorption ionisation mass spectrometry
Microtome	Instrument for cutting extremely thin section of material
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MRSI	Magnetic resonance spectroscopic imaging
MRS	Magnetic resonance spectroscopy
MS	Mass spectrometry
N stage	Regional lymph node involvement, part of TNM staging
N-cadherin	Neural cadherins, markers for mesenchymal cells
NCWP-EMT	Non-canonical Wnt pathway and epithelial to mesenchymal transition
NGS	Next-generation DNA sequencing
NOESY	Nuclear Overhauser effect spectroscopy
ODC1	Ornithine decarboxylase
PCA	Principal component analysis
PGLS	6-phosphogluconolactonase
PH	Proportional hazard
PLS-DA	Partial least squares discriminant analysis
ppm	Parts per million
PRESS	Point-resolved spatially localised spectroscopy
Prostatectomy	Surgical removal of the prostate gland
PSA	Prostate specific antigen
RBKS	Ribokinase

RF	Radio frequency
RNA-seq	RNA-sequencing - next generation
RNA	Ribonucleic acid – the transcript of DNA
SAT1	Spermine/spermidine N1-acetyltransferase 1
SFRP	Secreted frizzled-related protein (1-5)
SI	Staining index
SMOX	Spermine oxidase
SMS	Spermine synthase
SRM	Spermidine synthase
ssGSEA	Single sample gene set enrichment analysis
T stage	Primary tumour extent, part of TNM staging
TCA	Tricarboxylic acid
TKT	Transketolase
TMA	Tissue microarray
TMPRSS2	Transmembrane protease serine 2
TNM	Tumour extent, Node involvement, Metastases - cancer staging system
TRUS	Trans-rectal ultrasound
TURP	Transurethral resection of the prostate

1. Introduction

Prostate cancer is a heterogeneous disease, ranging from slow growing and indolent, to very aggressive. One of the major unsolved clinical challenges is to accurately separate indolent from harmful cancer at an early time point. This causes substantial overtreatment of patients with harmless cancers, as well as undertreatment of patients with aggressive disease. Increased understanding of the molecular mechanisms of prostate cancer progression is therefore needed in order to develop of new prognostic biomarkers for improved risk stratification of patients.

1.1 Prostate Cancer Characteristics

Normal Anatomy and Function

The prostate is a walnut sized exocrine gland of the male reproductive system. It surrounds the uppermost part of the urethra, and is located between the bladder neck and the pelvic floor, close to the ventral wall of the rectum (Figure 1.1). The prostate is divided into four histological regions: the peripheral zone, central zone, transition zone, and anterior fibromuscular stroma, where the peripheral zone compromises approximately 70% of the gland in young men (Figure 1.1) [1]. The main function of the prostate is to produce and secrete prostatic fluid, one of the components of semen along with spermatozoa and seminal vesicle fluid. The prostatic fluid has high concentrations of the metabolites citrate and polyamines, which are essential for energy supply and motility of the spermatocytes [2–4].

Epidemiology

Prostate cancer is the second most common malignancy, and the fifth leading cause of cancer related death among men, worldwide [5]. In 2015, 29% of all cancer diagnosed among men in Norway were prostate cancers, making it the most frequent cancer in men [6]. In Norway, approximately one in seven men will develop prostate cancer by the age of 75 (cumulative risk of 13.6%). The incidence in Norway has increased 4-fold in the past 60 years, however,
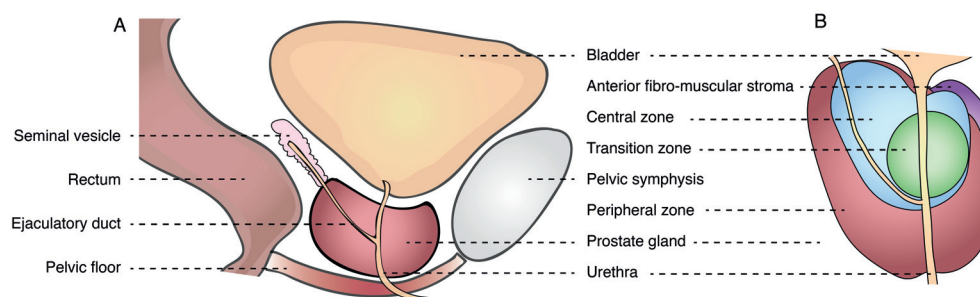
Figure 1.1 Anatomical location and zones of the prostate gland.
Figure A shows the anatomically location of the prostate, below the bladder and above the pelvic floor in men. Figure B shows the four zones of the prostate, namely the peripheral zone, transition zone, central zone, and anterior fibro-muscular stroma. Figure A and B are both sagittal views of the prostate. *The figure is an illustration.*

an incidence plateau has been reached over the last 10 years (Figure 1.2). Similar trends are seen in the Western world. In addition to longer life expectancy, this can mainly be attributed the increased use of PSA testing from the late 1980s, resulting in earlier detection as well as increased diagnosis of asymptomatic and indolent disease [5, 7].

Norway has one of the highest prostate cancer mortality rates in Europe [7, 8] with 1 093 deaths in 2014, and in the same year prostate cancer was for the first time surpassing lung cancer in age-standardised mortality rates [6]. The mortality rate has, however, slowly declined from the mid 1990s as shown in Figure 1.2. This decline can be explained by earlier detection and improvements in curative treatment [6, 8]. The combination of increased incidence rate, detection of indolent cancer, and reduced mortality rate is reflected in the five-year survival rate, which has increased from 53.4% in 1976, to 92.9% in 2015 (Figure 1.2).

Risk Factors

Age, ethnicity, and family history are established risk factors for prostate cancer. Considering all malignancies, prostate cancer has one of the strongest relationships with age, and the median age at diagnosis in Norway is sixty-nine [9, 10]. Men of African descent have a higher risk of developing prostate cancer compared with white men across the world [9, 11, 12]. Although the reason behind this disparity is poorly understood, the global extent implicates genetic susceptibility [12]. Asian countries have the lowest incidence in the world [13], and this is likely caused by differences in diagnostic practise, genetic susceptibility, as well as environmental and lifestyle factors [14]. Family history is an important risk factor for prostate cancer; men
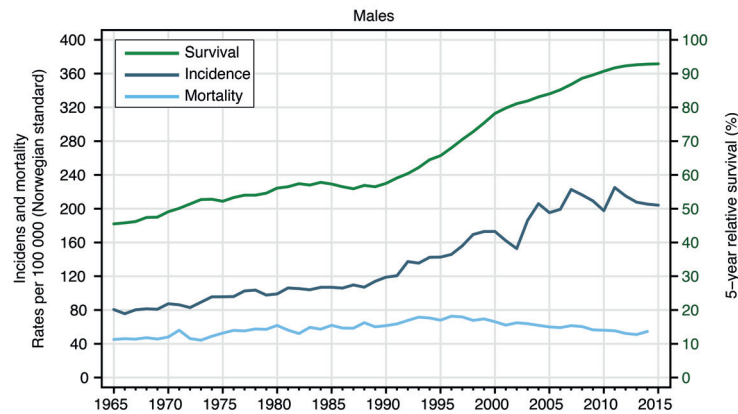
Figure 1.2 Trends in incidence, mortality, and survival of prostate cancer.
Incidence, mortality, and 5-year relative survival rate of prostate cancer for the last 50 years in Norway. From the 1990s, the incidence and survival have increased, whereas the mortality has declined. *Adapted from "Cancer in Norway 2015" [6], with permission from the Cancer Registry of Norway.*

with one affected first degree relative have a twofold increased risk, and the risk increases further with additional affected relatives [15]. While this evidence suggests a strong genetic component in prostate cancer, identification of specific gene mutations and alterations has proven challenging. Mutations of *BRCA1* or *BRCA2*, breast and ovarian cancer susceptibility genes, and the homeobox 13 (*HOXB13* gene), have been associated with an increased risk of prostate cancer [16–19]. However, these genetic mutations can only explain a small proportion of the family clustering. Although other loci have also been identified, there has been limited success in identifying high-risk susceptibility genes, reflecting the complexity of the disease.

Clinical Presentation and Diagnostic Tools

Cancers of the prostate are frequently asymptomatic at the time of diagnosis. Local progression of cancer may give symptoms such as lower urinary tract obstruction and haematuria, although other causes are more likely. Bone pain may be the presenting symptom in men with metastatic disease, however, initial presentation of metastatic prostate cancer is rare today, and reported to be ~7% of prostate cancer patients in Norway [6]. Similar numbers are seen in the United States, where 90% of new incidents have been reported as localised or regional cancers [20].

The main diagnostic tools for detection of prostate cancer are blood serum PSA and digital rectal examination. The PSA blood test was originally introduced as a biomarker to monitor prostate cancer recurrence and progression following treatment [21]. In the mid-1980s, PSA was adapted

as a test for initial detection of prostate cancer. This resulted in drastically improved diagnostics, and reduced the number of men with metastatic disease at initial presentation [22–24]. Normal glandular prostate cells produce and excrete PSA into their luminal space for the prostatic fluid. In cancer, tissue barriers may be disrupted, causing PSA to leak into the circulating blood, thus increasing serum PSA levels. However, indolent cancers, as well as non-cancerous conditions such as benign prostate hyperplasia (BPH), urinary retention, and prostatitis can also elevate PSA levels. The PSA blood test is therefore not a specific marker of prostate cancer [25], and screening for early detection is thus controversial. Many comprehensive studies have investigated the benefits and disadvantages of PSA screening, including the 2013 Cochrane review, which showed no significant decrease in prostate cancer-specific and overall mortality rates as a result of PSA screening [26]. Furthermore, the benefits of screening are shown to be outweighed by the risk of overdiagnosis, overtreatment and the associated morbidity [27, 28]; mathematical models have estimated that 23-42% of all PSA screening detected cancers are overdiagnosed [29]. The Norwegian and European Guidelines, as well as the United States Preventive Services Task Force, all recommend against population based PSA screening [30–32]. PSA screening after informed decision by patients is still widely practised, and, in Norway, an increasing number of prostate cancer patients are initially referred to secondary care (urologists) due to elevated PSA levels alone [10]. Additionally, the preoperative PSA level is not a satisfactory indicator of aggressiveness, and only shows a poor correlation with postoperative histopathological grade [33]. In fact, the poorly differentiated, thus aggressive, cancer cells may lose their PSA producing properties, and lower levels of PSA have been detected in very aggressive prostate cancer [34].

The other main diagnostic tool for detection of prostate cancer is digital rectal examination, which can detect tumours in the posterior and lateral part of the gland. However, this technique has several shortcomings; approximately 25% of cancers arise in non-palpable zones of the prostate, localised cancers are usually non-palpable, and digital rectal examination is highly subjective with poor inter-examiner reliability [35]. Digital rectal examination is still of importance as some clinically aggressive cancer are detectable by this method, without having elevated PSA levels [36].

Histopathological confirmation of the diagnosis is performed on trans-rectal ultrasound (TRUS) guided biopsies. Negative biopsies do not exclude prostate cancer; the false negative rate of a 12-core biopsy procedure was reported to be above 30% [37]. As a result of this, many patients undergo repeated biopsy sessions. Another weakness of TRUS biopsies is that they do not necessarily represent the most aggressive part of the tumour, and a meta-analysis showed that 30% of cancers were histopathologically upgraded after surgery [38]. To improve the accuracy

4

of diagnostic biopsies, it has become more common to perform a magnetic resonance imaging (MRI) examination before the biopsy procedure. The MRI can then guide the biopsy procedure either by cognitive fusion, software based MRI-ultrasound fusion, or directly in the MRI scanner [39].

Histopathological Evaluation

Most cancers of the prostate are adenocarcinomas arising from the glandular prostate components. In 1966, Donald F. Gleason described a grading system based on the tissue architecture [40], and this system was revised in 2005 by consensus of the International society of urological pathologist [41]. The Gleason grading system consists of histopathological patterns graded from well-differentiated grade 1 to poorly-differentiated grade 5 (Figure 1.3), where grade 1 and 2 are not considered to be cancer and are rarely used. Prostate cancer is often heterogeneous; and a scoring system of the first and the second most dominant Gleason grades are used to obtain a Gleason score. A less common, but higher grade pattern, is reported as the secondary grade in needle biopsies, and as tertiary grade in prostatectomy specimens, as these have additional prognostic value [42]. Gleason score is one of the strongest predictors of prostate cancer progression [43–45], where cancers with Gleason score 8-10 have high metastatic potential [46]. Risk prediction for Gleason score 7, representing the major proportion of cancers, is more challenging. Although a division into Gleason score 3+4 and 4+3 have shown some prognostic differences [47, 48], this challenge still remains. Gleason score 6 (3+3) has low metastatic potential [49, 50], and there is debate as to whether Gleason score 6 should be defined as cancer [51–55]. However, for patients, a Gleason score of 6 out of 10 may appear as a high number, and this is a flaw of the Gleason grading system. A new system using the terminology Grade Group was proposed by the International society of urological pathologist in 2016 [56]. In this new system, the Grade Groups 1-5 will correspond to the old Gleason scores 6, 3+4, 4+3, 8, and 9-10, respectively. Thus, Gleason score 6 will be Grade Group 1, potentially lowering fear and overtreatment. Gleason score 7 (3+4 and 4+3) will also be distinguished into Grade Group 2 and 3, respectively, acknowledging their prognostic differences.

Staging, Risk Stratification, and Treatment

As prostate cancer ranges from indolent to lethal, correct classification and risk stratification are highly important to provide the appropriate treatment for each patient. Prostate cancer is staged according to the primary tumour (T), regional lymph node (N), and distant metastasis (M) – TNM classification system (Table 1.1). The clinical T stage is based on digital rectal examination, number and sites of positive TRUS biopsies, and, when available, MRI of the prostate. To further
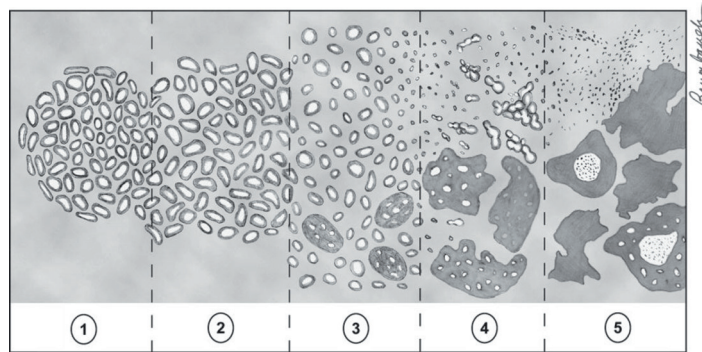
Figure 1.3 Gleason grading system as modified in 2005.
Pattern 1-2 consist of well to moderate differentiated medium-sized glands, and are rarely reported in biopsies. Pattern 3 has moderately differentiated, still recognisable glands, typically smaller than pattern 1 and 2, but varies in size and shape, and infiltrates in and amongst benign acini. Pattern 4 has poorly differentiated, ill-defined, and often fused glands, with poorly formed lumina. Pattern 5 has no glandular differentiation. *Reprinted from Epstein et al. [41], with permission from The American Journal of Surgical Pathology (Wolters Kluwer Health).*

Table 1.1 Tumour Node Metastasis (TNM) classification of prostate cancer.

Primary tumour (**T**)	
T0	No evidence of primary tumour
T1	Clinically inapparent tumour not palpable or visible by imaging
T2	Tumour confined within prostate
T2a	Tumour involves one-half of one lobe or less
T2b	Tumour involves more than one-half of one lobe, but not both lobes
T2c	Tumour involves both lobes
T3	Tumour extends through the prostatic capsule
T4	Tumour fixed or invades adjacent structures other than seminal vesicles
Regional lymph nodes (N)	
N0	No regional lymph node metastasis
N1	Metastasis in regional lymph node(s)
Distant metastasis (M)	
M0	No distant metastasis
M1	Distant metastasis

The table is adapted and simplified from the AJCC Cancer Staging manual [57].

Table 1.2 Risk stratification of prostate cancer.

Risk	cT stage		PSA		Gleason score
Low risk	$\leq$ T2a	AND	<10 ng/mL	AND	$\leq$ 6
Intermediate risk	T2b-c	OR	$10-20$ ng/mL	OR	=7
High risk	$\geq$T3a	OR	$\geq$20	OR	8-10

cT stage – Clinical tumour stage. PSA – prostate specific antigen.
The table is adapted from the Norwegian prostate cancer guidelines [30].

assist treatment decision, tables combining clinical T stage with Gleason score and PSA value are used to stratify cancers as low, intermediate, and high risk (Table 1.2). For treatment selection, the risk, age, and general health condition of the patient, as well as the patient's own preferences are taken into account. Treatment options generally include radical treatment for curable patients (surgical radical prostatectomy or radiation therapy), active surveillance for a selected group of patients with very low risk, and palliative treatment for patient with advanced and disseminated disease [30, 31].

Potential errors in risk classification arise due to shortcomings of the methods. Briefly summarised, important limitations include underestimation of T stage by digital rectal examination, under sampling for TRUS biopsies affecting both T stage and Gleason score, and low interobserver reproducibility for Gleason score assessment. In a recent study of almost 26 000 prostate cancer patients, Caster et al., highlighted some of these challenges [58]. They detected, among others, that 43% of patients with low risk cancer (biopsy Gleason score of 6, and pre-treatment PSA of <10 ng/mL), were pathologically upgraded after surgery [58]. Although today's risk stratification system improves treatment selection, the shortcomings imply a need for more accurate variables for optimal treatment selection for prostate cancer patients.

Recurrence after Surgical Treatment

If all prostate tissue is removed by radical prostatectomy, the PSA serum level is expected to be undetectable within six weeks after surgery, an absence of this or a detection of increased PSA levels, may indicate the presence of remaining prostate tissue, or loco-regional and systemic cancer recurrence [59]. Regular PSA measurements are therefore an important part of patient follow-up after surgical treatment. Biochemical recurrence is defined as serum PSA$\geq$ 0.02 ng/mL in two independent measurements [30, 31]. Approximately 90% of biochemical recurrences occur within the first five years following surgery, where 20-30% of all patients experience biochemical recurrence [31, 60]. Due to the low mortality and long survival times of prostate

cancer patients, biochemical recurrence is a frequently used surrogate endpoint in statistical survival analyses. For interpretation of such analyses, it is important to recognise that only a minority of patients with biochemical recurrence will develop clinical recurrence or die of prostate cancer [31, 61].

1.2 Molecular Alterations in Prostate Cancer

Epithelial-Mesenchymal Transition (EMT)

In the prostate, it is the glandular epithelial cells that may give rise to prostate cancer (adenocarcinoma). Epithelial cells are well structured due to cell-to-cell adhesion, among others formed by the adhesion protein E-cadherin (epithelial), whereas mesenchymal cells are loosely organised. Epithelial-mesenchymal transition (EMT) is a process where epithelial cells become more motile and acquire mesenchymal properties and markers such as N-cadherin (neural) (Figure 1.4). EMT is essential during embryogenic development of different types of cells [62]. However, cancer cells of epithelial origin may later take advantage of the same transition for tumour invasion and metastasis [62]. In prostate cancer, there is increasing evidence associating EMT with cancer aggressiveness [63]. In particular, a switch from E-cadherin to the N-cadherin has been linked to poor prognosis in prostate cancer patients when investigated by immunohistochemistry in prostatectomy specimens [64].
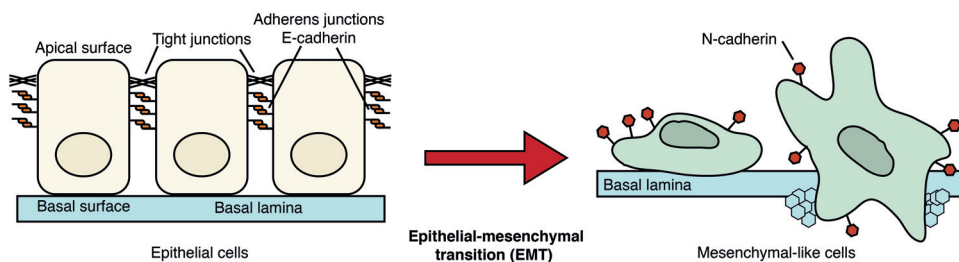
Figure 1.4 Epithelial-mesenchymal transition (EMT) in cancer cells.
Glandular epithelial cells in the normal prostate are systematically arranged with the basal surface towards the basal lamina, and the apical surface towards the lumen; forming well structured glands. The cells are connected by adherens and tight junctions. E-cadherin is a crucial part of the adherens junction. During EMT, the cell-to-cell cohesion is lost, E-cadherin is downregulated, and the cells start to express mesenchymal markers (such as N-cadherin), and can invade the basal lamina.

The Wnt Signalling Pathway

In addition to being a regulator of EMT, the Wnt signalling pathway is critical in embryogenesis and tissue homeostasis [65, 66]. Malfunction of the Wnt pathway has been established in numerous diseases, including cancer development, progression, and metastasis [67]. The discovery of Wnt signalling started with two independent identifications of the Wnt pathway ligand, Wnt1; first in embryogenesis in 1980 [68], then as a proto-oncogene in 1982 [69], which in 1987 were proven to be identical genes [70]. Today, a total of 19 Wnt ligands that can activate the pathway have been identified in humans [71].

Wnt signalling can be divided into canonical and non-canonical pathways. The canonical pathway is frequently called β-catenin dependent, due to β-catenin's crucial role in canonical signalling. Activation of the canonical pathway inhibits the destruction complex in labelling β-catenin for proteasomal degradation, and as a result β-catenin is stabilised and translocated to the nucleus. Nuclear expression of β-catenin is therefore a hallmark of canonical Wnt activation. The nuclear β-catenin activates specific transcription factors promoting EMT, as well as cell proliferation, differentiation and survival (Figure 1.5A). The importance of the canonical pathway in carcinogenesis was first discovered in colorectal cancer, where somatic and inherited mutations of the adenomatous polyposis coli (*APC*) gene are common [72] and cause structural changes of the APC protein. APC is one of the main components of the β-catenin destruction complex, and such changes can make the complex defective, hence activate the downstream pathway [73, 74]. *APC* mutations are rare in human prostate cancer [75], however, dysregulation and activation of the canonical pathway have still been detected in prostate cancer cell lines, where canonical signalling has been associated with advanced, metastatic and androgen resistant features [76–78]. Studies of human prostate cancer tissue samples have been less consistent. Chen et al. detected increased cytoplasmic and nuclear β-catenin immunohistochemistry staining, indicating activation of the pathway, in 36 % of the prostate cancers [79], whereas Bismar et al. observed no nuclear staining of β-catenin [80]. Further investigations of the canonical Wnt pathway in human prostate cancer is therefore needed.

The non-canonical Wnt pathway is commonly divided into two sub-pathways, the planar cell polarity and the Wnt/Calcium pathway (Figure 1.5B-C). The non-canonical signalling pathways have been less thoroughly studied in prostate cancer, however, a study by Wang et al. detected increased activity in the Wnt/Calcium pathway to be associated with cytoskeleton remodelling and cell motility in prostate cancer cell lines [81]. In addition, the non-canonical Wnt ligand, WNT5A, has been suggested as a prostate cancer biomarker, and was reported to be upregulated

9

in several studies of human prostate cancer tissue [82–84]. There are, however, inconsistent results of the prognostic association of WNT5A expression; one study found a correlation between WNT5A expression and poor prognosis [82], whereas other studies have detected an association with good prognosis [83–85].

An additional non-canonical pathway, Wnt/Fzd2, was discovered by Gujral et al. in 2014 (Figure 1.5D) [66]. Activation of this pathway, by WNT5 binding to the frizzled2 (FZD2) receptor, was reported to induce tumour progression and epithelial-mesenchymal transition in breast, colon, lung, and hepatocellular cancers [66]. Gujral et al. also identified that knockdown of *Fzd2* in mice models resulted in reduced tumour growth, and that a signature of central Wnt5/Fzd2 genes could accurately predict metastasis and survival of hepatocellular carcinoma patients. The Wnt/Fzd2 pathway has not previously been investigated in prostate cancer.
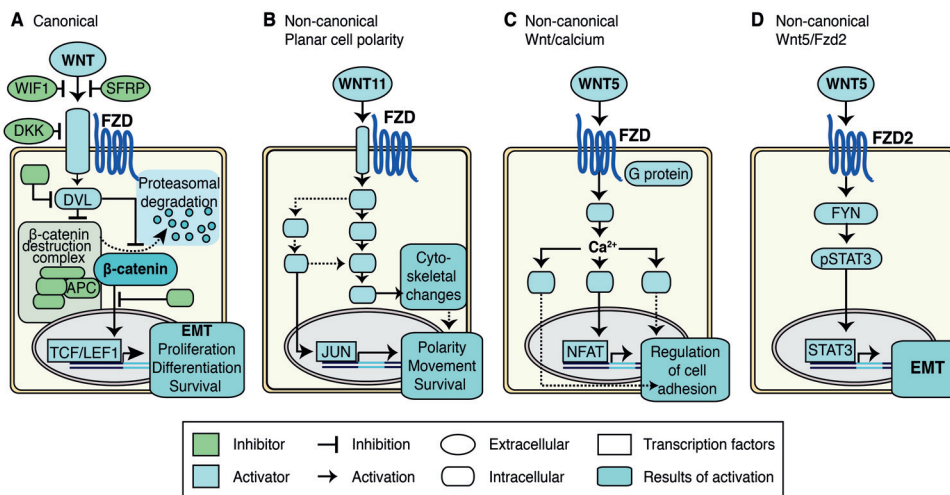
Figure 1.5 Schematics of the Wnt signalling pathways
(**A**) Activation of the canonical Wnt pathway causes nuclear translocation of β-catenin, promotes epithelial-mesenchymal transition and proliferation, and regulates cell survival. (**B**) The non-canonical planar cell polarity pathway regulates cell polarity, movement and survival. (**C**) Signalling from the non-canonical Wnt/Calcium pathway affects cell adhesion. (**D**) Activation of the non-canonical Wnt5/Fzd2 pathway induces epithelial-mesenchymal transition. The figure is simplified, for full figure with all protein/gene symbols, see paper II in this thesis. *The figure is modified from Paper II with permission/Creative Commons Attribution License* [86].

Secreted Frizzled-Related Protein 4 (SFRP4)

The family of secreted frizzled-related proteins (SFRP) are extracellular modulators of Wnt signalling (Figure 1.5A). Following their discovery in 1996 [87], five different members of the SFRP family have been identified in humans (SFRP1–5). The SFRPs can bind to Wnt ligands and frizzled receptors, both of which may inhibit Wnt signalling. This Wnt antagonist role of the SFRPs, combined with frequent silencing of their genes in cancer, have made SFRPs putative tumour suppressors. The SFRPs have, however, also been found to interact in other cell signalling pathways where they may have more aggressive properties [88]. Additionally, it has been suggested that some SFRPs may even activate Frizzled receptors, thus triggering Wnt signalling [88, 89].

SFRP4 is the largest member of the SFRP family, and is structurally the most different from the other family members [90]. In cancers, *SFRP4* is frequently hypermethylated and downregulated, and it is investigated as a possible therapeutic agent for cancers [91]. In prostate cancer, however, an opposite pattern of *SFRP4* is seen, where *SFRP4* gene expression has not only been detected upregulated, but also associated with aggressive and recurrent disease [92–94]. Thus, SFRP4 in prostate cancer does not seem to follow the presumed tumour suppressor role. Only two prostate cancer study cohorts have been investigated for protein expression of SFRP4 by immunohistochemistry, and the results are conflicting. Horvath et al. reported increased membranous expression to be associated with good prognosis [95, 96], whereas Mortensen et al. reported cytoplasmic expression to be associated with worse prognosis [94]. There is a need for clarifying the role of SFRP4 in prostate cancer.

TMPRSS2-ERG Gene Fusion

In 2005, Tomlins et al. identified a recurrent gene fusion in prostate cancer [97]. This fusion was between the promotor of the transmembrane protease serine 2 (*TMPRSS2*), and the coding region of the erythroblast transformation-specific (*ETS*) transcription factor ETS-related gene *ERG*, combined termed TMPRSS2-ERG. Normal prostate express *TMPRSS2*, and the promoter of the gene is positively regulated by androgens [98]. ERG is essential during embryogenesis, and continues to regulate systems such as angiogenesis in adults, however, *ERG* is not normally expressed in prostate epithelial cells [99–101]. The TMPRSS2-ERG gene fusion results in the TMPRSS2 promotor activating transcription of *ERG* (Figure 1.6). Overexpression of *ERG* is a frequent finding in prostate cancer, and the prevalence of TMPRSS2-ERG fusion has been reported in 30-80% of cancers [102]. The overexpression of *ERG* has been shown to induce EMT through the canonical Wnt pathway, in fusion positive prostate cancer cell lines [103]. A

11

gene expression signature for TMPRSS2-ERG has been established, and has shown potential for subtyping of prostate cancers [104, 105].

Although the discovery of TMPRSS2-ERG generated much enthusiasm and hope as a marker of aggressiveness in prostate cancer, studies on prognostic outcome related to the gene fusion have been inconsistent. Several studies have identified an association between the gene fusion and markers of poor prognosis in prostate cancer [106, 107]; a study of 445 conservatively treated prostate cancer patients reported a cause-specific 8-year survival of only 25% in fusion positive patients, contrasting to 90% 8-year survival of fusion negative patients [108]. However, a large meta-analysis of more than 5 000 patients, did not find TMPRSS2-ERG fusion to be associated with recurrent disease or cancer-specific death [109].

Another branch of TMRPSS2-ERG studies, focuses on understanding the mechanisms of prostate cancer heterogeneity by looking into differences between positive and negative gene fusion cancers. The TMPRSS2-ERG gene fusion has been linked to fatty acid oxidation, and increased glucose uptake [110, 111]. However, the metabolic alterations associated with TMPRSS2-ERG fusion are still largely unknown. Further insight into the molecular processes, such as reprogramming of metabolism, may increase the understanding of the gene fusion.
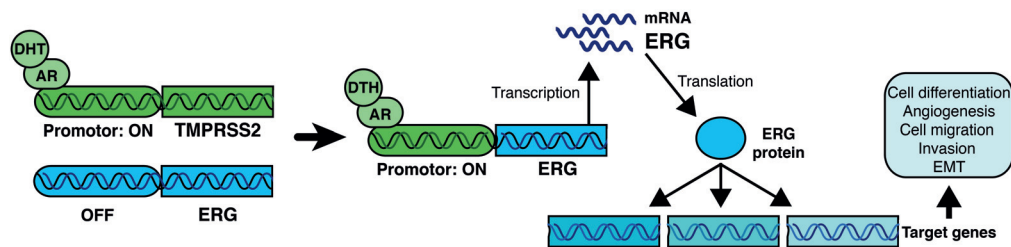
Figure 1.6 TMPRSS2-ERG gene fusion.
The promotor for *TMPRSS2* expression is regulated by androgens, and is normally activated in prostate cells, while *ERG* is not normally expressed. A fusion of the *TMPRSS2* promotor to the coding *ERG* gene, results in active transcription of *ERG*, which increases the synthesis of the ERG protein. ERG can then activate transcription of several genes, and thereby induce processes such as EMT. *AR – androgen receptor, DHT – dihydrotestosterone, EMT – epithelial-mesenchymal transition.*

1.3 Prostate Cancer Metabolism

Reprogramming of metabolism is one of the hallmarks of cancer development [112]. Increased production of energy and building blocks, as well as biochemical homeostasis, are necessary for cancer cell survival and proliferation [113, 114]. Metabolic alteration in cancer was first identified in 1924 by Otto Warburg [115], where he described increased glucose utilisation via aerobic glycolysis in cancer cells (Figure 1.8) [116]. The research field of cancer metabolism is still growing [114], among others because oncogenes and tumour suppressors have been shown to induce metabolic changes in cancer [113, 117]. Additionally, there is growing evidence implicating regulatory mechanism between metabolic reprogramming and cancer epigenetics [118]. Increased knowledge about metabolic alterations in cancer can contribute to better understanding of tumour progression, identification of metabolic biomarkers, as well as provide opportunities for cancer intervention and targeted therapy.

In prostate cancer, alterations of several metabolites and metabolic pathway have been identified [119–121]. Metabolic alterations connected to the TMPRSS2-ERG gene fusion, Wnt pathway and SFRP4 are still largely unexplored. Here, an introduction are given to the metabolism of choline phospholipids, citrate, and polyamines in prostate cancer. It should be noted that other metabolic alterations have been identified in prostate cancer, amongst other in the alanine, lactate and lipid metabolism [121, 122].

Choline Phospholipid Metabolism

Cancer cells need increased synthesis of cell membranes to grow and divide. The phospholipid phosphatidylcholine is the major component of cell membranes and is synthesised from choline by the Kennedy pathway (Figure 1.7). The choline phospholipid metabolism is altered in several cancers [123], including prostate cancer [124, 125]. The need of additional choline is met by increased expression and synthesis of choline transporters in prostate cancer cells [125]. An increase in total choline-containing-compounds, as well as individual increase in free choline, phosphocholine and glycerophosphocholine, are well documented in prostate cancer [124, 126, 127].

Citrate Metabolism

Citrate production and storage is one of the main functions of the prostate gland, where citrate in the prostatic fluid is used as energy by the spermatozoa. Prostate cells achieve net citrate production by truncating the tricarboxylic acid (TCA) cycle. This is facilitated by inhibition
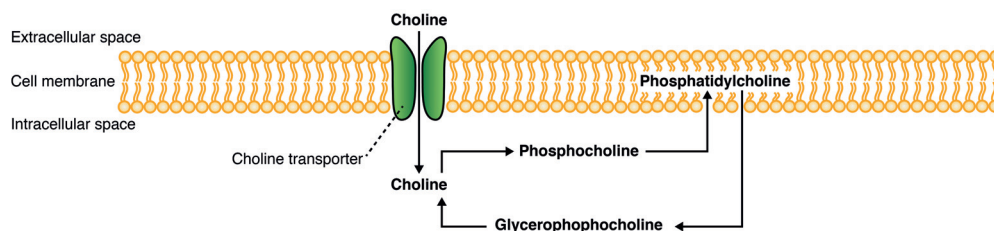
Figure 1.7 The choline phospholipid metabolism.
The choline phospholipid metabolism is important for synthesis of cell membranes. Choline enters the cell through choline transporters, and can be converted to phosphocholine, and further to phosphatidylcholine, i.e. cell membrane. Breakdown of phosphatidylcholine results in glycerophosphocholine. Enzymes are needed to facilitates the pathway.

of aconitase (ACON), the enzyme that converts citrate to isocitrate. This inhibition is caused by accumulation of zinc in prostate cells (Figure 1.8A) [119]. Malignant prostate cells lose their zinc accumulating abilities, resulting in higher ACON activity, and oxidation of citrate in the TCA cycle (Figure 1.8B) [119]. Adenosine triphosphate (ATP)-citrate lyase (ACLY) is the enzyme converting citrate to a precursor of the fatty acid synthesis. The gene expression of *ACLY* is elevated in prostate cancer [128], and found to be anti-correlated with citrate levels [124]. This suggests increased use of citrate in fatty acid synthesis in prostate cancer.

Reduced level of citrate in prostate cancer is well recognised, and has been detected by magnetic resonance spectroscopy both *in vivo* and *ex vivo* [129–131]. Further reduction in citrate concentration has been identified in cancer with high Gleason score [126], and a negative correlation between citrate and PSA level has been detected [132].

Although reprogramming of the citrate metabolism in prostate cancer is well described, it has also been hypothesised that the detection of reduced citrate in tissue is mainly due to the morphology of prostate cancer, with suppression of the luminal space where citrate is stored [133]. For a comprehensive molecular understanding, separating between metabolic reprogramming and changes in morphology is important for understanding of mechanisms.

Polyamine Metabolism

The polyamine metabolism is frequently dysregulated in cancer [134]. Prostate tissue has one of the highest concentrations of polyamines in the body, and their metabolism is therefore of particular interest in prostate cancer. Putrescine, spermidine and spermine are the three polyamines synthesised in mammalian cells, where the enzyme ornithine decarboxylase (ODC1)
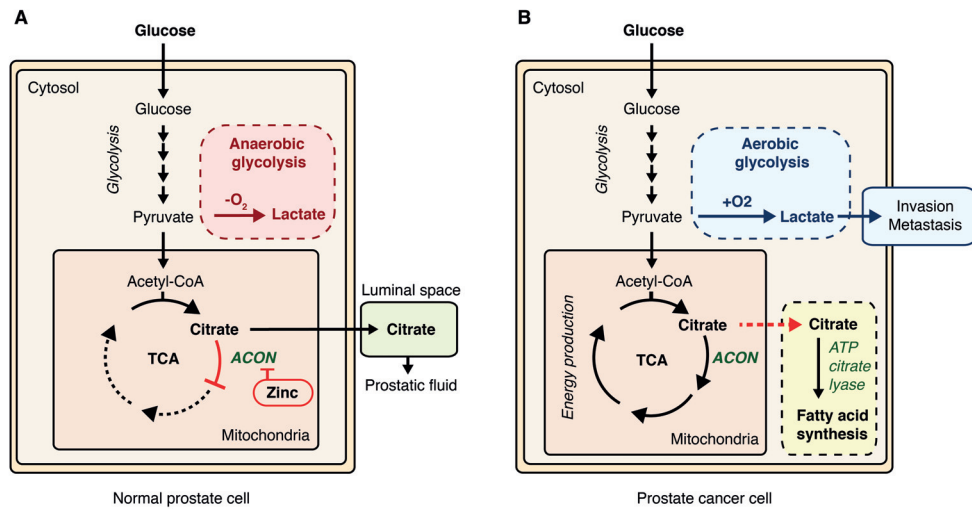
Figure 1.8 Glycolysis and citrate metabolism in (A) normal and (B) cancer cell.
In cancer cell reprogramming of the glucose metabolism towards glycolysis can occur even if oxygen is present. Normal prostate cells produce and secrete citrate, and this is altered in cancer cells where citrate is used for energy production in the tricarboxylic acid (TCA) cycle, and for fatty acid synthesis.

catalyses the first step, by converting ornithine to putrescine (Figure 1.9). ODC1 can be activated by androgens [135, 136], and is classified as an oncogene. Furthermore, the expression of ODC1 has been associated with cell transformation and proliferation [137]. In prostate cancer specimens, an overexpression of ODC1 has been found [138, 139], suggesting increased biosynthesis of putrescine. On the other hand, the last polyamine, spermine, has been identified to inhibit cell proliferation [140], and the expression of the enzyme converting spermine back to spermidine, spermine oxidase (*SMOX*), has been shown upregulated in prostate cancer compared to benign tissue [141]. Spermine/spermidine N1-acetyltransferase 1 (*SAT1*) is, however, the rate limiting enzyme of both spermine and spermidine catalyse (Figure 1.9). Previous magnetic resonance spectroscopy studies of prostate tissue have detected reduced levels of spermine in prostate cancer compared with both benign and BPH prostate tissue [126, 142–144], and a further decrease has been noted in high Gleason score samples [126]. Putrescine has also been found in lower concentration in prostate cancer compared with normal prostate tissue [126].

1.4   *Omics* Sciences

Omics is a collective term for a broad discipline of high throughput research exploring the characteristics and interactions of biological molecules. This includes studies of genes (genome),
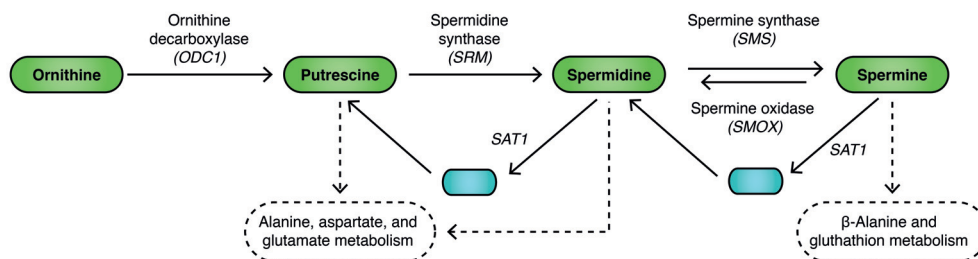
Figure 1.9 The polyamine metabolism.
The polyamines putrescine, spermidine, and spermine are synthesised in the polyamine metabolism. Note that the three polyamines can be converted both forward and backward, and they can be precursors for other metabolic pathways. Different enzymes are needed for catalyse of the different steps. Gene symbols of the enzymes are in parentheses. *SAT1* – Spermine/spermidine N1-acetyltransferase 1.

gene transcripts (transcriptome), proteins (proteome), and small metabolites (metabolome). Figure 1.10 illustrates the general direction of the *omics* cascade, however, interactions may occur between all levels. Integrating different steps in the *omics* cascade makes it possible to observe the intricate relationship between them. There is a range of analytical techniques for each *omics* level, and an introduction to the technologies most relevant for this thesis follows.
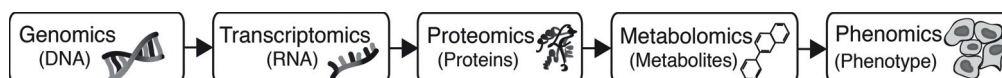
Figure 1.10 Schematics of the *omics* cascade.
The *omics* cascade goes from genomics to phenomics, via transcriptomics, proteomics, and metabolomics, where the latter is closest to the phenotype.

Transcriptomics

Transcriptomics is defined as the study of the complete set of ribonucleic acids (RNA), also called the transcriptome. In 1997, the first paper was published using whole transcriptome analysis [145]. This was followed by rapid advancement of the technology, further accelerated by the successful sequencing of the human genome in 2001 [146, 147]. Today, over 40 000 transcripts can be simultaneously measured using microarray gene chip technology, and RNA sequencing is emerging as more precise method.

Gene expression can be used to improve the understanding of cancer genesis, classification of molecular subtypes, and identification of diagnostic and prognostic biomarkers [148]. One of the most successful stories is found in breast cancer, where five clinically relevant subtypes

have been identified by gene expression analyses [149, 150]. In prostate cancer, differences in gene expression with Gleason score and patient outcome have been described [151], where, among others, the expression disparity of *ERG* led to the discovery of the TMPRSS2-ERG gene fusion [97]. Several gene expression signatures have been established, and shown suitable for molecular subtyping of prostate cancer [104, 105], including three commercial available signatures for prostate cancer aggressiveness (Prolaris [152], OncotypeDx [153] and Decipher [154]). However, the clinical applications of these signatures are still unclear. Further insight and validation of gene expression in prostate cancer are therefore of interest.

DNA Microarray

Deoxyribonucleic acid (DNA) microarray technology evolved from Southern blotting as a high throughput method for quantification of a large number of expressed genes [148]. The core principle of the technique is DNA hybridisation, where two complementary single DNA strands (and DNA/RNA) will anneal together. There exist several variations of the procedure, but the main principles are the same and are shown in Figure 1.11. Before analysis, the RNA is extracted from the tissue samples, and the quality of the RNA is measured, and reported as RNA integrity number (RIN), ranging from 1 (low quality) to 10 (perfect quality) [155]. DNA microarray is relatively inexpensive, however, one of the main drawback is lack of absolute quantification.

RNA-Sequencing

Next-generation DNA sequencing (NGS), is much quicker compared with conventional sequencing, and is expected to transform genomic and transcriptomic research [156, 157]. In general, DNA sequencing is an approach where the order on the DNA nucleotides (thymine, adenine, cytosine and guanine) are determined. RNA-sequencing (RNA-seq) is based on the principles for NGS, where RNA is converted to complementary DNA (cDNA) before the sequencing process. Although more expensive, RNA-seq has several advantages compared with microarray analysis for gene expression, including the possibility for absolute quantification, low background noise, higher sensitivity, and the possibility for detection of new transcripts [157–159].

Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) is a powerful tool for analysis of previously defined microarray gene sets [160]. By looking at the collective expression of several genes, more precise information can be obtained on biological processes involving multiple genes such as activation of pathways. GSEA are frequently used for analyses between samples of two classes, however, an extension of the method allows for single samples GSEA (ssGSEA) [161]. In
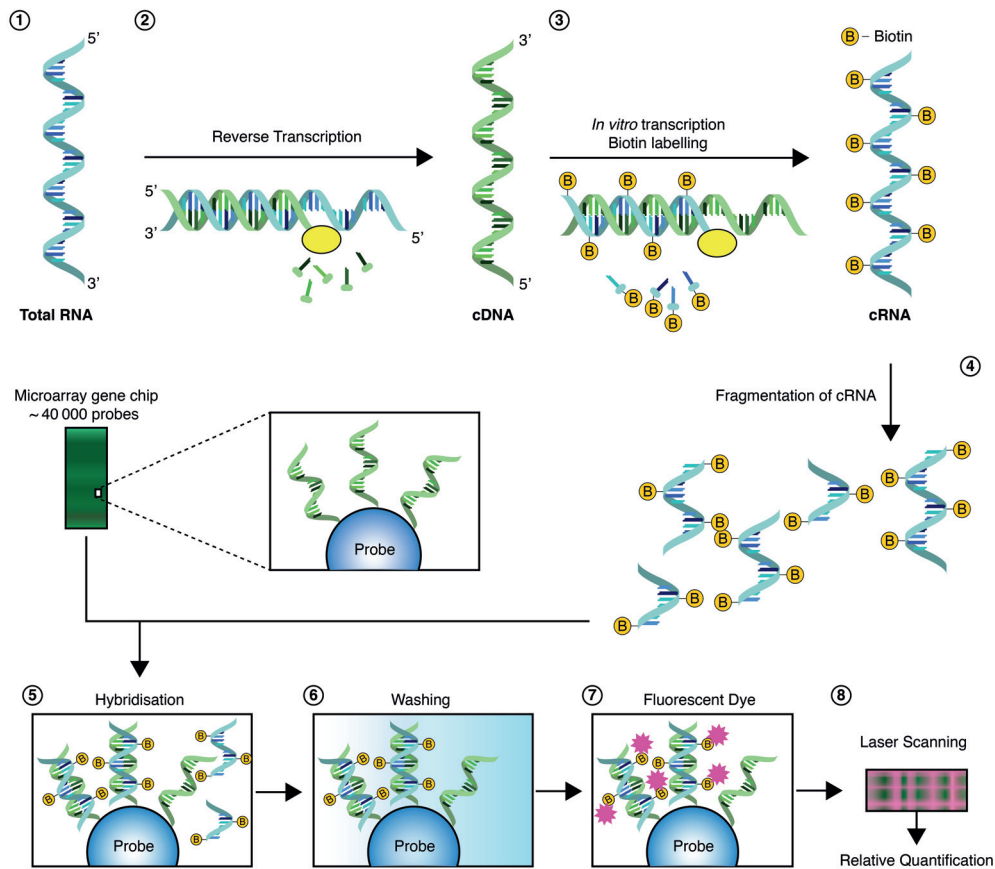
Figure 1.11 Schematic illustration of the steps involved in DNA microarray.
1) Total RNA is extracted from tissue sample. **2**) RNA is copied to complementary DNA (cDNA) by reverse transcription **3**) and back to cRNA labelled with biotin through *in vitro* transcription. **4**) Fragmentation of cRNA before **5**) it is added to the microarray gene chip, containing several thousand specific DNA sequences (probes), where the cRNA will hybridise with the matching DNA sequences. **6**) The non-bonding cRNA is washed away before **7**) the chip is stained with fluorescent molecules sticking to biotin. **8**) The signal intensity is dependent on the amount of cRNA attached to each individual probe, and can be measured by a laser scanner.

ssGSEA, the score of each sample represents the coordinated up- or downregulation of the assigned gene set within one sample, and the score can be compared across the samples.

Immunohistochemistry

The mRNA can be translated into proteins in the cells, and analyses of tissue protein expression can therefore be used investigate and validate if the gene expression is reflected in protein abundance. Immunohistochemistry (IHC) is a common method for visualising proteins in tissue sections, and is frequently used in the clinic as a complement technique to morphologic histopathology [162]. The possibility of immunological staining of antigens in tissue sections was discovered in 1941 [163], and has since been used for characterisation of a wide range of diseases, including in cancer diagnosis, prognosis, and therapeutic decision making [162]. The principle of antigen to antibody interaction is exploited in IHC, where defined antigens of proteins in the tissue sections can be visualised by adding specific antibodies that are marked with a staining agent [164]. Tissue sections stained by IHC can be evaluated in a normal light microscope, and protein abundance can be reflected in staining intensity. The staining can also be identified to cell types, as well as location within the cells, which is an advantage when such distinctions are important, as for example for nuclear translocation of β-catenin during canonical Wnt pathway activation [165]. For prostate cancer, IHC staining is not used in the clinical routine. However, several markers, including the EMT markers E-cadherin and N-cadherin, have shown potential for additional prognostic information for prostate cancer patients [64].

Fluorescence *In Situ* Hybridisation (FISH)

Altered gene expression may be induced by changes in the genome, such as gene fusions. To detect such translocation of specific DNA sequences, fluorescence *in situ* hybridisation (FISH), a cytogenetic technique, can be used. The method was developed in the early 1980s, and, as for DNA microarray, the principle of DNA hybridisation is utilised [166]. FISH can be performed on fixed cells or in tissue, where the DNA in the chromosomes is denatured into single strands before adding fluorescence labelled DNA probes. The probes will hybridise with matching DNA sequences in the tissue, while excess probes are washed away. A fluorescence microscope is used to analyse the existence and the physical location of the sequences of interest [167]. Although gene expression signatures for TMPRSS2-ERG gene fusion have been developed, FISH analysis is considered to be the gold standard for detection [97].

Metabolomics

Metabolites are products or intermediates of the metabolism, and are essential for the cells and tissue as energy, building blocks, and signalling molecules. The production of metabolites is dynamic, and interacts with genetic transcription, proteins, and the environment. Alterations in the concentration of metabolites or in the metabolic fingerprint can be indicative of abnormal processes, such as cancer development and progression [168]. The term metabolomics refers to the study of metabolites in an organism, cell, or tissue. Different analytical technologies for studying metabolomics exist, including magnetic resonance spectroscopy, a reproducible method to gain information about the metabolic situation in tissue samples [169].

The Principles of Magnetic Resonance Spectroscopy (MRS)

The fundaments of nuclear magnetic resonance spectroscopy (MRS) were discovered in the 1940s as a way to determine chemical structures of materials [170–172]. In MRS, the physical properties of nuclei and electrons are exploited to obtain information about chemical composition of the material in question. All nuclei have spin and are precessing around their own axis. Nuclei with an odd number of protons, such as ^{1}H, are accessible for MRS as the movement of their positive proton charge produces a small magnetic field. When placed in an external magnetic field (B_0), the nuclei will align and precess parallel (low energy) or anti-parallel (high energy) to B_0. Most of the nuclei will be in the low energy state, thus there will be formed a net longitudinal magnetisation parallel to B_0. The precession of the nuclei are random and cancel each other out. Thus there is no transversal magnetisation in this state (Figure 1.12A). A radio frequency (RF)-pulse with the same frequency as the precession of the nuclei, will supply the nuclei with energy. This will facilitate a low energy state nuclei to enter the high energy state, decreasing the longitudinal net magnetisation. The energy from the RF-pulse will also synchronise the precession of the nuclei, creating a new precessing transversal magnetisation vector (Figure 1.12B). When the RF pulse stops, the nuclei will relax back to their normal state. The longitudinal magnetisation will then start to increase, and the precessing transversal magnetisation vector decrease. This produce a sum magnetisation vector with a spiralling motion. The moving magnetic field is the free induction decay (FID) signal, and can be registered by a receiver coil (Figure 1.12C). The relaxation times of the longitudinal and transversal magnetisation is dependent on the tissue properties, thus different components (water, lipids etc.) will produce FID signals with slightly different signal intensity (longitudinal) and signal broadening (transversal).

In addition to protons, the electrons will also move in response to the B_0 magnetic field. This creates a local and much smaller magnetic field, which affects the nuclei (electron shielding).

Because of this effect, nuclei in different chemical structures will send out FID signals with slightly different frequencies (chemical shifts).

The FID signal can be Fourier transformed from the time domain to the frequency domain. However, the frequencies are dependent on the B_0 field strength, and the frequency domain is therefore usually converted to the absolute scale of parts per million (ppm). This provides the MR spectrum with peaks along the ppm axis, making it possible to identify and quantify different metabolites (Figure 1.13).
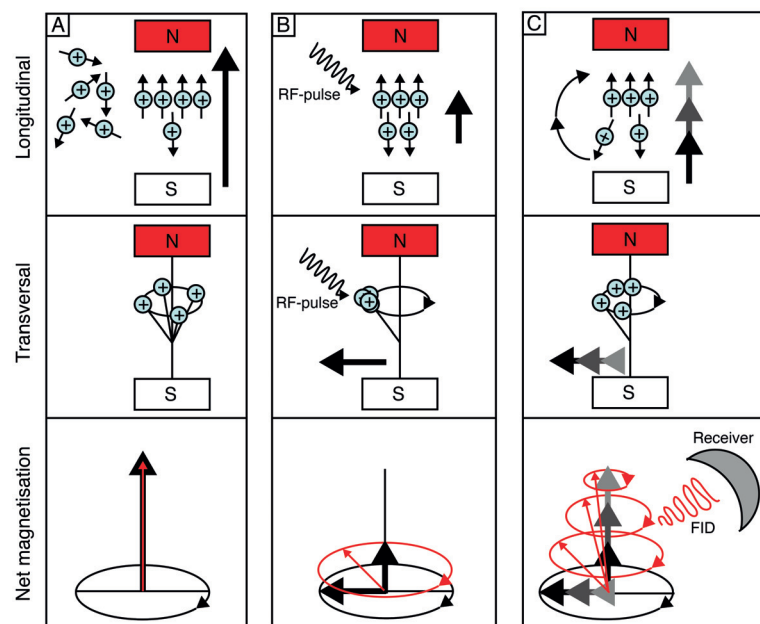
Figure 1.12 Principles of magnetic resonance.
(A) Charged nuclei will align with an external magnetic field, resulting in a net longitudinal magnetisation (B) A radio frequency (RF)-pulse that resonates with the nuclei, results in reduced longitudinal magnetisation and a spinning transversal magnetisation. (C) When the RF-pulse is turned off, relaxation of the nuclei cause a spiralling net magnetisation, detectable as a free induction decay (FID) signal.

High-Resolution Magic Angle Spinning Magnetic Resonance Spectroscopy (HR-MAS MRS)

Detection of metabolites in tissue can be achieved *ex vivo* by high-resolution magic angle spinning MRS (HR-MAS MRS). Tissue is considered a semisolid material with reduced molecular mobility. This induces large dipole-dipole interactions and chemical shift anisotropy, causing line broadening of the spectra, thus hiding metabolic information. The line broadening can be

reduced by rapidly spinning the samples at 54.7 degrees to the static magnetic field (Figure 1.13). This is called the magic angle, and was discovered by Andrew et al. [173] and Lowe [174] in the late 1950s. HR-MAS MRS was first used to study tissue specimens in 1996 [175], and has since been widely used in metabolomics studies of cancer [176], including prostate cancer [122, 126, 127, 131]. The method is non-destructive, enabling integration of the metabolic information with results from subsequent analyses, such as gene expression, histopathology, and immunohistochemistry. Other advantages include simple and highly standardised sample preparation, as well as the possibility for absolute quantification of the metabolites [177–179]. The main drawback compared to other methods for metabolomics, such as mass spectroscopy, is the relatively low sensitivity of HR-MAS MRS, where metabolites needs to be in millimolar concentrations for detection, compared to picomolar for mass spectroscopy.
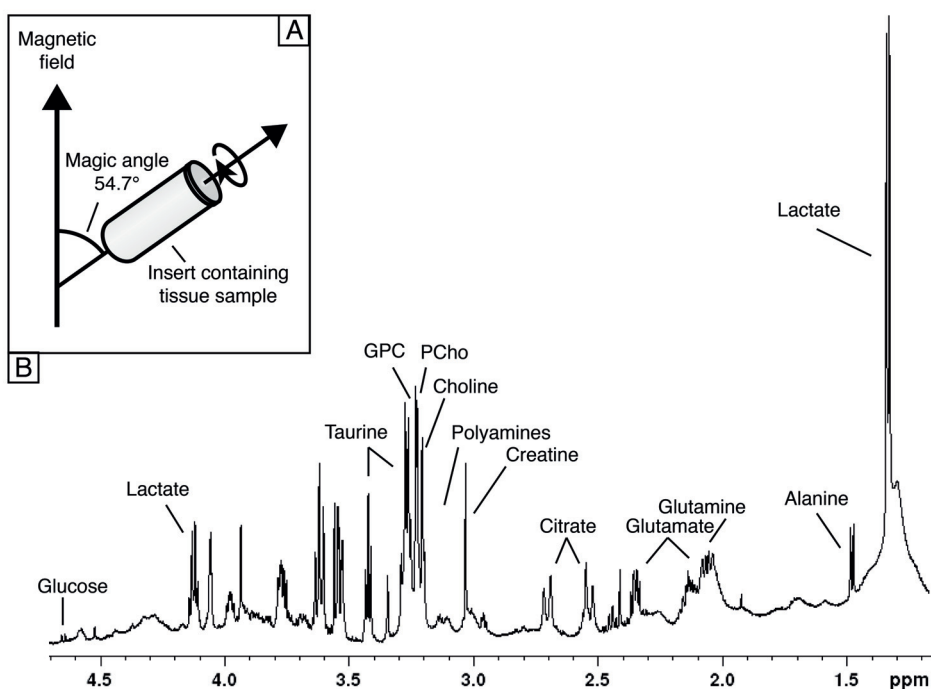
Figure 1.13 HR-MAS MRS of prostate tissue.
(A) For HR-MAS MRS, the tissue samples are tilted to the magic angle and are spun around their own axis. These techniques reduce the line broadening and increase the resolution of the spectrum. (B) Chemical shift spectrum of prostate tissue. The main metabolites are assigned to their peaks as an overview. *ppm – parts per million, GPC – Glycerophosphocholine, PCho – Phosphocholine.*

Magnetic Resonance Spectroscopic Imaging (MRSI)

Prostate metabolism can also be investigated using traditional MRI scanner by *in vivo* MRS imaging (MRSI). Brown et al. introduced MRSI in 1982 [180], which led to a rapid development in both equipment and acquisition methods [181]. Today, MRSI of the prostate can detect the metabolites citrate, choline, creatine and polyamines/spermine on a standard 3 tesla MRI scanner, and has shown potential for detection, localisation, and assessment of prostate cancer aggressiveness [182]. However, prostate MRSI is not frequently used in the clinic today, which may be due to the additional technical aspects as well as time required for the examination ($\sim$10 minutes), and the lack of a good system for interpretation combined with unclear clinical benefits.

Quantification - Linear Combination of Model Spectra (LCModel)

The area under the peaks of the MR spectra is proportional to the number of nuclei in the metabolites creating the peaks. Several methods for MRS quantification has been developed [176], including Linear Combination of Model spectra (LCModel) [183]. For this method, the frequency-domain data is fitted to model spectra from individual metabolites, and a semi-parametric algorithm is used to calculate the metabolite concentrations [183, 184]. The LCModel method was originally designed for *in vivo* MRSI, but has been successfully adapted for *ex vivo* MR spectra [126, 185].

1.5 Statistical Analyses

Data Transformation

Continuous biomedical variables often do not fulfil the normality assumption of many statistical tests. Although several non-parametric analyses exist for skewed, non-normal distributed data, more statistical power can be achieved by transforming the data to be closer to normal distribution. Logarithmic and square-root transformation are common methods for positive variables [186]. A small constant is frequently added to the logarithmic function to avoid the problem of values close or equal to zero.

Linear Mixed Model

Linear mixed model is a statistical analysis particularly useful for repetitive measurements, missing data, or datasets with several measurements per subject. The model is an extension of

the linear regression model, but uses both fixed and random effects (mixed effects) [187, 188]. Linear mixed models can describe the relationship between a dependent variable, e.g. metabolite concentration or gene expression, and explanatory variable(s), e.g. sample classes such as normal and cancer samples (fixed-effects). Random effects are not of primary interest, but may be important to account for, such as multiple samples per patient.

Multivariate Analyses

Large number of variables as observed in MR spectra and gene expression, combined with a relatively low number of samples, is challenging with traditional statistical analysis. Multivariate analysis is a specialised approach to investigate such data, and can be used for data reduction, identification of biomarkers, and for discrimination between groups.

Principal Component Analysis (PCA)

Principal component analysis (PCA) is an unsupervised (no added knowledge) multivariate analysis, where the dimensionality of the variance in the data is reduced by linear transition into principal components [189]. The first principal component explains most of the variance in the data set, whereas the second maximise the remaining variance, and so on. These principal components are uncorrelated to each other. PCA can reveal hidden structures of the data, and is frequently visualised by score plots, where the principal components form the axis.

Partial Least Squares Discriminant Analysis (PLS-DA)

Partial least squares discriminant analysis (PLS-DA) is a similar method to PCA, but additionally exploits known information in the response variable(s) (supervised). In this way, the relationship between a response variable (such as sample class or patient information) and the experimental data (such as MRS spectra) can be investigated. PLS-DA analysis results in uncorrelated latent variables [190, 191], and similar to PCA, PLS-DA is often visualised by score plots, where the latent variables form the axis.

Survival Analyses

Time to an event, such as cancer-specific death or surrogate endpoints including biochemical recurrence and distant metastasis for prostate cancer, can be studied by survival analyses. Censoring in survival analyses makes it possible to deal with variation in patient follow-up time [192].

Kaplan-Meier Estimator and Log-Rank Test

In 1958, Kaplan and Meier described a non-parametric method to estimate the survival distribution of time-to-event data, the Kaplan-Meier Estimator [193]. The survival distribution is frequently presented as a Kaplan-Meier plot, where the survival curves can be visualised and compared between groups. The log-rank test is commonly used to test the hypothesis of differences in survival distribution between groups of patients [194], however, the log-rank test does not give information about the effect size between survival of the groups.

Cox Proportional Hazard Model

Cox Proportional Hazard (PH) model is a regression based analysis of time-to-event data [195]. The method bears similarities to logistic regression model, but with the added time variable and censoring. Cox PH model estimates the hazard ratio (HR) as an effect size. If HR is above or below 1, the hazard is increased or decreased, respectively. As an example, in a Cox analysis comparing two groups, a HR of 2 would indicate twice the rate of an event per unit of time in the reference group. Instead of groups, the covariate in a Cox PH model may be continuous, such as gene expression and metabolite concentration, the HR will then reflect each unit increase of the covariate, and is thus scaled to the range of the covariate.

One covariate can be investigated on its own by univariable Cox PH model. However, more accurate evaluation of the usefulness of each prognostic factor can be made when other known or likely prognostic factors are controlled for. This can be done by multivariable Cox PH models, where more covariates are modelled together.

Meta-Analysis

Meta-analysis is a statistical method integrating the results from several studies, to obtain a pooled result with higher statistical power [192, 196]. The method can also be used to combine results from several cohorts within a study, where the raw data cannot be directly combined into one analysis. This is the case for microarray based gene expression data, where normalisation is performed within each cohort, and absolute quantification of the gene expression cannot be obtained.

The effect size in a statistical analysis, for instance the mean difference in two groups, is dependent on the underlying population of the study. To combine studies in a meta-analysis, the effect size in each study needs to be standardised. Cohen's *d* is a commonly used standardised

effect size, and is obtained by presenting the group mean difference (effect size) in units of standard deviation [197]. Cohen's *d* effect size can be classified as very small (0.01), small (0.20), medium (0.50), large (0.80), very large (1.20), and huge (2.0) [197, 198]. Hazard ratios for gene expression can also be converted to a standardised effect size by multiplying the natural logarithm of the hazard ratio (β) with its standard deviation [199]. The standardised hazard ratio can be interpreted as the unstandardised hazard ratio, with increased and decreased hazard when values are above and below 1.0, respectively.

To estimate the overall effect in meta-analysis, precise studies (narrow confidence intervals) are weighted more than less precise studies (wide confidence intervals). The study weight can be calculated by two different models, *fixed effects* and *random effects*. The *fixed effects* model assumes that the effect size should be equal in all studies. This is often unreasonable, as the underlying populations of the studies frequently differ. Thus, the *random effects* model, taking additional variation into account, is usually more appropriate. However, using a *random effects* model will give larger confidence intervals of the overall effect. Test of heterogeneity or the dissimilarity in the effect size of all included studies, can guide selection between *fixed effects* and *random effects* model, where the latter is advocated if the heterogeneity test is significant. The results of a meta-analysis are commonly graphically displayed as a forest plot (Figure 1.14) [192, 196].
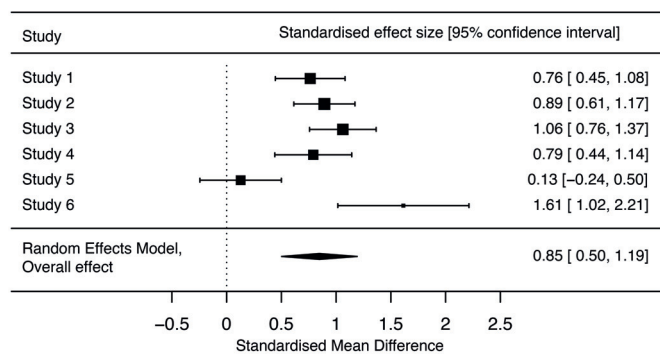
Figure 1.14 Example of a meta-analysis forest plot.
A forest plot visualises the results for a meta-analysis. For each study, the standardised effect with 95% confidence intervals are plotted, as well as the combined effect. Notice that the box sizes in the forest plot reflect the weight of each study in the meta-analysis.

Correction for Multiple Testing

One of the challenges with high throughput technologies, is the vast number of measured variables (i.e. expressed genes or spectral points), and many hypotheses can be investigated in explorative studies. Multiple hypothesis testing with relative small sample sizes increases the probability of false discoveries. To identify the true relationships and reject false discoveries, several methods for correction of the p-values have been developed, including the Benjamini-Hochberg procedure [200]. In this method, an adjusted p-value is calculated for each test, dependent on the total number of tests. The adjusted p-values are obtained by ranking the unadjusted p-values from 1 to n (number of tests). Each p-value is multiplied by n and divided by its assigned rank. The false discovery rate is often set to 0.05, where the adjusted p-values below this limit are recognised as true discoveries.

2. Objectives

Overall Aim

The overall aim of this thesis was to obtain high resolution multi-level molecular information to identify candidate biomarkers and signatures for improved risk stratification of prostate cancer patients.

Specific Aims

To integrate omics technology (transcriptomics and metabolomics) with histopathology and immunohistochemistry to study molecular prostate cancer pathways in human tissue biobank material.

To validate specific and promising biomarkers and signatures in publicly available human prostate cancer cohorts and investigate their association with aggressiveness and recurrent disease.

To validate the possible metabolic biomarkers in a small *in vivo* MRSI patient cohort, to verify *ex vivo* metabolomics results.

Specific objectives for each paper

Paper I

Identify metabolic alterations associated with the presence of TMPRSS2-ERG gene fusion in prostate cancer tissue, and investigate its association with aggressiveness and recurrent disease.

Paper II

Identify and validate Wnt signalling and epithelial-mesenchymal transition (EMT) in prostate cancer tissue, and investigate its association with metabolic reprogramming, aggressive and recurrent disease.

Paper III

Identify and validate the association between *SFRP4* gene expression and aggressive and recurrent prostate cancer.

3. Materials and Methods

In this thesis, tissue samples from prostate cancer patients have been analysed by several methods, and patient follow-up data have been collected. An overview of cohorts, methods, and analyses performed for each paper is given in Table 3.1.

Table 3.1 Overview of cohorts and methods used in the papers included in this thesis.

		Paper I	Paper II	Paper III
Cohorts	Main cohort (N=41, n=129)	✓	✓	✓
	IHC cohort (N=40, n=40)	✓	✓	✓
	Validation cohorts	–	5 cohorts (n=1519)	8 cohorts (n=2001)
Methods	Transcriptomics	Microarray	Microarray	Microarray
	Metabolomics	HR-MAS MRS MRSI	HR-MAS MRS MRSI	HR-MAS MRS
	Other	Histopathology FISH	Histopathology Immunohistochemistry	Histopathology Immunohistochemistry
Data and statistical analyses	Metabolite	LCModel Multivariate analyses	LCModel	LCModel
	Gene expression	ssGSEA INMEX	Differential expression ssGSEA	Differential expression Log-fold change
	General statistics	T-test Pearson correlation LMM	T-test Spearmans rho LMM Chi-squared	T-test Pearsons correlation Fisher exact test
	Multivariate	PCA/ PLS-DA (metabolomics)	PCA (transcriptomics)	–
	Survival analyses	Kaplan-Meier plot Log rank test	Kaplan-Meier plot Log-rank test Cox PH model	Cox PH model
	Correction for multiple testing	Benjamini-Hochberg false discovery rate	Benjamini-Hochberg false discovery rate	–
	Meta-analysis	–	–	✓

3.1 Ethics Statement

The studies included in this thesis were approved by the Regional Committee of Medical and Health Research Ethics (REC), Central Norway (case numbers: 010-04, 2009/1161 (4.2007.1654), and 4.2007.1890). All included patients gave an informed written consent.

3.2 Materials

Patients

The three papers included in this thesis used tissue samples from patients diagnosed with prostate cancer and treated by radical prostatectomy at St. Olav's Hospital, Trondheim, between March 2007 and February 2010. The patients made up two separate cohorts, the *main cohort* and the *immunohistochemistry (IHC) cohort* (termed "*validation cohort*" in paper I). The inclusion of patient and samples is presented in Figure 3.1, and characteristics of the patients are given in Table 3.2.
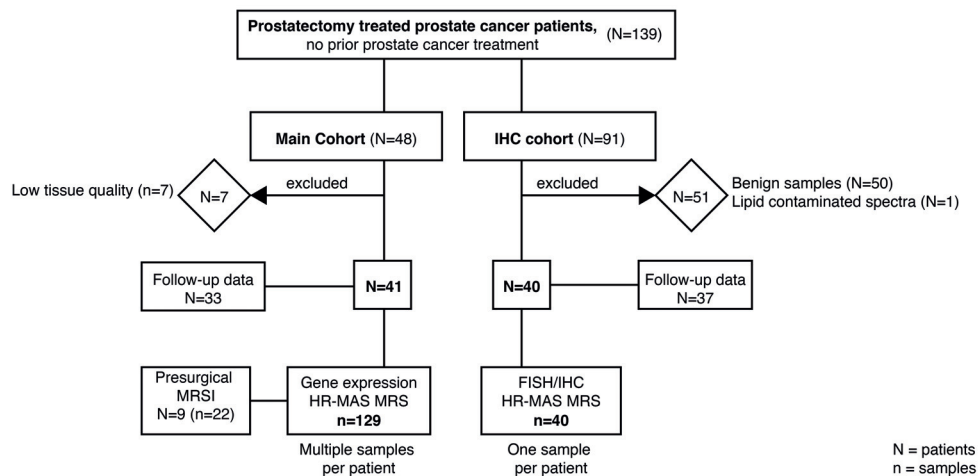
Figure 3.1 Patient and sample inclusion diagram.
Tissue samples from prostate cancer patients were included in two independent cohorts, the *main cohort* and the *IHC cohort*. Patients were excluded due to low tissue or RNA quality in the *main cohort*, this cohort also included multiple samples per patients. In the *IHC cohort* benign samples were excluded, as well as one sample with lipid contaminated HR-MAS MRS spectra. The *IHC cohort* only included one sample per patient.

Table 3.2 Characteristics of the included patients.

Patients		**Main cohort** (n=41)	**IHC cohort** (n=40)
Age at diagnosis *(median, range)*	Years	64 (48-69)	62 (48-73)
PreOp PSA *(median, range)*	(ng/mL)	9.1 (4.0-45.8)	8.9 (5.2-18.0)
Pathological T stage *(patients)*	T2	28	27
	T3	13	12
	Unknown	–	1

PreOp PSA – Preoperative measurement of serum PSA.

Follow-up

At least five-year follow-up data were collected in both cohorts, including date of last negative serum PSA measurement, date of biochemical recurrence (serum PSA of at least 0.2 ng/mL), prostate specific death, as well as information about prostate cancer specific treatment. In paper I, II, and III follow-up data were successfully obtained for 33 of the 41 patients in the *main cohort*. Whereas follow-up data were available for 37 of the 40 patients in the *IHC cohort* for paper III.

Validation Cohorts

For validation of the results in paper II and III, additional cohorts were downloaded from open, online databases. These *validation cohorts* included gene expression data, histopathology, and patient follow-up information. All samples were from radical prostatectomy specimens, except for the Sboner et al. cohort, which was from incidental discovered prostate cancer by transurethral resection of the prostate (TURP). More information about the *validation cohorts* is given in Table 3.3.

Tissue Sample Harvesting

The tissue samples from the prostatectomy specimens in the *main* and *IHC cohort* included in this thesis, were harvested by two different approaches.

Fresh Frozen Slices (*main cohort*)

The tissue samples of the *main cohort* were initially harvested for a study by Bertilsson et al. [124]. Whole mount clinical histopathological sections above and below a fresh frozen prostatectomy specimen tissue slice, were used to identify locations for sample collection. Normal, non-cancer, samples were harvested as far away from the cancerous area as possible (Figure 3.2A). Only tissue slices with more than 5% cancer, and cancer in both the adjacent whole mount histopathological sections were included. Multiple samples were collected from

Table 3.3 Overview of the validation cohorts used in paper II and paper III.

Cohort	Access number	Gene expression	Cancer samples	Normal samples	Follow-up endpoint	Paper II	Paper III	Reference
TCGA-PRAD	TCGA PRAD	RNA Sequencing	RP n=497	Same patients n=52	BCR	✓	✓	[201]
CAM Ross-Adams et al.	GSE70768	Microarray Illumina HT12v4	RP n=112	Same patients n=74	BCR	–	✓	[202]
STK Ross-Adams et al.	GSE70769	Microarray Illumina HT12v4	RP n=94	–	BCR	–	✓	[202]
Wang et al.	GSE8218	Microarray Affymetrix U133A	RP n=65	Same patients n=67 Autopsy n=4	BCR	✓	✓	[203–205]
Sboner er al.	GSE16560	Microarray Illumina DASL Assay	TURP n=281	–	Prostate cancer-specific death	✓	✓	[206]
Taylor et al.	GSE21034	Microarray Affymetrix Human Exon 1.0 ST	RP n=131	Same patients n=29	BCR	✓	✓	[207]
Mortensen et al.	GSE46602	Microarray Affymetrix U133 Plus 2.0	RP n=36	Bladder cancer patients n=14	BCR	–	✓	[94]
Erho et al.	GSE46691	Microarray Affymetrix Human Exon 1.0 ST	RP n=545	–	Metastatic progression	✓	✓	[154, 208]

RP- Radical prostatectomy, n – number of samples, TURP – Transurethral resection of the prostate, BCR – biochemical recurrence (PSA $\geq$ 0.02 ng/ml)

each slice (median 3, range 1-6 samples per slice).

The fresh frozen prostate slices that were used for tissue samples collection in the *main cohort* are routinely collected from consenting prostatectomy patients at St. Olavs Hospital, Trondheim University Hospital, by the regional biobank of Central Norway, BioBank1. The method is highly standardised, and, as described by Bertilsson et al., a 2 mm thick slice is obtained from the middle of the gland using a double bladed knife, while the prostate is stabilised in a plastic holder (Figure 3.2B) [177]. The slice is snap frozen between two pre-cooled aluminium plates embedded in liquid nitrogen, and further stored in a mechanical freezer at -80 °C (Figure 3.2B). The average freezing time from surgical removal by this method is previously reported to be 15±4 minutes [177]. Bertilsson et al. also described a method for collecting smaller tissue samples (tissue cores) from the slices. The technique is design to reduce thawing of the tissue, where the tissue slice is placed on a cooled aluminium plate in direct contact with liquid nitrogen, and samples are harvested by a modified drilling device (Figure 3.2C) [177]. This method was used for harvesting the tissue samples in the *main cohort*.

Fresh Frozen Needle Biopsies (*IHC cohort*)

The *IHC cohort* consisted of needle biopsies collected within ~2 minutes after surgical removal of the prostate gland, from consenting patients. Two biopsies were taken from each prostate specimen, and were immediately frozen in liquid nitrogen (-196°C), and further stored in cryotanks (-196°C) in a local biobank administered by the MR Cancer Group, NTNU. Although two samples were collected for each patient, only one sample per patient was included in this thesis. To increase the likelihood of cancer tissue, the samples were chosen according to the following inclusion criteria: The needle biopsy was taken from the area of previously positive TRUS biopsies, where the histopathological reported cancer area in the TRUS biopsy was at least 1 mm.

3.3   Histopathology

Preparation, Sectioning, and Staining

In both cohorts, the tissue composition of the samples was evaluated by histopathology. Cryosectioning (-20°C) was used for the tissue cores in the *main cohort*, and was performed prior to HR-MAS MRS analysis. To prevent contamination, the samples were attached to the microtome
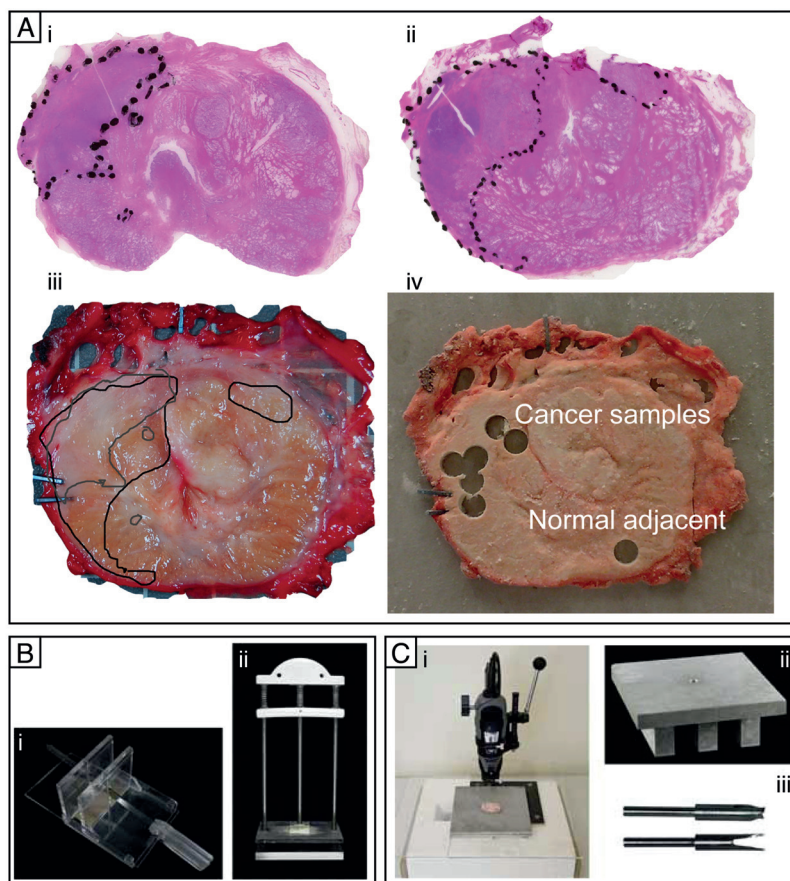
Figure 3.2 Tissue harvesting for the *main cohort*.
(A) Cancer regions from whole mount histopathology section directly below (A_i) and above (A_{ii}) the slice, are transferred to a digitised photo of the slice taken before freezing (A_{iii}), and used for selection of area for smaller sample collection. (A_{iv}) The frozen tissue slice after extraction of samples. (B_i) For harvesting of the fresh tissue slice, the prostate gland is stabilised in a plastic holder, and the slice is obtained using a double bladed knife. (B_{ii}) The slice is immediately frozen in liquid nitrogen, while it is placed between two pre-cooled aluminium plates. (C_i) The workstation for harvesting smaller tissue samples from the slices, (C_{ii}) equipped with liquid nitrogen cooled aluminium plates, (B_{iii}) and a modified drill with a 3 mm bore. *Figure A is reproduced from Bertilsson [209] with permission. Figure B and C are adapted from Bertilsson et al. [177], with permission from John Wiley and Sons.*

by saline water only, and a 4 μm slice was sectioned from each sample. The sections were stained with Haematoxylin and Eosin (HE) (Figure 3.3A).

The tissue samples of the *IHC cohort* were formalin fixed and paraffin embedded after HR-MAS MRS analysis. Due to the rapid spinning of the samples during the MRS acquisition (see section 3.4), the samples will naturally curl up in formalin, making it hard to get representative sections for histopathology. This issue was avoided by uncurling and stretching out the biopsies, before attaching them to lab sheets and cork plates by staples before formalin fixation. A total of ten ~4 μm paraffin sections were initially cut from each biopsy, and the first and the last sections were stained with Haematoxylin Eosin Saffron (HES) (Figure 3.3B). The other tissue sections were used for FISH and immunohistochemistry staining in paper I and paper II, respectively (Section 3.6). Additional tissue sections were later obtained from the same paraffin blocks to accommodate immunohistochemistry staining for paper III.

Evaluation and Scoring

The histopathological sections of both cohorts were evaluated by the same uropathologist (St. Olav's Hospital, Trondheim University Hospital). Percentages of benign epithelium, stroma, and cancer were reported for each sample, and cancer samples were scored according to the clinical Gleason system (see Section 1.1) [41]. The distribution of Gleason score of the samples is given in Table 3.4. The samples were further divided into two groups, *low Gleason* and *high Gleason*, where the *low Gleason* samples had a Gleason score $\leq$ 3+4, and the *high Gleason* samples had a Gleason score $\geq$ 4+3. For the *IHC cohort*, the cancer regions of each biopsy were outlined on digitised photos of the sections (Figure 3.3B), and this was later used for assisting FISH and immunohistochemistry evaluation (Section 3.6).
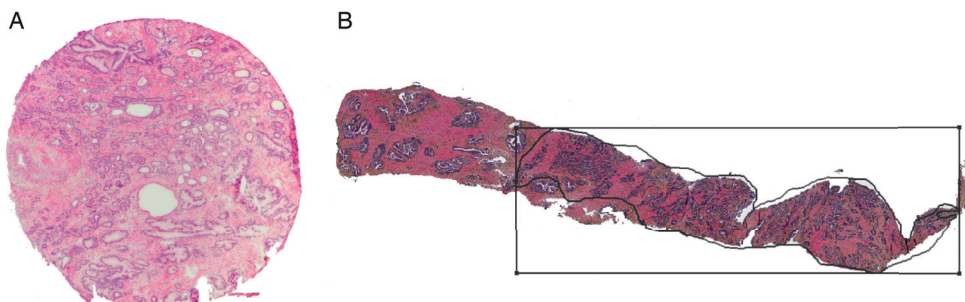
Figure 3.3 Sections for histopathological evaluation.
(A) Cryosection from the *main cohort* stained with HE. (B) Paraffin section from the *IHC cohort* stained with HES. The cancer area in each biopsy samples was outlined by a pathologist.

Table 3.4 Characteristics of the samples in the cohorts.

Tissue samples		**Main cohort** (n=129)	**IHC cohort** (n=40)
Samples weight *(mean, range)*	mg	12.7 (3.0–21.9)	12.6 (7.6–21.0)
Cancer samples	*n*	95	40
Normal samples	*n*	34	–*
Gleason score	*n*		
	3+3	24 (25%)	5 (12.5%)
Low Gleason ↑	3+4	21 (22%)	16 (40%)
High Gleason ↓	4+3	20 (21%)	9 (23%)
	4+4	15 (15%)	5 (13%)
	4+5/5+4	15 (15%)	4 (10%)
	5+5	–	1 (2%)
Luminal space[a] *(median, range)*	*percent*	6.2 (0.0–31.6)	3.4(0.0–14.3)

*Benign samples of the *IHC cohort* were not used in this thesis.

[a]Luminal space were only measured in cancer samples.

Reproducibility

The reproducibility of the histopathological evaluation was assessed in the *main cohort*. All the tissue sections were independently evaluated by an additional experienced uropathologist (University Hospital of North Norway). The pathologist was blinded of the results from the previous histopathological evaluation. For the two pathologists, the overall kappa coefficient for interobserver agreement between normal, *low Gleason* ($\leq$3+4) and *high Gleason* ($\geq$4+3) tissue sections was 0.66, indicating substantial agreement. The first reading was used for the studies, so that the scoring of the *main cohort* and *IHC cohort* were performed by the same pathologist. In addition, there were signs of degradation of the cryosection staining before the second evaluation, making this reading more uncertain.

Luminal Space Measurement

To measure the proportion of luminal space, the HE and HES stained sections were first digitised with 40x magnification, using a camera equipped microscope (Olympus BX41 and DP26, Japan). The proportion of luminal space in each sample was identified by the positive pixel count algorithm (ImageScope v8.0, Aperio Technologies), a colour-based segmentation method [210]. Tissue pixels were identified based on colour, using a hue setting of 0.7 and window setting of 0.39. Not identified pixels were considered to be luminal space, and the percentage was calculated as the proportion of total pixels (tissue and luminal space). The fraction of luminal space in the cohorts is given in Table 3.4.

3.4 Magnetic Resonance Spectroscopy (MRS)

High Resolution Magic Angle Spinning (HR-MAS) MRS

Sample preparation

Sample preparation was performed on an in-house designed workstation, cooled by liquid nitrogen to minimise tissue thawing and degradation. A deuterium oxide (D_2O) based solution (3 μl) was added to leak-proof, disposable HR-MAS inserts (30 μl, Bruker Biospin, Germany), and pipetting errors were checked by weight measurements. For the *main cohort*, the added solution was phosphate-buffered saline, containing trimethylsilyl 3-propionic acid sodium salt (5mM) and sodium formate (25mM). Due to changes and standardisation of the lab protocols, the solution used for the *IHC cohort* only contained sodium formate (25mM). To fit into the inserts, the samples in the *main cohort* were sectioned using a sterile 2 mm biopsy punch, and to remove remnants of blood or lipids in the *IHC cohort*, the edges of the biopsy samples were excised by a sterile scalpel. The tissue samples were placed in the inserts, and the sample weight was registered (Table 3.4). Finally, the inserts were placed into 4 mm zirconium rotors with spinning caps (Bruker Biospin, Germany). This standardised method of sample preparation has previously been described in further details by Giskeødegård et al. [179].

Spectral acquisition

Spectral acquisition was performed using a Bruker Avance DRX600 (14.1 T) spectrometer (Bruker BioSpin, Germany), equipped with a ^{1}H/^{13}C MAS probe. To minimise tissue degradation, the probe temperature was fixed at ~5°C. Proton (^{1}H) spectra were acquired as described in (Table 3.5). The spectra were Fourier transformed with 0.30 Hz line broadening, chemical shifts were referenced to the left peak of the lactate doublet at 1.336 ppm, and a linear baseline correction was applied (Topspin 3.1, Bruker Biospin, Germany). After acquisition, the tissue samples were immediately refrozen and later prepared for gene expression or histopathology analysis in the *main* and *IHC cohort*, respectively.

Magnetic Resonance Spectroscopic Imaging (MRSI)

Acquisition

As a part of a different study [211], nine of the patients in the *main cohort* had a MRSI acquisition included in their pre-surgical MRI examination, and this data were included in paper I and II. The MRSI was performed on a 3 T system (MAGNETOM Trio, Siemens Medical Solutions, Germany), with a 6-channel phased array body coil (Body Matrix coil, Siemens). Saturation

Table 3.5 HR-MAS MRS parameters.

Pulse sequence	Main cohort		IHC cohort	
	Single pulse	CPMG Spin-echo	1D NOESY	CPMG Spin-echo
Bruker ID	ereticpr.drx	cpmgpr	noesygppr	cpmgpr
Temperature	4 °C	4 °C	5°C	5°C
Spin rate	5kHz	5kHz	5kHz	5kHz
Acquisition time	3.28 s	3.28 s	2.74 s	3.07 s
Number of scans	128	128	128	256
Paper	I, II & III	I	I , II& III	I

NOESY - Nuclear Overhauser effect spectroscopy, CPMG - Carr-Purcell-Meiboom-Gill.

slabs were positioned around the prostate to saturate periprostatic lipid signals, and manual shimming was performed. A ^{1}H MRSI point-resolved spatially localised spectroscopy (PRESS) sequence optimised for the prostate was used [212], with a nominal voxel size of 7.1x7.1x7.5 mm^3.

Matching of HR-MAS MRS and MRSI

To analyse the MRSI and gene expression data together, the MRSI voxels were matched with the equivalent tissue samples. To identify the best corresponding MRSI slice, anatomical landmarks of the MRI images were compared with the whole mount HE stained sections below and above the fresh frozen tissue slice. The location of the small tissue cores harvested for HR-MAS MRS and gene expression, were transferred to a digitised photo of the fresh tissue slice, and matching MRSI voxels were identified by transparent overlay of the images in Photoshop (Adobe Photoshop Elements 4.0). The matching of tissue samples used in this thesis was initially performed to compare MRSI and HR-MAS MRS in a study by Selnæs et al. [211].

Metabolite Quantification

LCModel was used to quantify both the HR-MAS MRS and MRSI spectra [183]. For the *main cohort*, a 23 metabolites basis set was simulated NMRSIM (Bruker BioSpin, Germany), and used to quantify the pulse-acquired HR-MAS MRS, as previously described by Giskeødegård [126]. The spectra in the *IHC cohort* were quantified by a similar procedure, using the NOESY spectra, and a further optimised basis set of 24 metabolites. In both cohorts, the known concentration of the added formate was used to achieve absolute quantification of the metabolites, which were reported in mmol/kg wet weight.

For the MRSI spectra, a basis set of four metabolites, citrate, choline, creatine, and spermine, was

simulated by NMRSIM. As there are no metabolites of known concentration in the MRSI spectra, only relative concentration could be quantified by LCModel. Creatine was considered stable, and metabolites to creatine ratios were used for the analyses in this thesis. The quantification of the MRSI spectra is previously described by Selnæs et al. [211].

3.5   Gene Expression

Gene Expression Measurement

In the *main cohort*, gene expression analysis was performed on the exact same tissue samples after HR-MAS MRS. The tissue was homogenised with tissue lysis buffer for 10-20 seconds, before manual extraction of RNA by using the mirVana™miRNA Isolation Kit (Ambion Inc.). The concentration and purity of RNA were measured by a spectrophotometer (NanoDrop Technologies, USA), and the integrity of the RNA (RIN score) was analysed with the 2100 Bioanalyzer (Agilent Technologies, USA). Illumina TotalPrep RNA Amplification Kit (Ambion Inc.) was used for RNA amplification before the microarray analysis.

Gene expression DNA microarray analysis was performed using an Illumina Human HT-12v4 Expression Bead Chip (Illumina), which provides a genome-wide expression analysis, containing more than 47,000 probes. To adjust for technical artefacts, the transcript values were filtered, $\log_2$ transformed and quantile normalised. The microarray service was provided by the Genomics Core facility – NTNU, and the Norwegian Genomics Consortium, and was originally obtained for a study by Bertilsson et al., where they investigated gene expression alterations associated metabolic reprogramming of citrate and choline in prostate cancer [124]. The microarray data has been published in an open database, Array Expression, with access number: E-MTAB-1041.

Gene Set Enrichment Analysis (GSEA)

The expression of specific sets of genes, called gene signatures, were analysed in paper I and II. To give each sample a score reflecting the enrichment of genes in the signatures, single samples gene set enrichment analysis (ssGSEA) were used. Briefly, all gene expression values were ranked in descending order and normalised within each sample. An enrichment score was then calculated based on the difference between the rank of the genes in the signature and the rank of the remaining genes. A high GSEA score reflects a collectively high expression level of the genes in the signature in the sample. Full calculation procedures and equations for ssGSEA have previously been described by Barbie et al. [161] and Rye et al. [105].

Balancing for Tissue Composition

The stroma content is usually lower in cancer tissue compared with normal prostate tissue, and as stroma has a different gene expression profile than epithelial cells, this is a source of error in differential expression analysis [213, 214]. In paper II of this thesis, a method to reduce such confounding signals was applied when analysing differential gene expression between normal and cancer tissue samples. In this method, the samples were dived into a *balanced* sample-set where cancer (n=47) and normal (n=17) samples had approximately the same average stroma content (37% and 45%, respectively), and an *unbalanced* sample-set, where the cancer samples (n=48) had low stroma content (14%), and the normal samples (n=17) had high stroma content (72%). In the two sample-sets (*balanced* and *unbalanced*), the differentially expressed genes were investigated between the cancer and normal samples. Simplified, the results of the *balanced* sample-set give information about changes in gene expression due to cancer development, whereas the *unbalanced* results give information on alterations caused by different fractions of stroma in normal and cancer samples. This method for balancing tissue composition in gene expression analysis has recently been published by Tessem et al. [213].

TMPRSS2-ERG fusion – Paper I

In paper I, an already established gene expression signature, termed ERG, was used to investigate the TMPRSS2-ERG fusion status of the samples in the *main cohort*. This signature consists of 27 genes (Table 3.6A), and was optimised by Markert et al. [104] from three previously proposed gene sets [215–217]. An ERG score for each sample was calculated by ssGSEA, and based on this score the cancer samples were classified as high probability of having the gene fusion, ERG_{high} (n=34), if the score was increased two-fold compared to the mean. The remaining samples were divided according to their ERG score, into two equally sized groups, ERG_{low} (n=30) and $ERG_{intermediate}$ (n=31).

Wnt Pathway and Epithelial-Mesenchymal Transition (EMT) – Paper II

In paper II, the activation of the Wnt pathway was investigated, and, unlike the TMPRSS2-ERG gene fusion, no gene signature has been established for this activation in prostate cancer. A total of 196 relevant genes for the Wnt pathway and EMT were chosen from publicly available pathway maps (KEGG per March 2015), and literature [66, 218–220]. Differential expression of the genes was analysed between normal and cancer samples using the tissue composition balancing method as described above. Additionally, the gene expression between *high* and *low* *Gleason* samples was investigated. After the differential gene expression analysis, the 48 most

Table 3.6 Genes included in the (A) ERG and (B) NCWP-EMT signatures.

A. ERG – TMPRSS2-ERG gene fusion					
AMPD3	ARHGDIB	CACNA1D	CADPS	COL2A1	COL9A2
EIF5	ERG	F5	GHR	HDAC1	HLA-DMB
ITPR3	KCNN2	KCNS3	KHDRBS3	LAMC2	MYO6
OCLN	PDE3B	PEX10	PLA1A	PLA2G7	RGS10
TLE1	UBE2E3	ZNF3			
B. NCWP-EMT					
CDH2	CDH3	CDH11	FYN	FZD2	LEF1
MMP9	NKD2	PLCB2	SFRP1	SFRP2	SFRP4
VIM	TCF4	WNT5A			

central and/or significant genes were selected for further multivariate analysis. A PCA score plot of the principal component 1 and 2 was used to reveal a set of 15 genes applicable for a gene expression signature (NCWP-EMT) (Table 3.6B). The clustering of the NCWP-EMT genes was validated by PCA analyses of the *validation cohorts*. Similar as for the ERG signature, ssGSEA was performed to score the cancer samples according to their enrichments of the NCWP-EMT genes. The cancer samples of the *main cohort* were divided into three equally sized groups depending on this score: *high* (n=32), *intermediate* (n=31), and *low* (n=32) NCWP-EMT score.

SFRP4 – paper III

In paper III, the continuous gene expression values of *SFRP4* was investigated, and differential expression analyses were performed between normal and cancer samples, as well as between *low* and *high* Gleason samples, in the *main* and in the *validation cohorts*. Meta-analyses were performed to obtain combined results of all cohorts. Due to the lack of detailed tissue composition of the *validation cohorts*, no balancing for tissue heterogeneity was performed in this paper.

3.6   Analyses of the *IHC cohort*

To validate the findings of the *main cohort*, the samples of the *IHC cohort* were prepared for FISH and immunohistochemistry analysis after HR-MAS MRS. The sample preparation and staining were performed by the Cellular and Molecular Imaging Core Facility, NTNU.

Fluorescence *in situ* Hybridisation (FISH)

The TMPRSS2-ERG gene fusion status of the samples was assessed by FISH analyses on 4 μm thick formalin fixed and paraffin embedded tissue sections, which were deparaffinised before

staining. A triple-labelled colour FISH break-apart assay was performed using a commercial probe, designed to detect deletion between TMPRSS2 and ERG at 21q22 (Kreatech Diagnostics, The Netherlands). By this assay, ERG is stained with a blue signal, TMPRSS2 with a red signal, and the proximal part of TMPRSS2 (2R1G2B) with a green signal, and the loss of green signal indicates TMPRSS2-ERG gene fusion (Figure 3.4). The sections were counterstained with DAPI (4',6-diamidino-2-phenylindole), which is a fluorescent staining binding to AT rich regions of the DNA. The results were visualised using a 100X oil immersion objective on a Nikon Eclipse 90i fluorescent microscope (Nikon Corp., Japan). For each sample, 25 well preserved, non-overlapping nuclei were evaluated in previously identified cancer regions (assessed by HES staining, see section 3.3). The samples were identified as fusion-positive if the deletion was detected in at least 80% of the evaluated nuclei.
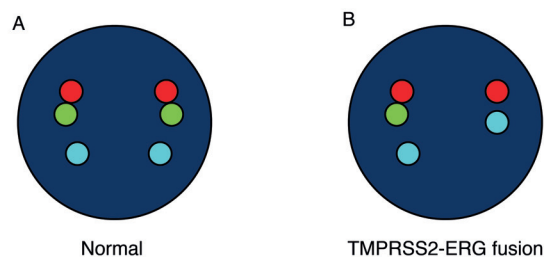
Figure 3.4 FISH break-apart assay for detection of TMPRSS2-ERG.
Triple-labelled colour FISH assay, where the loss of green signal indicates TMPRSS2-ERG gene fusion. (A) Normal nucleus without the gene fusion – two copies of each of the three colours. (B) Deletion of the green signal in one chromosome indicates fusion of TMPRSS2 and ERG. *Blue signal – ERG, red signal – TMPRSS2, and green signal – proximal part of TMPRSS2 (2R1G2B)*

Immunohistochemistry

For immunohistochemistry staining, the 4 μm thick formalin fixed paraffin embedded tissue sections were deparaffinised, and embedded with a solution of TRS (Target Retrieval Solution, high pH, Dako) for 20 minutes at 97 °C. The sections were incubated for 60 minutes at room temperature with primary mouse monoclonal antibodies against E-cadherin (Dako, clone NCH-38, dilution 1:100), N-Cadherin (Dako, clone 6G11, dilution 1:30), and Wnt5a (Sigma-Aldrich, clone 3A4, dilution 1:50), and polyclonal rabbit antibodies against β-catenin (PRESTIGE antibodies Sigma, dilution 1:300), and SFRP4 (Protein Tech catalogue: 15328-1-AP, dilution 1:200). Immunoreactive proteins were visualised using an EnVision Peroxidase/DAB+ Rabbit/Mouse (Dako), with 30 minutes incubation time. After washing, all the sections were counterstained with haematoxylin for 30 seconds. Positive and negative controls were processed

Table 3.7 Immunohistochemistry scoring for staining index (SI).

Score	0	1	2	3
Staining intensity	No detectable staining	Weak staining	Moderate staining	Strong staining
Percentage of positive cells	0%	1-10%	11-50%	>50%
Staining index	0	1,2	3,4,6	9
Staining classification	Negative	Weak	Moderate	Strong

The staining index (SI) was obtained by multiplying the score of staining intensity and the score of percentage of positive cells.

for each antibody.

Assessment of the immunohistochemistry sections was performed manually, and cancer regions were identified form the HES-stained sections (Section 3.3). The average staining signal intensity in cancer cells (0-3) multiplied by the percentage of positive cancer cells (0-3), was used to obtaining a total staining index (SI) (0-9) (Table 3.7). Examples of different staining intensities of the antibodies used in this thesis are shown in Figure 3.5. For the evaluation of β-catenin, membrane, cytoplasmic, and nuclear localisation of the staining was noted. In paper II, the scoring was validated by a pathologist experienced in immunohistochemistry, whereas in paper III, the SFRP4 scoring was performed by two independent readers, which of one was an experienced pathologist, and consensus was reached when scoring differed.

3.7 Integrated Statistical Analyses

Linear mixed models were used to investigate alterations in metabolite concentrations between the groups of ERG and NCWP-EMT scores in paper I and II, respectively. The models were built with adjustment for multiple samples per patients, and additional models were developed with correction for tissue heterogeneity (fraction of luminal space, stroma, cancer and benign glandular tissue), and Gleason grade. In paper I, metabolic alterations between the ERG score groups were also tested by multivariate analyses, using PCA and PLS-DA. For paper III, the correlation between *SFRP4* gene expression values and the concentration of the metabolite citrate and spermine were investigated by Pearson correlation coefficient, and the other members of the NCWP-EMT gene signature were investigated in the same way for comparison.

In paper I and II, the relationships between gene expression and biochemical recurrence were assessed by selecting the highest signature score of each patient. In paper III, one sample was selected by random for each patient in the cohorts with multiple samples per patient. For the
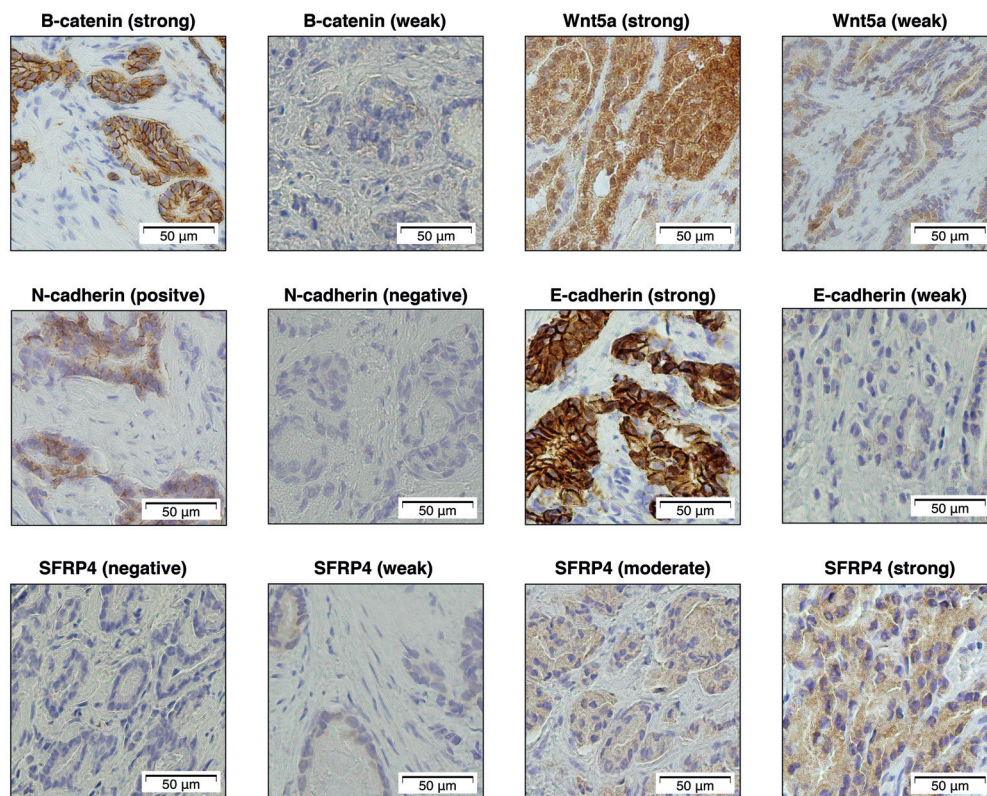
Figure 3.5 Immunohistochemistry staining examples.
Examples of immunohistochemistry staining intensities of β-catenin, Wnt5a. N-cadherin, E-cadherin, and SFRP4. *The figure is modified from Paper II with permission/Creative Commons Attribution License [86], and paper III of this thesis (unpublished).*

categorised ERG and NCWP-EMT scores in paper I and II, Kaplan-Meier plots and log-rank test were performed to investigate the differences in biochemical recurrence between the signature groups. In paper II and III, Cox PH models were used for further investigation of the relationship between gene expression and biochemical recurrence, as well as other follow-up endpoints. The continuous values of the ssGSEA score of the gene signature and the continuous expression level of *SFRP4* were used in these Cox PH models.

4. Summary of Papers

Paper I

Presence of TMPRSS2-ERG is associated with alterations of the metabolic profile in human prostate cancer

Ailin F. Hansen, Elise Sandsmark, Morten B. Rye, Alan J. Wright, Helena Bertilsson, Elin Richardsen, Trond Viset, Anna M. Bofin, Anders Angelsen, Kirsten M. Selnæs, Tone F. Bathen, May-Britt Tessem

The aim of paper I was to identify metabolic alterations associated with the presence of TMPRSS2-ERG gene fusion in prostate cancer tissue, and investigate its association with aggressiveness and recurrent disease.

Integrated *ex vivo* metabolomics, gene expression, and histopathological data were obtained from prostate tissue samples (n=129) in a cohort of 41 patients. Scores representing the likelihood of gene fusion in each sample for TMPRSS2-ERG (ERG) gene fusion was calculated based on a previously published gene expressions signature [104]. Based on this score samples were categorised into three groups: ERG_{low}, $ERG_{intermediate}$, and ERG_{high}. Differences between the metabolite levels, gene expression levels of metabolic enzymes, and frequency of biochemical recurrence were compared between the groups. Validation was performed in an independent prostate cancer cohort (n=40) using fluorescence *in situ* hybridisation (FISH) analysis to categorise the samples as TMPRSS2-ERG gene fusion positive or negative.

The study detected significant alterations across the ERG groups for the metabolites citrate, spermine, ethanolamine, glucose, glycine, phosphocholine, phosphoethanolamine, and putrescine. In addition, significant lower concentrations of citrate and spermine were detected in ERG_{high} compared with ERG_{low} samples (Figure 4.1A-B), and these alterations were more pronounced in *low Gleason* ($\leq$3+4) samples. The reduced concentrations of citrate and spermine were also

confirmed in the independent validation cohort (Figure 4.1C-D). A similar trend of reduced citrate and spermine levels was detected by *in vivo* magnetic resonance spectroscopic imaging (MRSI), indicating potential for clinical translation of the metabolic biomarkers. Furthermore, the gene expression of several key enzymes connected to citrate and spermine metabolism were significantly different between ERG$_{high}$ and ERG$_{low}$ samples. Decreased levels of citrate and spermine have previously been associated with more aggressive disease, and the findings therefore suggest TMPRSS2-ERG gene fusion to be an aggressive feature. However, no significant difference in the frequency of biochemical recurrence was detected between the ERG groups.

In conclusion, the TMPRSS2-ERG gene fusion in prostate cancer was associated with a distinct metabolic profile previously associated with aggressive disease, and this was supported by alterations in gene expression of key metabolic enzymes. The TMPRSS2-ERG gene fusion, as well as citrate and spermine, may therefore be potential candidates for improved risk stratification of prostate cancer patients, particularly in the clinical challenging group of patients with low Gleason score.
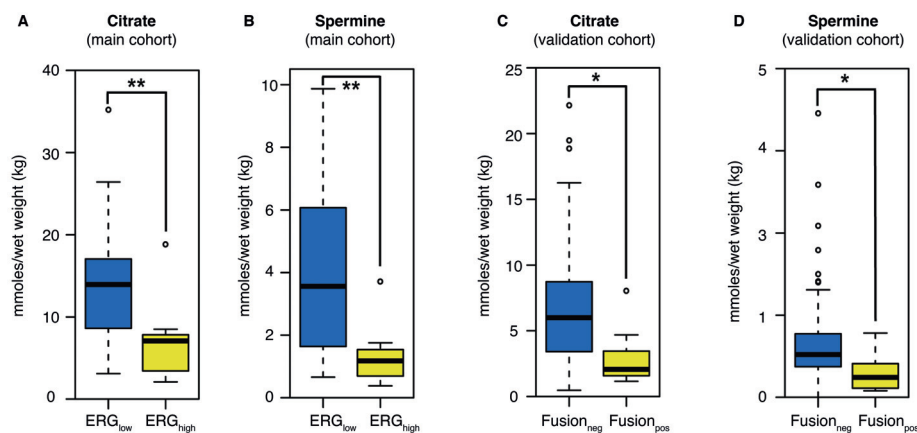
Figure 4.1 TMPRSS2-ERG status and citrate and spermine concentrations.
(A-B) Box-plots for citrate (A) and spermine (B) comparing ERG$_{low}$ and ERG$_{high}$ samples in the *main cohort*, where both metabolites were detected in significant lower concentrations in the ERG$_{high}$ samples. (C-D) Box-plots for citrate (C) and spermine (D) comparing TMPRSS2-ERG fusion in negative and positive samples in the *validation cohort*, where both metabolites were detected in significant lower concentration the fusion positive samples. *Abbreviations: pos – positive, neg – negative, * $p < 0.05$, ** $p < 0.001$. The figure is adapted from Paper II [221] under the Creative Commons Attribution License (CC BY).*

Paper II

A novel non-canonical Wnt signature for prostate cancer aggressiveness

Elise Sandsmark, Ailin F. Hansen, Kirsten M. Selnæs, Helena Bertilsson, Anna M. Bofin, Alan J. Wright, Trond Viset, Elin Richardsen, Finn Drabløs, Tone F. Bathen, May-Britt Tessem, Morten B. Rye

The aim of paper II was to identify and validate Wnt signalling and epithelial-mesenchymal transition (EMT) in prostate cancer tissue, and investigate its association with metabolic reprogramming, aggressive and recurrent disease.

Analyses were performed using integrated transcriptomic, *ex vivo* and *in vivo* metabolomics, and histopathology of a cohort of radical prostatectomy tissue samples (n=129/N=41). At least five-year follow-up data were collected for the patients (n=33). For validation, five publicly available prostate cancer gene expression cohorts were investigated (total n=1519). Additionally, an independent tissue cohort (n=40) was analysed by integrated histopathology, immunohistochemistry, and *ex vivo* metabolomics. Clinical translation of metabolic markers was investigated by *in vivo* MRSI in a small cohort (n=22/N=9).

The study detected no alterations in gene expression and immunohistochemistry indicating activation of the canonical Wnt pathway in prostate cancer. However, an increased expression of the non-canonical Wnt pathway and EMT markers were detected in high Gleason score ($\leq$3+4) cancer samples. This suggests non-canonical signalling to be the most common mode of Wnt activation in prostate cancer. The transcriptional association between the non-canonical Wnt pathway and EMT markers was confirmed in the five validation cohorts, and a novel gene expression signature for this concordant expression was developed (NCWP-EMT) (Figure 4.2A). The NCWP-EMT signature was significantly associated with metastatic events and shown to be a significant predictor of biochemical recurrence after prostatectomy (Figure 4.2B). The prediction of biochemical recurrence was strongest in patients with low Gleason score ($\leq$7) cancer, suggesting the signature to be a candidate for risk stratification in this clinical challenging patient group. The signature was also associated with decreased concentrations of the metabolites citrate and spermine, which have previously been linked to aggressive prostate cancer. Reduced citrate and spermine levels were further validated by *in vivo* MRSI, indicating a potential for clinical translation.

In conclusion, this paper demonstrates the importance of non-canonical Wnt signalling and EMT in prostate cancer aggressiveness, and the novel NCWP-EMT gene expression signature may improve risk stratification and molecular subtyping of prostate cancer patients.
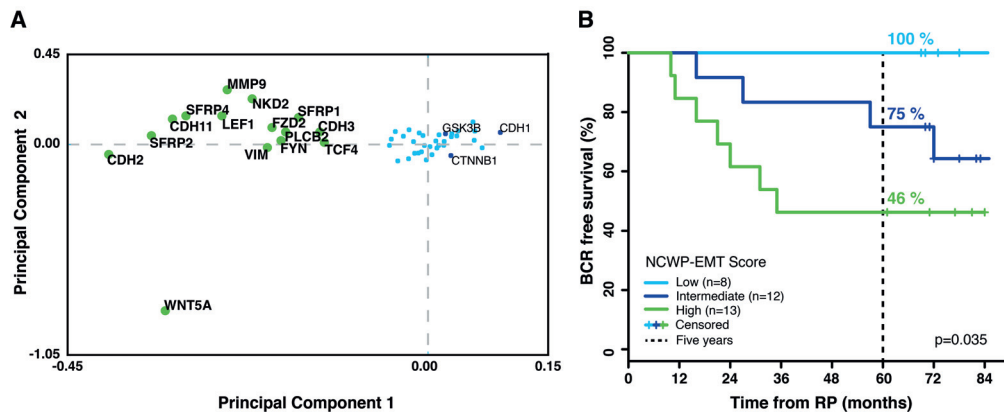
Figure 4.2 NCWP-EMT gene expression signature and its association with biochemical recurrence. (A) Principal component analysis revealed a group of 15 of 48 genes, consisting of components of the non-canonical Wnt pathway, epithelial-mesenchymal transition (EMT), and inhibitors of the canonical Wnt pathway, collectively termed NCWP-EMT (*CDH2, CDH3, CDH11, FYN, FZD2, LEF1, MMP9, NKD2, PLCB2, SFRP1, SFRP2, SFRP4, VIM, TCF4 WNT5A*). (B) Kaplan-Meier plot and log-rank statistic showed significant separation in biochemical recurrence free survival between the *low, intermediate* and *high* NCWP-EMT signature groups. The signature score was also shown to be an independent predictor of biochemical recurrence in multivariable Cox proportional hazard analysis. *Abbreviations: BCR – biochemical recurrence, RP – radical prostatectomy. The figure is adapted from Paper II [86] under the Creative Commons Attribution License (CC BY).*

Paper III

SFRP4 gene expression is increased in aggressive prostate cancer

Elise Sandsmark, Maria K. Andersen, Anna M. Bofin, Helena Bertilsson, Finn Drabløs, Tone F. Bathen, Morten B. Rye, May-Britt Tessem

Manuscript

Secreted frizzled-related protein 4 (SFRP4) is a modulator of the cancer associated Wnt pathway, and has previously been suggested as a potential marker for prostate cancer aggressiveness. In paper III, the aim was to identify and validate the association between *SFRP4* gene expression and aggressive and recurrent prostate cancer.

The study was performed by analysing *SFRP4* gene expression, concentrations of citrate and spermine, histopathology and patient follow-up data from a cohort of prostate cancer patients. The results were validated in eight independent publicly available gene expression cohorts of prostate cancer patients, which all included follow-up information (total n=2197 samples, N=1884 patients). Meta-analyses were used to get combined results for all the cohorts. Additionally, immunohistochemistry protein expression of SFRP4 was evaluated in an independent cohort with metabolomics and follow-up data (N=40).

By differential expressions and meta-analyses of all the cohorts, a significantly higher *SFRP4* gene expression was detected in cancer compared with normal samples (Figure 4.3A), and in high ($\geq$4+3) compared with low ($\leq$3+4) Gleason score samples (Figure 4.3B). The continuous *SFRP4* gene expression was a significant predictor of biochemical recurrence after prostatectomy in six of seven cohorts, and in the overall meta-analysis. Expression of *SFRP4* was also a significant predictor of metastatic events after surgery. Additionally, a significant negative correlation was seen between *SFRP4* expression values and concentrations of the metabolites citrate and spermine, two previously suggested aggressive markers in prostate cancer. Immunohistochemistry of SFRP4 was not associated with any markers for prostate cancer aggressiveness.

In conclusion, *SFRP4* gene expression was shown to be associated with aggressive prostate cancer and recurrent disease after prostatectomy. The results show *SFRP4* to be a potential biomarker candidate for prostate cancer aggressiveness, and *SFRP4* deserves further attention in prostate cancer studies.
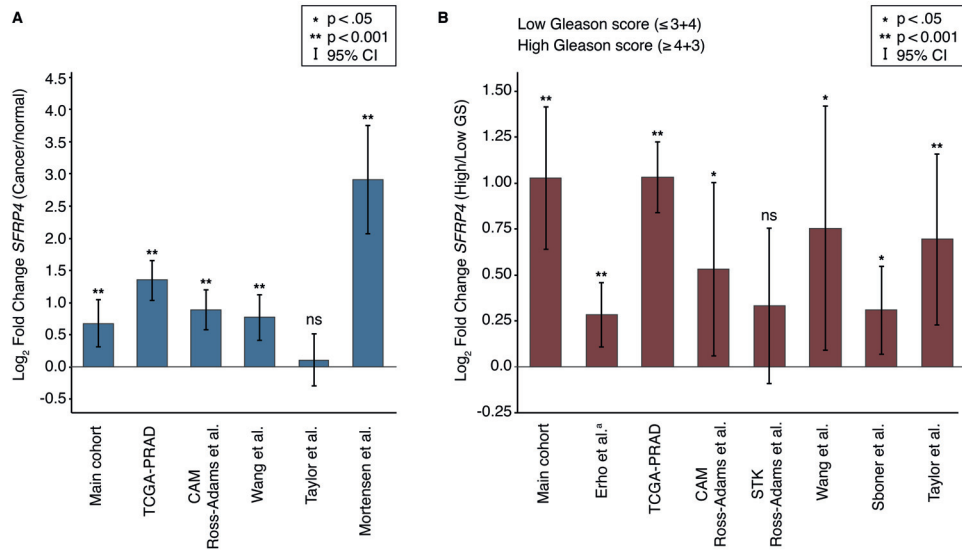
Figure 4.3 SFRP4 gene expression in prostate cancer.
(A) Log_2 fold change of *SFRP4* gene expression in the cohorts. (B) Log_2 fold change of *SFRP4* gene expression in high Gleason score ($\geq$4+3) compared with low Gleason score ($\leq$3+4) samples in the cohorts. *Abbreviations: GS – Gleason score, CI – confidence interval. [a]In the Erho et al. cohort low Gleason score was defined as $\leq$7 and high Gleason score as $\geq$8. The figure is adapted from Paper III [Sandsmark et al. unpublished].*

5. Discussion

5.1 Methodological Considerations

For scientific work, the study design is important for the validity and the interpretation of the results. Patient inclusion, sample collection, choice of experimental methods, data analyses, and statistics are all factors that may influence the outcome of a study. The strengths and limitations of the methods used, as well as their potential impact on the results and conclusions of the work in this thesis are discussed in this section.

Patient Inclusion

The work in this thesis is based on tissue from prostate cancer patients treated with radical prostatectomy at St.Olav's Hospital. In Norway, 46% of all patients diagnosed with prostate cancer (<75 years) undergo surgical treatment [10]. The treatment selection is based on risk stratification as described in Section 1.1, in combination with patient factors such as age, life expectancy and overall medical condition. Patients selected for other types of prostate cancer management, such as active surveillance, radiotherapy, and palliative care, were not investigated in the papers of this thesis. In Norway, 79% of patients undergoing radical prostatectomy treatment were preoperatively categorised to have intermediate risk prostate cancer, and this number was 90% for surgeries performed at St.Olav's Hospital [10]. However, of all patients diagnosed with prostate cancer in Norway, only 68% were categorised as intermediate risk, and the same number was 73% for patients in the Central Norway health region, for whom surgical treatment is offered at St.Olav's Hospital [10]. The cohorts in this thesis therefore have a bias towards inclusion of intermediate risk prostate cancer patients, and this is important to consider when interpreting the results of the studies. However, this study design was considered the most ethically acceptable, as the harvesting of tissue samples after prostatectomy gave no additional procedures or side-effects for the included patients. In addition, the radical prostatectomy patient group was considered highly suitable for investigation of the research questions in paper I-III,

where the overall goal was to identify molecular markers or signatures which can help in risk stratification and treatment selection for prostate cancer patients.

Validation Cohorts

To validate and strengthen the results of the studies in this thesis, sample sizes were increased by the use of validation cohorts. Five and eight independent cohorts were downloaded from publicly available databases in paper II and III, respectively (Section 3.2). All cohorts were based on prostatectomy samples, apart from Sboner et al. [206], used in paper III, which were prostate tissue from trans-urethral resection of the prostate (TURP). The patient selection, information, and methods used were out of our control. Although most of the necessary information could be gathered from the databases and from previously published papers, this lack of complete overview and control is still a limitation that should be kept in mind. However, as the validation cohorts were harvested and analysed by different methods, the universality and robustness of the findings were increased.

Patient Follow-up

Patient follow-up data were included for the *main cohort* in all papers, for the *IHC cohort* in paper III, for two of the validation cohorts in paper II, and all eight validation cohorts in paper III. This allowed for statistical analyses of the relationship between the molecular findings and patient outcome. However, the relative low number of patients in the *main cohort*, as well as the *IHC cohort*, limited the conclusions that could be drawn from the patient follow-up analyses. This was especially true for paper I, which did not include any validation cohorts, partly for paper II, and for the *IHC cohort* in paper III. Whereas the meta-analysis of biochemical recurrence data from several independent cohorts made it possible to make a stronger conclusion in paper III.

There are several confounding factors affecting biochemical recurrence and clinical failure after prostatectomy, including pre-surgical PSA level, tumour stage, capsular invasion, surgical margins, and adjuvant treatment [222, 223]. Furthermore, for interpretation of the results, it is important to recognise that only a minority of patients with biochemical recurrence will develop clinical recurrence or die of prostate cancer [31, 61]. Other follow-up measurements, such as quality of life, may be of high importance. Collection of this type of follow-up data should be considered in future studies.

Tissue Harvesting

Harvesting of tissue cores of the *main cohort* were directly guided by adjacent histopathological sections, whereas the needle biopsies of the *IHC cohort* were more blindly aimed at the location of pre-surgical positive TRUS biopsies (section 3.2). After applying a selection criteria to increase the number of cancer samples in the *IHC cohort*, only 45% of the samples contained cancer tissue. Furthermore, the cancer percentage within the samples of this cohort was relatively low, with an average of 38% (range 5–80%). A low cancer fraction could be a confounding factor for the analysis of metabolites by HR-MAS MRS, as an average of the metabolites in the whole tissue sample is measured. The low proportion of cancer in some of the samples was also a challenge for the immunohistochemistry evaluation where some stained tissue sections had to be excluded due to insufficient or lack of tumour cells. A stricter selection criteria of a larger cancer area in the TRUS biopsies could be an alternative, but this will increase the bias towards collection of patients with larger tumours. The method in the *main cohort* was more successful in harvesting samples with high cancer content (63%) where the histopathological evaluation of cryosections was performed prior to other analyses. This can be regarded as a favourable approach because unsuitable samples can be discarded from further analyses. However, the simplicity of the tissue harvesting method used in the *IHC cohort* was also an advantage, where no special equipment or particular skills were required, making it highly reproducible. These needle biopsies could also be harvested and snap frozen at the surgical department, and the short freezing time is ideal to prevent tissue degradation. In the *main cohort*, the prostatectomy specimens had to be transferred to the pathology department for tissue slice harvesting, but the use of a pneumatic tube system made the freezing time relative short, and has previously been reported to average at 15 minutes [177]. Alterations in metabolites associated with glycolysis (alanine, glucose and lactate) have been detected in rat brain tissue after 30 minutes of storage in room temperature (20°C) [224], and no changes of individual metabolites were detected for 30 minutes of freezing time in a study of breast cancer xenografts [225]. These results are likely translational to prostate tissue, where the freezing time of both cohorts was less than 30 minutes, and thus unlikely to affect the metabolomics analysis. The quality of RNA has been reported to handle several hours before degrading [226], and the freezing time of the samples in the *main cohort* was therefore not considered to affect the gene expression.

A field effect of altered gene expression has been detected in benign prostate tissue adjacent to cancer tissue [227, 228]. The harvesting method of the *main cohort* allowed for normal samples to be extracted as far away as possible from the cancer areas (Figure 3.2A). This method was expected to make the cancer field effect less extensive, and the histopathological confirmed

non-cancer samples were therefore used for as normal prostate tissue for comparison of gene expression. One advantage of using adjacent samples, is that the tissue harvesting was performed under the exact same conditions for both cancer and non-cancer samples. Some studies have tried to avoid the field effect by using autopsy biopsies as normal control samples [203, 204], but different tissue handling is a limitation for this method. Another possible study design is to use normal prostate tissue samples from surgical specimens of radical cystectomy treated bladder cancer patients. This approach was used in one of the validation cohorts included in paper III, Mortensen et al [94]. However, both bladder cancer itself and radical surgical treatment of this patient group are less frequent than prostate cancer [6], and patient inclusion may therefore be more time consuming. Although cancer samples were compared to normal samples in all three papers of this thesis, the main focus was on the differences within cancer samples, and the harvesting method used for non-cancer tissue samples was therefore considered satisfactory.

Quality of Gene Expression Analysis

The gene expression profiling in the *main cohort* of paper I-III was obtained by microarray technology. Gene expression analysis is highly dependent on the quality of the RNA transcripts, where degradation or fragmentation of mRNA will affect the measurement. RIN (RNA integrity number) from 1-10 is used to measure the RNA quality [155]. A RIN above 7 is often considered acceptable for transcriptomics studies, but there is no consensus and RIN thresholds as low as 3.95 [229] and as high as 8 [230] have been proposed. The average RIN of the samples in the *main cohort* was 9.1, with a standard deviation 1.2 [124], and the RNA quality was therefore considered to be very good. For the validation cohorts included in paper II and III, the quality of the RNA was reported as median RIN value of 5.9 (range 3.9–9.7) in the Mortensen et al. cohort [94], and samples with a RIN above seven were included in the Taylor et al. cohort [207]. Information on RNA quality were not available for the other cohorts. The relatively low RIN in the Mortensen et al. and possibly other cohorts could be a limitation for the results in the validation cohorts.

The microarray technique for measuring gene expression is dependent on the accuracy of the DNA probes. The specific binding of the targeted transcript is essential, because the cross-hybridisation of similar transcripts can be a source of error [231]. The signal intensity of low abundance transcripts could be indistinguishable from non-specific bindings (background noise), and both are therefore commonly filtered out before further analyses. Filtering was performed on gene expression data in the *main cohort* of this thesis, reducing the number of transcripts from ~47 000 to ~23 500. This resulted in genes without any measurements, and missing data

was observed for 30 of the selected 196 relevant genes in paper II, representing a limitation of the study. Microarray platforms frequently includes several probes for each gene, which may represent different slice variants of the gene. In paper III, some cohorts had two probes for the *SFRP4* gene. Several strategies for selection of probes have been proposed [232]. In general, summarisation of the probes is not recommended, as alternative transcripts or slice forms of the gene may not correlate. In paper III, the probe with the highest variance of gene expression values was selected for further analysis. The use validation cohorts with gene expression measured by different microarray platforms, reduced the likelihood of poorly produced probes to affect the overall results.

The gene expression in microarray analysis cannot be absolutely quantified [233], and can therefore not be directly compared between cohorts. This is not a problem for studies exploring the enrichment of gene sets or differential gene expression. However, for clinical translation, absolute quantification may be necessary for analysis and interpretation in individual samples. An alternative to microarray, RNA-sequencing (RNA-seq), a newer technology, allows for more accurate, quantitative and higher resolution measurement of the transcripts [157]. RNA-seq was not available at out facility at the time of analysis of the *main cohort*. However, one of the validation cohorts included in paper II and III, was based on RNA-seq technology (TCGA PRAD [201]). The agreement of the results between this cohort and the microarray based cohorts is a sign of accuracy. RNA-seq is currently more expensive, and requires extensive skills for processing and analysis compared to microarray [234]. Although several advantages of RNA-seq for gene expression exists, microarray was considered a reasonable approach for the research questions of the papers included in this thesis.

Tissue Heterogeneity in Gene Expression

Differences in the transcriptome of stroma and epithelial prostate cells are well acknowledged [235], and tissue type heterogeneity is an important challenge for differential gene expression in prostate tissue [213, 214]. In the *main cohort*, a difference in stroma content of the normal (mean 57%) and cancer (mean 28%) samples was observed (t-test $p<0.001$), which may introduce a systematic bias. To approach this issue, a method for balancing tissue composition as described in Section 3.5, was performed in paper II for detection of differentially expressed genes between normal and cancer samples. By this strategy, stroma confounding could largely be identified, and eliminated. However, a limitation is the subdivision of the samples into datasets, which reduces the sample size for the statistical analysis. Furthermore, detailed histopathology is required of

the exact same samples as used for gene expression, and this is rarely available. In paper III, the main focus was on changes within cancer samples, and this, combined with the lack of detailed histopathology of the validation cohorts, were the reasons why tissue composition balancing was not performed in this paper. However, the same finding in several cohorts and different types of samples proves *SFRP4* to be a robust marker in cancer.

Another method to overcome the challenges caused by tissue heterogeneity is laser microdissection of the tissue before gene expression profiling [236]. By this method, molecular profiles from different cell types such as stroma, epithelium and cancer can be identified. In paper III, laser microdissection was performed in the Mortensen et al. cohort [94]. This may explain why this cohort had the highest log fold change of *SFRP4* gene expression between cancer and normal samples (Figure 4.3A). However, the normal samples of this cohort were from bladder cancer patient without prostate cancer, which may also explain the high log fold change. Methods for spatial gene expression are starting to emerge, showing possibilities for localisation, visualisation, and quantification of gene expression in tissue sections, and this is a promising prospective potential for transcriptomics for both research and clinical applications [237, 238].

Transcriptome vs. Proteome

The protein expression is the product of gene expression, where the genetic information in the transcripts (mRNA) are decoded into amino acids sequences, forming proteins. The transcriptome is, however, not directly proportional to the proteome (protein expression), and the observed Spearman's rank correlation between mRNA and protein expression has been reported between 0.45 and 0.76 [239]. Although gene transcription is important in the regulation of protein expression, additional complex and diverse mechanisms regulates the abundance of proteins. One of the main regulatory steps in protein synthesis is the ribosomal translation of mRNA, and one single transcript can be translated multiple times, or not at all [240]. The half-lives and intracellular degradation of proteins will further affect the protein expression [241]. Furthermore, post-translation modifications, such a phosphorylation, can regulate the functions of proteins without increasing the transcription or translation [242]. Gene expression is still highly valuable for understanding molecular mechanism of cancer, however, the mentioned differences in transcriptome and proteome are important to remember when interpreting gene expression data. Combining gene expression with high-throughput techniques for analysing the protein expression, such as gel electrophoresis [243] or mass spectrometry [244], can give a more comprehensive understanding of cancer progression. Such proteomics analysis was not performed for the work

in this thesis, however, protein expression of the most important genes in paper II and III were investigated by immunohistochemistry.

Immunohistochemistry (IHC)

In paper II and III, protein expression of the most relevant genes was validated by immunohistochemistry (IHC) in an independent cohort, the *IHC cohort*. Using two different cohorts only allowed for a general comparison of the gene and protein expression in prostatectomy tissue samples. Further development of the tissue harvesting method used in the *main cohort*, with additional tissue sections for IHC (and FISH) analyses of the exact same sample, can allow for direct comparison of expressions of genes, proteins, and metabolites.

IHC is a relatively effective and simple method for visual 2D evaluation of protein expression. The different cells and tissue types can be evaluated separately, which made it possible to specifically study the expression in tumour cells in this thesis. In addition, the protein staining can be localised within the cells (membranous, cytoplasmic, and nuclear staining), which may give important additional information on the function of the proteins. This advantage of IHC was particularly valuable for β-catenin expression in paper II, where translocation to the nuclei is a hallmark of canonical Wnt pathway activation [245]. However, there are limitations associated with the IHC method, including poor reproducibility with a lack of standardisation in antibodies and staining protocols, as well as high reader subjectivity [246, 247]. To reduce some of these limitations, positive and negative controls were processed for all antibodies, and the evaluation were performed under guidance from an experienced pathologist in paper II, and by two readers in paper III.

Tissue microarray (TMA) cores are frequently used in IHC research studies, and can be as small as 0.6 mm in diameter [248]. The IHC staining of the tissue in a TMA section is relatively homogeneous due to its small size. However, the tissue sections of the *IHC cohort* were from 16 Gauge needle biopsies (1.7 mm diameter) with length up to ~30 mm. Staining heterogeneity within the samples were therefore a challenge for the IHC evaluation in both paper II and III. However, the size of the tissue sections in the *IHC cohort* was more similar to a clinical sample, and the challenges of staining heterogeneity are therefore highly important to acknowledge when evaluating possible clinical translations of IHC staining. Another important limitation of non-targeted biopsies in IHC analysis of cancer, is that the samples are not necessarily representative of the most aggressive part of the tumours. This may especially be a source of error for statistical analyses comparing the IHC staining with clinical parameters and follow-up status of the patients,

as performed in paper III.

A method currently emerging in the field of tissue section pathology is multicolour multiplex immunohistochemistry [249]. This method has the potential to reduce some of the presented limitations of IHC, such as poor reproducibility and reader subjectivity, by, among others, standardised and quantitative image analysis [249].

Metabolomics – HR-MAS MRS

In this thesis, HR-MAS MRS was used for detection and quantification of the metabolites in the tissue samples. The advantages of this technique include simple sample preparation, semi-automatic and high through-put acquisition [250], as well as established protocols for tissue harvesting, sample preparation, and acquisition [177–179]. Together, this ensures a high reproducibility of the method. In paper I-III, the use of HR-MAS MRS allowed for absolute quantification of metabolites by LCModel [183, 184], which permitted advanced statistical analyses of the metabolite concentrations, as well as opening for possible comparison with other studies and cohorts. In addition, multivariate analyses, a commonly used method for statistical analyses of HR-MAS MRS spectra [176], were performed in paper I, and the agreement of the results by quantification and multivariate analyses, gives extra confidence.

The clinical translation of HR-MAS MRS to *in vivo* patient MRSI, as demonstrated in paper I and II of this thesis, and previously shown by Selnæs et al [211], makes HR-MAS MRS highly relevant for identification of clinically useful biomarkers. The ongoing introduction of ultra-high-field (7 Tesla) clinical MRI scanners, will offer increased spectral resolution and higher signal-to-noise ratio of MRSI [251], and this may further increase the translational potential of *ex vivo* MRS findings.

Another main advantage is the non-destructiveness of HR-MAS MRS, where the exact same tissue samples can be further analysed after acquisition. This was demonstrated in this thesis by gene expression, histopathology, immunohistochemistry, and FISH analyses. The conservation of the tissue was an important advantage, especially in paper I where the main aim was to identify metabolic associations with the TMRPSS2-ERG gene fusion in prostate cancer. This could be assessed with a higher degree of certainty when both analyses were performed on the exact same tissue sample, avoiding the problem of cancer heterogeneity seen when using adjacent samples. The integrated analyses also allowed for the incorporation of metabolomics, hence gave a broader molecular understanding, in paper II and III.

Tissue degradation can be regarded as a possible limitation of the HR-MAS MRS technique, where the high spin rate and acquisition time may be important factors. This has been investigated in prostate tissue by Taylor et al. whom detected distortion of the ductal structures of prostate tissue after using a spin rate of 3 kHz [252]. However, in the same study, the tissue degradation did not affect the histopathological evaluation of the samples [252]. In this thesis, a 5 kHz spin rate was applied during HR-MAS MRS acquisition. The spinning was not observed to hinder the detailed histopathological, immunohistochemistry, or FISH evaluation, nor affect the RNA [177]. Tissue degradation caused by HR-MAS MRS acquisition was therefore not regarded as an issue in the work of this thesis.

One of the drawbacks of HR-MAS MRS is that only metabolites of relatively high abundance can be detected (milimolar concentrations), whereas the most commonly used alternative metabolomics technology, mass spectrometry (MS), has a higher sensitivity (picomolar concentrations) [253–255]. MS offers quantitative analyses and good separation of metabolites, however, MS requires more intricate and destructive sample preparations, and subsequent analyses of the same tissue sample can no be performed. In addition, the sample preparation can cause loss and discrimination of metabolites, as well affecting the repeatability and reproducibility of the measurements [256]. In the work included in this thesis, the non-destructiveness combined with clinical translation potential were considered to be of high importance, and HR-MAS MRS was considered the most suitable technology for metabolomics analysis. However, when interpreting the findings, it is important to recognise that the relatively low sensitivity of the technique does not offer a full overview of the metabolic status in the tissue.

For future studies, matrix-assisted laser desorption ionisation (MALDI) MS imaging could be an interesting and relevant technique for investigating the metabolism in prostate cancer. MALDI MS is commonly used in proteomics studies and is currently emerging as an analytic tool for metabolomics [257, 258]. This technique gives high sensitivity, but only requires a thin tissue section, and directly adjacent sections can be used for other tissue analyses such as histopathology and gene expression. The main advantage of the MALDI MS imaging technique is the possibility to localise the analysis to specific tissue types, such as normal epithelium, cancer, and stroma tissue.

Sample Classification

Based on histopathological grading, the cancer samples were divided into two subgroups of low Gleason score ($\leq 3+4$) and high Gleason score ($\geq 4-3$) in all included cohorts in this thesis. The reasons for choosing this cut-off were the previously detected prognostic differences Gleason score 3+4 and 4+3 [47, 48], as well as getting relatively equal sized groups for statistical analysis. Furthermore, the new Grade Group system for prostate cancer samples separates Gleason score 3+4 and 4+3 into the Grade Group 2 and 3, respectively [56]. The low and high Gleason score groups used in the papers of this thesis, is therefore in accordance with the new Grade Groups.

The samples in the *main cohort* were also divided according to ssGSEA score of the gene expression signatures. In paper I, the samples were divided into three groups depending on the ssGSEA score of the established ERG-fusion gene signature. As the gene fusion have been reported in a range from 30-80% of cancers, the use of three groups increased the probability of the samples in the ERG_{high} group to be true positive, and the ERG_{low} to be true negative. In paper II, the samples were divided into equally sized groups depending on the ssGSEA score of the developed NCWP-EMT gene signature. The frequency of the non-canonical Wnt pathway activation in prostate cancer was not previously known, however the immunohistochemistry results indicated activation in less than 50 percent of the samples. To increase the likelihood of activation in the *high* score group, and at the same time maintaining large enough sample size statistical analysis, three groups were therefore found the most appropriate for the NCWP-EMT signature.

5.2 Biological Interpretation

Understanding the molecular alterations in prostate cancer can enable identification of biomarker candidates and signatures for improved risk stratification for patients, as well as help the selection of more personalised treatment strategies. In this thesis, the TMPRSS2-ERG gene fusion, Wnt pathway, and *SFRP4*, as well as their association with aggressive disease and metabolic alterations, were investigated in human prostate cancer tissue samples. In this section, the biological interpretation and the possible clinical impact of the findings are discussed.

TMPRSS2-ERG Gene Fusion

In paper I of this thesis, the TMPRSS2-ERG gene fusion and its associations with metabolism was the focus, and the study design was optimised for this. Further the relationship between TMPRSS2-ERG and biochemical recurrence was also investigated. Previously there has been

inconsistent results regarding the association of TMPRSS2-ERG gene fusion and prostate cancer aggressiveness. A large meta-analysis of 48 different studies concluded the gene fusion not to be a strong predictor of recurrence or mortality in prostatectomy treated patients [109]. This is in agreement with the findings of paper I, where no significant difference in biochemical recurrence between patients with ERG_{high} and ERG_{low} score was detected. When restricting the analyses to low Gleason score ($\leq$3+4) samples, none of the patients in the ERG_{low} group experienced biochemical recurrence during time of follow-up. However, the statistical comparison of the biochemical recurrence in between patients with ERG_{high} and ERG_{low} was not significant. Due to the relatively low number of patients, especially when restricting the analysis to low Gleason score samples, the study did not have the statistical power to make any conclusions. Further investigation of the association between TMPRSS2-ERG and clinical outcome in prostate cancer patients with low Gleason score may be of interest, as improved risk stratification is needed in this patient group for selection of patients for active surveillance.

Several metabolic differences in ERG_{high} and ERG_{low} tissue samples were detected in paper I. Of particular interest, the metabolites citrate and spermine were significantly lower in ERG_{high} samples. These metabolic alterations have previously been associated with high Gleason score [126], and, recently, biochemical recurrence [259]. This may suggest ERG fusion to be associated with a more aggressive metabolic pattern. Interestingly, the alterations of citrate and spermine were more profound when separately investigating *low Gleason* samples. This may suggest *low Gleason* ERG-fusion positive prostate cancer to have a more similar metabolism to *high Gleason* prostate cancer, however, further validation in larger cohorts are needed. The metabolic alterations where further supported by changes in key metabolic enzymes of the citrate and polyamine metabolism, suggesting these metabolic pathways to possibly be regulated differently in prostate cancer possessing the TMPRSS2-ERG gene fusion. The alterations in metabolites and enzymes for each metabolic pathway observed for the TMPRSS2-ERG gene fusion are further discussed in the separate metabolomics part of the discussion (Section 5.3).

Wnt Signalling Pathway

The Canonical Wnt pathway

Increased activation of the canonical Wnt pathway has previously been linked to aggressive features in prostate cancer [260], and drugs targeted to inhibit Wnt signalling have shown promising results in prostate cancer cell lines [261, 262]. When investigating the Wnt pathway in paper II, the expected finding was therefore signs of increased activation of the canonical Wnt pathway, by upregulation of relevant genes and immunohistochemical detection of nuclear translocation

of β-catenin. However, the findings of paper II did not confirm this, neither in cancer compared with normal samples, nor in high Gleason compared with low Gleason samples. Some suggested reasons for the discrepancy between the result of paper II and previous findings are therefore discussed below.

First, the samples used in paper II were from prostatectomy patients diagnosed with local or locally advanced prostate cancer, whereas the canonical Wnt pathway has mostly been associated with advanced disease, such as androgen resistant prostate cancer [78], and metastatic disease [79]. The canonical Wnt pathway may therefore still be important in advanced and metastatic prostate cancer. The findings in paper II suggest the canonical Wnt pathway to be inappropriate for early risk stratification or early targeted treatment in prostate cancer patient.

Secondly, most of the previous studies of the canonical Wnt pathway in prostate cancer have been performed on cell lines [77, 78]. In cancer research, cell lines are powerful model systems to obtain understanding of the mechanisms of pathway activity. Nonetheless, the cells are frequently derived from advanced types of cancer, and may be genetically modified to obtain features such as immortality, and the primary cancer properties might have been changed [263]. The discrepancy between cell lines studies and the findings in paper II regarding the canonical Wnt pathway, may therefore reflect the differences between cell lines and human prostate cancer tissue. This highlights the importance of validation of cell lines findings, in primary cells, but also in human tissue as the tumour cell environment cannot be completely reproduced *in vitro*.

Finally, balancing the tissue samples for stroma fraction in paper II, revealed substantial stroma confounding in several of the central canonical Wnt pathway genes. Previous studies of differential expression between prostate cancer and benign prostate tissue may therefore be affected by the natural differences in stroma content, further explaining discrepancies from previous studies of the canonical Wnt pathway in prostate cancer tissue.

Non-Canonical Wnt Pathway

Increased expression of several of the components in the non-canonical Wnt pathway, particularly matching the newly discovered Wnt5/Fzd2 pathway, was detected in a subset of prostate cancer samples in paper II. Furthermore, concordant increased expression of epithelial-mesenchymal transition (EMT) markers was identified. This concordant expression was validated in five independent validation cohorts, and a gene expression signature for non-canonical Wnt pathway EMT (NCWP-EMT) was developed. This signature represents the central components in the

non-canonical Wnt pathway, and increased expression of the signature suggests activation of the pathway, but this should be further validated by functional studies in cell cultures.

The continuous NCWP-EMT signature score was shown to be a predictor of biochemical recurrence by Cox Proportional Hazard analysis. This was further demonstrated by Kaplan-Meier analysis, where patients with samples classified as *high* NCWP-EMT score, had significantly higher rates of biochemical recurrence compared to both *intermediate* and *low* NCWP-EMT score. In fact, none of the patients with *low* NCWP-EMT score experienced biochemical recurrence during follow-up. This could, however, not be validated in an independent validation cohort (n=131, Taylor et al. [207]), although a non-significant similar pattern, separating *low, intermediate* and *high* NCWP-EMT in the Kaplan-Meier plot was shown. However, some shortcomings of this validation cohort may have affected the result. This cohort had only one sample per patient, samples were not necessarily extracted from the most aggressive cancer foci, and many patients were lost early during follow-up. In a larger validation cohort (n=545, Erho et al. [154]), samples with *high* NCWP-EMT score was significantly associated with metastatic progression after surgery, and this further supported the NCWP-EMT signature to be associated with worse prognosis. In the *main cohort*, there was also a non-significant, but visual separation of biochemical recurrence in *low, intermediate* and *high* NCWP-EMT in patients with a Gleason score ≤7. This patients group also had a higher hazard ratio for biochemical recurrence in Cox PH analysis of the continuous NCWP-EMT score compared with patients having a Gleason score ≥8. This may indicate a potential for clinical risk stratification in the challenging group of patients with low Gleason score. However, as for TMPRSS2-ERG, the low number of patients reduced the statistical power, and studies in larger patient cohorts are necessary.

The NCWP-EMT signature was significantly associated with the concentration of the metabolites citrate and spermine. Reduced concentrations of these metabolites have previously been associated with aggressive prostate cancer [126, 142, 259], and these findings further supports the NCWP-EMT to be associated with worse prognosis. However, possible mechanisms between citrate and spermine concentrations and non-canonical Wnt pathway were not investigated in this study.

In general, the results of paper II points towards non-canonical Wnt5/Fzd2 Wnt pathway activation, combined with EMT, to be associated with aggressive prostate cancer. This is in agreement with the findings by Gujral et al. where the Wnt5/Fzd2 pathway was detected to be a predictor of metastasis and survival in hepatocellular carcinoma patients [66]. However, larger cohorts are

needed for validation and refinements of the NCWP-EMT signature, as well as for evaluation of the causal relation and mechanisms of pathway activation in prostate cancer.

The role of WNT5A

WNT5A is a ligand which may activate the non-canonical Wnt pathway, and this ligand was a part of the NCWP-EMT gene expression signature developed in paper II. The reported role of WNT5A in prostate cancer has been inconsistent, where it has been associated with both good [83–85] and worse prognosis [82]. In paper II, *WNT5A* gene expression seemed to be an aggressive marker, as it was increased in high compared with low Gleason score cancer samples. However, *WNT5A* gene expression was actually higher in normal samples compared to low Gleason cancer samples. Previously, WNT5A has been detected as a tumour promoter in colon and thyroid cancer [264, 265]. However, it has also been shown to antagonise and inhibit canonical Wnt signalling [266, 267], and a tumour suppressor role of WNT5A has been observed in several cancers including melanoma, pancreatic and gastric cancer [264, 265]. A hypothesis could therefore be that in normal prostate cells, WNT5A has a tumour suppressing role, perhaps by inhibiting the canonical Wnt pathway. Therefore, WNT5A expression in cancer may be associated with good prognosis. On the other hand, if WNT5A expression increases during tumour progression, this may suggest activation of the non-canonical Wnt pathway and worse prognosis. This hypothesis of shifting roles of WNT5A, could explain the disagreement in the previous studies of prognostic outcome associated with its expression in prostate cancer. When using the NCWP-EMT signature as a biomarker, rather than *WNT5A* alone, the potential problem of the hypothetical dual roles of WNT5A may be reduced. This is because the gene signature relies on overexpression of several genes of the non-canonical Wnt pathway and EMT markers. Further investigation of the role of WNT5A in prostate cancer is warranted.

Secreted Frizzled-Related Protein 4 (SFRP4)

Of the genes in the NCWP-EMT signature, secreted frizzled-related protein 4 (*SFRP4*) had the highest negative correlation with concentrations of the metabolites citrate and spermine. SFRP4 is classified as a tumour suppressor due to its inhibition of the Wnt pathway [268]. Decreased gene expression of *SFRP4* has previously been detected in several types of cancers, including endometrial, ovarian, bladder and oesophageal cancer [91]. However, some studies of prostate cancer tissue have implied a possible opposite role of *SFRP4*, where expression has been associated with more aggressive disease [92, 94]. Both the metabolic correlation and the contradictory findings of *SFRP4* expression in prostate cancer compared with other cancers, made further investigation and validation of *SFRP4* expression intriguing, and resulted in the

work presented in paper III of this thesis.

In paper III, significantly higher gene expression of *SFRP4* was detected in prostate cancer compared with normal tissue. This is in agreement with previous findings of *SFRP4* in two small studies of human prostate tissue (n=16 and n=56) [92, 93]. The multiple cohort and large sample size (n=1237) in paper III, added substantial validation for *SFRP4* expression to be increased in prostate cancer. Additionally, significantly higher *SFRP4* expression was detected in high Gleason score (≥4+3) compared with low Gleason score (≤3+4) cancer samples. The continuous *SFRP4* values were detected to be a predictor of biochemical recurrence and metastasis after radical prostatectomy. This suggest *SFRP4* expression to be associated with more aggressive prostate cancer, which is in concordance with previous studies of *SFRP4* in prostate cancer [92, 94]. Furthermore, the results of paper III supports the inclusion of *SFRP4* as a part of previously developed signatures for prostate cancer aggressiveness, including two signature developed by Mortensen et al. [94], the commercially available Oncotype DX prostate signature [153], and the NCWP-EMT signature from paper II of this thesis [86].

Although *SFRP4* expression in prostate cancer tissue seems to be associated with aggressive disease, a few cell line studies have supported tumour suppressor properties of SFRP4 also in prostate cancer. This includes association with reduced cellular proliferation [95, 96] and reduced expression in cancer compared with normal control cells [269]. As discussed for the canonical Wnt pathway, this disagreement may be attributed to the differences between cancer tissue and cell lines. However, another cell line study was in accordance with the findings in tissue, where SFRP4 was detected upregulated in all prostate cancer cell lines (LNCaP, PC3, DU145 and 22Rv1) compared with control cells [270].

The results of paper III, indicate *SFRP4* expression to be a possible tissue biomarker for prostate cancer aggressiveness, however, direct clinical application of *SFRP4* was not assessed in this thesis. Opportunities may include absolute quantification of *SFRP4* expression by real time PCR in tissue biopsies for risk stratification of patients. A recent conference abstract indicated increased SFRP4 in urine as a method for detection of prostate cancer [271], and a recently published patent included *SFRP4* gene expression in serum as a marker for predicting prostate cancer aggressiveness [272]. This suggest potential for SFRP4 to be a biomarker for prostate cancer also by less invasive methods, and *SFRP4* deserves further attention in prostate cancer studies.

5.3 Metabolic Reprogramming in Prostate Cancer

Reprogramming of metabolism is one of the hallmarks of cancer [112], and in this thesis, metabolic alterations were associated with TMPRSS2-ERG gene fusion (paper I), non-canonical Wnt pathway and EMT (paper II), as well as *SFRP4* expression (paper III). In this section, the metabolic alterations of all papers are interpreted together by each metabolic pathway.

Citrate, Energy, and Fatty Acid Metabolism

Reduced concentration of citrate was detected in prostate cancer tissue samples with high signature scores (ERG and NCWP-EMT). In addition, a negative correlation between citrate concentration and *SFRP4* gene expression was detected in paper III. This may indicate a loss of the excessive citrate production of normal prostate cells. Previously, a loss of zinc accumulation has been shown in prostate cancer, which in turns activate the enzyme ACON (*ACO1/2*), and as a result citrate may be transformed to isocitrate in the TCA cycle and used for energy production (Figure 1.8 and 5.1) [273, 274]. However, opposite of expected, reduced expression values of both ACON genes (*ACO1/2*) was detected in ERG$_{high}$ compared with ERG$_{low}$ samples in paper I (Figure 5.1). This is in agreement with a previous study detecting significant positive covariance between citrate level and *ACON* expression, hence suggesting low citrate levels to be associated with reduced *ACON* expression [124]. This may imply that low concentration of citrate is not due to increased utilisation and energy production by the TCA cycle.

Citrate can also be a precursor for fatty acid synthesis, which has been associated with aggressive features of prostate cancer [121]. Increased fatty acid synthesis may therefore be another hypothesis for the reduced citrate concentration detected in cancer samples with high ERG and NCWP-EMT signatures scores and increased *SFRP4* expression. In paper I, increased gene expression of key lipogenic enzymes, including acetyl-CoA carboxylase alpha (*ACACA*) and fatty acid synthase (*FASN*) were observed in ERG$_{high}$ compared with ERG$_{low}$ cancer samples (Figure 5.1). This may indicate increased fatty acid synthesis in cancer possessing the TMPRSS2-ERG gene fusion.

Furthermore, a significantly increased expression of key enzymes of the pentose phosphate pathway were detected in ERG$_{high}$ compared with ERG$_{low}$ samples (Figure 5.1). This may suggest glucose to be used for nucleotide and fatty acid production by the pentose phosphate pathway, possibly instead of citrate production. Aerobic glycolysis is also a common pathway for increased glucose utilisation in cancer cells (Figure 1.8B) [116], however, lactate concentration

was not altered across the signature scores in paper I and II, indicating no differences in aerobic glycolysis within the cancer samples.
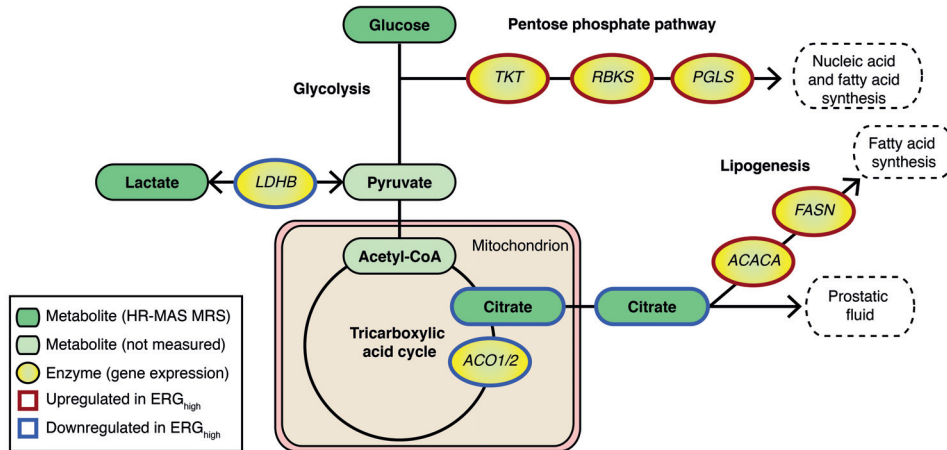
Figure 5.1 TMPRSS2-ERG relation to the citrate, energy and fatty acid metabolism.
Schematic representation of pathways and gene expression levels of associated key enzymes altered due to presence of the TMPRSS2-ERG gene fusion. *Gene/protein names: ACO1/2 – aconitase 1/2, ACACA – acetyl-CoA carboxylase alpha, FASN – fatty acid synthase, PGLS – 6-phosphogluconolactonase, RBKS – ribokinase, and TKT – transketolase. Blue = downregulation, red = upregulation in ERG$_{high}$ compared with ERG$_{low}$ tissue samples.*

Polyamine metabolism

In paper I-III of this thesis, reduced spermine concentration was associated with ERG$_{high}$, *high* NCWP-EMT, and higher *SFRP4* expression in prostate cancer samples. Reduced level of spermine has previously been observed in prostate cancer compared with normal prostate tissue [142], and a further decrease has been detected in prostate cancer with high Gleason score [126]. The mechanisms of spermine reduction in prostate cancer is not completely understood, however, the genes of the polyamine pathway were generally observed to be upregulated in ERG$_{high}$ compared with ERG$_{low}$ cancer samples in paper I (Figure 5.2). This may indicate an upregulation and a high flux through the polyamine pathway. This high flux together with the reduced spermine concentration in ERG$_{high}$ samples might be explained by the strong upregulation of the *SAT1* gene expression in the same samples (Figure 5.2). The SAT1 is the rate-limiting enzyme of spermine and spermidine catalysis, and has previously been shown to reduce intracellular concentration of polyamines [275].

Additionally, ERG$_{high}$ samples were associated with lower concentration of putrescine, fur-

ther supporting this high flux theory. In paper II, no alteration in putrescine concentration were detected between *high* and *low* NCWP-EMT score samples, and this may indicate slightly different mechanism for spermine reduction in prostate cancer with TMPRSS2-ERG and non-canonical Wnt pathway activation, however, the mechanism was not further investigated in this thesis. The androgen regulated ODC enzyme controls the rate-limiting step of the polyamine metabolism; conversion of ornithine to putresine (Figure 5.2). ODC has been described as an oncogene, and increased gene expression of *ODC* has been reported in prostate cancer tissue [138]. In paper I, when comparing ERG_{high} with ERG_{low} samples, the *ODC1* expression was only slightly upregulated in contrast to the strong upregulation of the other enzymes in the pathway, and this may explain the depletion of putrescine in ERG_{high} samples.
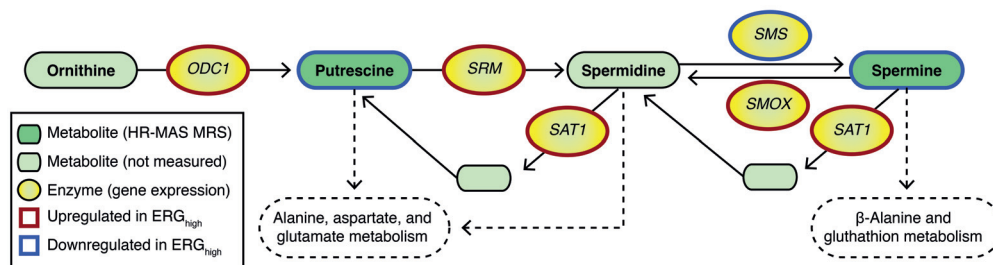
Figure 5.2 TMPRSS2-ERG gene fusion and the polyamine metabolism.
Schematic representation of the polyamine pathway and gene expression levels of associated key enzymes altered due to presence of TMPRSS2-ERG gene fusion. *Gene/protein names: ODC1 – ornithine decarboxylase 1, SRM – spermidine synthase, SMS – spermine synthase, SAT1 – spermidine/spermine N1-acetyltransferase 1. Blue – downregulation, red – upregulation in ERG$_{high}$ compared with ERG$_{low}$ tissue samples.*

Choline Phospholipid Metabolism

The choline phospholipid metabolism is crucial for biosynthesis of cell membranes which is needed by proliferating cells (Figure 1.7) [276]. In paper I, there were significant increasing concentrations of the metabolites phosphocholine and phosphoetanolamine with increasing ERG scores. These metabolites have previously been detected to be upregulated in prostate cancer compared with normal prostate tissue [126, 131, 277]. A hypothesis could therefore be that TMPRSS2-ERG fusion positive cancer has higher proliferation than fusion negative prostate cancer. In paper II, no alterations were detected of the metabolites in the choline phospholipid metabolism when comparing *high* with *low* NCWP-EMT score, possibly suggesting proliferation to be less important for the aggressiveness associated with the non-canonical Wnt pathway.

Luminal Space

In this thesis, TMPRSS2-ERG gene fusion in paper I, non-canonical Wnt pathway and EMT activation in paper II, and SFRP4 expression in paper III, were associated or correlated with reduced concentration of citrate and spermine. However, citrate and spermine are stored in the luminal space of the prostate glands, and it is debated if such reduced concentrations represent true alterations of the metabolism in cancer cells, or mainly reflects morphological changes with fewer and smaller glands. The effect of luminal space on citrate and spermine concentrations were therefore investigated in paper I and II, where moderate correlations between citrate and spermine concentrations and the fraction of luminal space were detected (r=0.369 and r=0.415, respectively). When correcting for luminal space fractions in the statistical analyses of metabolite concentrations across gene expression signature groups, highly significant reductions of citrate and spermine were still shown in both paper I and II. These results indicate the observed alterations of citrate and spermine to be a combination of morphological changes and true reprogramming of metabolism. This is in agreement with a study by Swanson et al., where citrate and spermine in luminal space could be investigated separately due to shorter MR relaxation time in the fluid-like environment [131].

Potential Metabolic Biomarkers

In paper II and III, the NCWP-EMT and *SFRP4* expression were associated with aggressive and recurrent disease, and their association with reduced concentration of citrate and spermine, further shows potential for these two metabolites to be prognostic biomarkers in prostate cancer. Furthermore, in paper II, possible *in vivo* translation was shown using MRSI, although significant, the cohort was too small to make any absolute conclusions. Further investigation of citrate and spermine as potential prognostic biomarkers in prostate cancer are therefore needed.

5.4   Clinical Implications

The overarching goal of all cancer research, including the work of this thesis, is to enable a future benefit for cancer patients. However, clinical implementation of basic research is not necessarily a straight forward process. The overall aim of this thesis was to identify candidates for molecular biomarker and signatures for improved risk stratification of prostate cancer patients. In this thesis, prostate cancer with TMPRSS2-ERG gene fusion was shown to be linked with a more aggressive metabolic pattern. Furthermore, a new gene expression signature was developed for non-canonical Wnt pathway and EMT, and this signature along with *SFRP4* expression was shown to be a predictor of biochemical recurrence in prostatectomy treated prostate cancer

patients. Citrate and spermine were also shown to be potential metabolic prognostic biomarkers. All these results gave increased molecular understanding of the differences between indolent and aggressive prostate cancer. However, further validation as well as investigation into how these findings can be used to improve risk stratification in prostate cancer patients in a clinical setting, are needed. The molecular understanding of prostate cancer progression might also be useful for selection of pathways to investigate for targeted drug therapy in prostate cancer. Although no direct clinical implication can be drawn, basic research, as performed in this thesis, is in general important for future progression of prostate cancer treatment and management strategies.

6. Concluding Remarks and Future Perspectives

The scope of this thesis was to obtain molecular information to identify biomarker candidates and signatures that may improve risk stratification of prostate cancer patients. Gene expression, MR-based metabolomics, detailed histopathology, immunohistochemistry, and fluorescence *in situ* hybridisation techniques were used on prostate tissue samples in an integrated fashion to reveal intricate, multi-level molecular relations. Follow-up of the patients allowed for investigation of the relationship between molecular alterations and cancer recurrence after surgery. The general aims of the presented work were to investigate two specific molecular alterations, the prostate cancer specific TMPRSS2-ERG gene fusion, and the cancer relevant Wnt signalling pathway. This thesis includes both an overview of the Wnt pathway activation, as well as a closer look into one of its important components, SFRP4.

The presence of TMPRSS2-ERG gene fusion in prostate cancer was associated with a distinct metabolic profile, where reduced concentrations of the metabolites citrate and spermine were the most prominent alterations. This was supported by concordant changes in the gene expression levels of key enzymes of the relevant metabolic pathways. The results indicated that prostate cancer with TMPRSS2-ERG gene fusion tended to differentiate towards a metabolic phenotype previously associated with aggressive prostate cancer.

The investigation of the Wnt signalling pathway revealed a gene expression pattern indicating activation of the non-canonical, rather than the canonical Wnt pathway in prostate cancer samples. This was combined with increased expression of epithelial-mesenchymal transition (EMT) markers, and a novel gene expression signature, NCWP-EMT, was developed for this concordant activation. The signature was shown to be a predictor of biochemical recurrence, and was associated with metastatic cancer progression after surgery. The NCWP-EMT signature may therefore be useful for risk stratification of prostate cancer patients. However, further refinement and validation of the signature in larger cohorts are necessary.

Concluding Remarks and Future Perspectives

Although the samples size was too small to make any conclusions, both the TMPRSS2-ERG gene fusion and the NCWP-EMT signature showed patterns indicating them as possible prognostic biomarkers in cancers with low Gleason score. For low Gleason, separation between patients in need of active treatment and patients suitable for active surveillance is a major clinical challenge. Thus, there is a need for new biomarkers for this patient group to prevent overtreatment of indolent and undertreament of aggressive cancers. Therefore, further investigation of the possible connection between cancer progression and both the TMRPSS2-ERG and NCWP-EMT, may be valuable in larger cohorts of patients with low Gleason score prostate cancers.

Gene expression of *SFRP4* was detected as a significant predictor of biochemical recurrence and metastasis in prostate cancer patients, and may therefore also be a potential biomarker for early prediction of prostate cancer aggressiveness. In addition to tissue sample measurements, *SFRP4* gene and protein expression may have a promising role for detection and risk stratification of prostate cancer by less invasive methods, such as serum and urine measurements. Further evaluation of potential clinical use of *SFRP4* is therefore required.

The mechanisms of the non-canonical Wnt pathway and SFRP4 in prostate cancer were not directly investigated in this thesis. Future studies, including functional studies of cell cultures, would be of great interest for validation and increased understanding of this activation. Additionally, increased knowledge of the signalling cascade and its function could lead to the discovery of potential targets for cancer therapy.

Reduced concentrations of the metabolites citrate and spermine were associated with all the molecular alterations detected in the work of this thesis: TMPRSS2-ERG gene fusion, non-canonical Wnt pathway activation by the NCWP-EMT signature, and expression level of *SFRP4*. Citrate and spermine may therefore be regarded as candidate tissue biomarkers for prostate cancer aggressiveness. Potential clinical translation of these metabolic biomarkers was shown by *in vivo* patient magnetic resonance spectroscopic imaging (MRSI), but the sample size was small, and further investigation is recommended. Additionally, functional studies investigating possible direct mechanisms between the gene expression and metabolic alterations are warranted.

Spatial transcriptomics and MALDI metabolomics are emerging techniques that make it possible to locate gene expression and metabolic alterations to different cells and tissue types. These methods could be beneficial for future wok following on from this thesis as a means to validate the existing finding, and to reduce the confounding factor of tissue type heterogeneity. Further-

more, adding high-throughput proteomics analysis of the same tissue samples may give a more complete understanding of the detected molecular alterations.

To summarise, the findings presented in this thesis suggest non-canonical Wnt pathway signalling and *SFRP4* expression to be potential candidates for improved risk stratification in prostate cancer patients. The gene fusion of TMPRSS2-ERG, activation of the non-canonical Wnt pathway, and increased *SFRP4* expression in prostate cancer were all associated with reduced concentrations of the metabolites citrate and spermine. These metabolites may therefore have potential as metabolic markers for early detection of prostate cancer, and stratification of phenotypes and aggressiveness. The TMRPSS2-ERG gene fusion, non-canonical Wnt pathway, *SFRP4*, as well as citrate and spermine deserve further attention in prostate cancer research.

References

1. J. E. McNeal, "The zonal anatomy of the prostate.," *Prostate*, vol. 2, no. 1, pp. 35–49, 1981.

2. J. P. Kavanagh, "Isocitric and citric acid in human prostatic and seminal fluid: implications for prostatic metabolism and secretion.," *Prostate*, vol. 24, no. 3, pp. 139–142, 1994.

3. A. Medrano, J. M. Fernandez-Novell, L. Ramio, J. Alvarez, E. Goldberg, M. Montserrat Rivera, J. J. Guinovart, T. Rigau, and J. E. Rodriguez-Gil, "Utilization of citrate and lactate through a lactate dehydrogenase and atp-regulated pathway in boar spermatozoa.," *Mol Reprod Dev*, vol. 73, pp. 369–378, Mar 2006.

4. P. L. C. Lefevre, M.-F. Palin, and B. D. Murphy, "Polyamines on the reproductive landscape.," *Endocr Rev*, vol. 32, pp. 694–712, Oct 2011.

5. L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," *CA Cancer J Clin*, vol. 65, pp. 87–108, Mar 2015.

6. *Cancer in Norway 2015 - Cancer incidence, mortality, survival and prevalence in Norway.* Oslo: Cancer Registry of Norway, 2016.

7. F. Bray, J. Lortet-Tieulent, J. Ferlay, D. Forman, and A. Auvinen, "Prostate cancer incidence and mortality trends in 37 european countries: an overview," *Eur J Cancer*, vol. 46, pp. 3040–52, Nov 2010.

8. E. Hernes, L. A. Johansson, S. D. Fosså, A. G. Pedersen, and E. Glattre, "High prostate cancer mortality in norway evaluated by automated classification of medical entities," *Eur J Cancer Prev*, vol. 17, pp. 331–5, Aug 2008.

9. R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," *CA Cancer J Clin*, vol. 65, no. 1, pp. 5–29, 2015.

10. *Årsrapport 2015 - Nasjonalt kvalitetsregister for prostatakreft.* Cancer Registery of Norway, 2016.

11. T. Lloyd, L. Hounsome, A. Mehay, S. Mee, J. Verne, and A. Cooper, "Lifetime risk of being diagnosed with, or dying from, prostate cancer by major ethnic group in england 2008–2010," *BMC medicine*, vol. 13, no. 1, p. 171, 2015.

12. F. T. Odedina, T. O. Akinremi, F. Chinegwundoh, R. Roberts, D. Yu, R. R. Reams, M. L. Freedman, B. Rivers, B. L. Green, and N. Kumar, "Prostate cancer disparities in black men of african descent: a comparative literature review of prostate cancer burden among black men in the united states, caribbean, united kingdom, and west africa," *Infectious agents and cancer*, vol. 4, no. 1, p. S2, 2009.

13. M. M. Center, A. Jemal, J. Lortet-Tieulent, E. Ward, J. Ferlay, O. Brawley, and F. Bray, "International variation in prostate cancer incidence and mortality rates," *European urology*, vol. 61, no. 6, pp. 1079–1092, 2012.

14. T. Kimura, "East meets west: ethnic differences in prostate cancer epidemiology between east asians and caucasians," *Chinese journal of cancer*, vol. 31, no. 9, p. 421, 2012.

15. G. D. Steinberg, B. S. Carter, T. H. Beaty, B. Childs, and P. C. Walsh, "Family history and the risk of prostate cancer," *Prostate*, vol. 17, no. 4, pp. 337–47, 1990.

16. D. Thompson, D. F. Easton, and Breast Cancer Linkage Consortium, "Cancer incidence in brca1 mutation carriers," *J Natl Cancer Inst*, vol. 94, pp. 1358–65, Sep 2002.

References

17. S. M. Edwards, Z. Kote-Jarai, J. Meitz, R. Hamoudi, Q. Hope, P. Osin, R. Jackson, C. Southgate, R. Singh, A. Falconer, D. P. Dearnaley, A. Ardern-Jones, A. Murkin, A. Dowe, J. Kelly, S. Williams, R. Oram, M. Stevens, D. M. Teare, B. A. J. Ponder, S. A. Gayther, D. F. Easton, R. A. Eeles, Cancer Research UK/Bristish Prostate Group UK Familial Prostate Cancer Study Collaborators, and British Association of Urological Surgeons Section of Oncology, "Two percent of men with early-onset prostate cancer harbor germline mutations in the brca2 gene," *Am J Hum Genet*, vol. 72, pp. 1–12, Jan 2003.

18. I. Agalliu, R. Gern, S. Leanza, and R. D. Burk, "Associations of high-grade prostate cancer with brca1 and brca2 founder mutations," *Clin Cancer Res*, vol. 15, pp. 1112–20, Feb 2009.

19. C. M. Ewing, A. M. Ray, E. M. Lange, K. A. Zuhlke, C. M. Robbins, W. D. Tembe, K. E. Wiley, S. D. Isaacs, D. Johng, Y. Wang, C. Bizon, G. Yan, M. Gielzak, A. W. Partin, V. Shanmugam, T. Izatt, S. Sinari, D. W. Craig, S. L. Zheng, P. C. Walsh, J. E. Montie, J. Xu, J. D. Carpten, W. B. Isaacs, and K. A. Cooney, "Germline mutations in hoxb13 and prostate-cancer risk," *N Engl J Med*, vol. 366, pp. 141–9, Jan 2012.

20. R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics, 2012," *CA: a cancer journal for clinicians*, vol. 62, no. 1, pp. 10–29, 2012.

21. L. D. Papsidero, M. C. Wang, L. A. Valenzuela, G. P. Murphy, and T. M. Chu, "A prostate antigen in sera of prostatic cancer patients," *Cancer research*, vol. 40, no. 7, pp. 2428–2432, 1980.

22. T. A. Stamey, N. Yang, A. R. Hay, J. E. McNeal, F. S. Freiha, and E. Redwine, "Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate," *N Engl J Med*, vol. 317, pp. 909–16, Oct 1987.

23. W. J. Catalona, D. S. Smith, T. L. Ratliff, K. M. Dodds, D. E. Coplen, J. J. Yuan, J. A. Petros, and G. L. Andriole, "Measurement of prostate-specific antigen in serum as a screening test for prostate cancer," *New England Journal of Medicine*, vol. 324, no. 17, pp. 1156–1161, 1991.

24. G. Aus, S. Bergdahl, P. Lodding, H. Lilja, and J. Hugosson, "Prostate cancer screening decreases the absolute risk of being diagnosed with advanced prostate cancer–results from a prospective, population-based randomized controlled trial," *Eur Urol*, vol. 51, pp. 659–64, Mar 2007.

25. R. B. Nadler, P. A. Humphrey, D. S. Smith, W. J. Catalona, and T. L. Ratliff, "Effect of inflammation and benign prostatic hyperplasia on elevated serum prostate specific antigen levels," *J Urol*, vol. 154, pp. 407–13, Aug 1995.

26. D. Ilic, M. M. Neuberger, M. Djulbegovic, and P. Dahm, "Screening for prostate cancer," *Cochrane Database Syst Rev*, vol. 1, p. CD004720, 2013.

27. T. J. Daskivich, K. Chamie, L. Kwan, J. Labo, R. Palvolgyi, A. Dash, S. Greenfield, and M. S. Litwin, "Overtreatment of men with low-risk prostate cancer and significant comorbidity," *Cancer*, vol. 117, pp. 2058–66, May 2011.

28. C. H. Bangma, S. Roemeling, and F. H. Schröder, "Overdiagnosis and overtreatment of early detected prostate cancer," *World J Urol*, vol. 25, pp. 3–9, Mar 2007.

29. G. Draisma, R. Etzioni, A. Tsodikov, A. Mariotto, E. Wever, R. Gulati, E. Feuer, and H. de Koning, "Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context," *J Natl Cancer Inst*, vol. 101, pp. 374–83, Mar 2009.

30. *Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølgning av prostatakreft*. Helsedirektoratet, 2015.

31. E. N. Mottet, J. Bellmunt, *Guidelines on Prostate Cancer*. European Association of Urology, 2015.

32. *Final Recommendation Statement: Prostate Cancer: Screening*. U.S. Preventive Services Task Force, 2016.

33. T. A. Stamey, I. M. Johnstone, J. E. McNeal, A. Y. Lu, and C. M. Yemoto, "Preoperative serum prostate specific antigen levels between 2 and 22 ng./ml. correlate poorly with post-radical prostatectomy cancer morphology: prostate specific antigen cure rates appear constant between 2 and 9 ng./ml," *J Urol*, vol. 167, pp. 103–11, Jan 2002.

34. B. B. McGuire, B. T. Helfand, S. Loeb, Q. Hu, D. O'Brien, P. Cooper, X. Yang, and W. J. Catalona, "Outcomes in patients with gleason score 8-10 prostate cancer: relation to preoperative psa level," *BJU Int*, vol. 109, pp. 1764–9, Jun 2012.

35. D. S. Smith and W. J. Catalona, "Interexaminer variability of digital rectal examination in detecting prostate cancer," *Urology*, vol. 45, pp. 70–4, Jan 1995.

36. O. T. Okotie, K. A. Roehl, M. Han, S. Loeb, S. N. Gashti, and W. J. Catalona, "Characteristics of prostate cancer detected by digital rectal examination only," *Urology*, vol. 70, pp. 1117–20, Dec 2007.

37. E. C. Serefoglu, S. Altinova, N. S. Ugras, E. Akincioglu, E. Asil, and M. D. Balbay, "How reliable is 12-core prostate biopsy procedure in the detection of prostate cancer?," *Can Urol Assoc J*, pp. 1–6, Mar 2012.

38. M. S. Cohen, R. S. Hanley, T. Kurteva, R. Ruthazer, M. L. Silverman, A. Sorcini, K. Hamawy, R. A. Roth, I. Tuerk, and J. A. Libertino, "Comparing the gleason prostate biopsy and gleason prostatectomy grading system: the lahey clinic medical center experience and an international meta-analysis," *Eur Urol*, vol. 54, pp. 371–81, Aug 2008.

39. C. M. Moore, N. L. Robertson, N. Arsanious, T. Middleton, A. Villers, L. Klotz, S. S. Taneja, and M. Emberton, "Image-guided prostate biopsy using magnetic resonance imaging–derived targets: a systematic review," *European urology*, vol. 63, no. 1, pp. 125–140, 2013.

40. D. F. Gleason, "Classification of prostatic carcinomas," *Cancer Chemother Rep*, vol. 50, pp. 125–8, Mar 1966.

41. J. I. Epstein, W. C. Allsbrook, Jr, M. B. Amin, L. L. Egevad, and ISUP Grading Committee, "The 2005 international society of urological pathology (isup) consensus conference on gleason grading of prostatic carcinoma," *Am J Surg Pathol*, vol. 29, pp. 1228–42, Sep 2005.

42. M. Adam, A. Hannah, L. Budäus, T. Steuber, G. Salomon, U. Michl, A. Haese, M. Fisch, C. Wittmer, S. Steurer, S. Minner, H. Heinzer, H. Huland, M. Graefen, G. Sauter, T. Schlomm, and H. Isbarn, "A tertiary gleason pattern in the prostatectomy specimen and its association with adverse outcome after radical prostatectomy," *J Urol*, vol. 192, pp. 97–101, Jul 2014.

43. J. I. Epstein, A. W. Partin, J. Sauvageot, and P. C. Walsh, "Prediction of progression following radical prostatectomy. a multivariate analysis of 721 men with long-term follow-up," *Am J Surg Pathol*, vol. 20, pp. 286–92, Mar 1996.

44. P. C. Albertsen, J. A. Hanley, D. F. Gleason, and M. J. Barry, "Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized prostate cancer," *JAMA*, vol. 280, pp. 975–80, Sep 1998.

45. J. I. Epstein, M. Amin, L. Boccon-Gibod, L. Egevad, P. A. Humphrey, G. Mikuz, D. Newling, S. Nilsson, W. Sakr, J. R. Srigley, T. M. Wheeler, and R. Montironi, "Prognostic factors and reporting of prostate carcinoma in radical prostatectomy and pelvic lymphadenectomy specimens," *Scand J Urol Nephrol Suppl*, pp. 34–63, May 2005.

46. L. Egevad, T. Granfors, L. Karlberg, A. Bergh, and P. Stattin, "Prognostic value of the gleason score in prostate cancer," *BJU Int*, vol. 89, pp. 538–42, Apr 2002.

47. T. Y. Chan, A. W. Partin, P. C. Walsh, and J. I. Epstein, "Prognostic significance of gleason +4 versus gleason score 4+3 tumor at radical prostatectomy," *Urology*, vol. 56, pp. 823–7, Nov 2000.

48. O. Alenda, G. Ploussard, P. Mouracade, E. Xylinas, A. de la Taille, Y. Allory, D. Vordos, A. Hoznek, C. C. Abbou, and L. Salomon, "Impact of the primary gleason pattern on biochemical recurrence-free survival after radical prostatectomy: a single-center cohort of 1,248 patients with gleason 7 tumors," *World J Urol*, vol. 29, pp. 671–6, Oct 2011.

References

49. H. M. Ross, O. N. Kryvenko, J. E. Cowan, J. P. Simko, T. M. Wheeler, and J. I. Epstein, "Do adenocarcinomas of the prostate with gleason score (gs) 6 have the potential to metastasize to lymph nodes?," *Am J Surg Pathol*, vol. 36, pp. 1346–52, Sep 2012.

50. J.-J. Liu, D. Y. Lichtensztajn, S. L. Gomez, W. Sieh, B. I. Chung, I. Cheng, and J. D. Brooks, "Nationwide prevalence of lymph node metastases in gleason score 3+3=6 prostate cancer," *Pathology*, vol. 46, pp. 306–10, Jun 2014.

51. H. B. Carter, A. W. Partin, P. C. Walsh, B. J. Trock, R. W. Veltri, W. G. Nelson, D. S. Coffey, E. A. Singer, and J. I. Epstein, "Gleason score 6 adenocarcinoma: should it be labeled as cancer?," *J Clin Oncol*, vol. 30, pp. 4294–6, Dec 2012.

52. J. C. Nickel and M. Speakman, "Should we really consider gleason 6 prostate cancer?," *BJU Int*, vol. 109, pp. 645–6, Mar 2012.

53. I. Kulac, M. C. Haffner, S. Yegnasubramanian, J. I. Epstein, and A. M. De Marzo, "Should gleason 6 be labeled as cancer?," *Curr Opin Urol*, vol. 25, pp. 238–45, May 2015.

54. P. C. Albertsen, J. A. Hanley, and J. Fine, "20-year outcomes following conservative management of clinically localized prostate cancer," *JAMA*, vol. 293, pp. 2095–101, May 2005.

55. H. U. Ahmed, M. Arya, A. Freeman, and M. Emberton, "Do low-grade and low-volume prostate cancers bear the hallmarks of malignancy?," *Lancet Oncol*, vol. 13, pp. e509–17, Nov 2012.

56. J. I. Epstein, L. Egevad, M. B. Amin, B. Delahunt, J. R. Srigley, P. A. Humphrey, G. Committee, *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," *The American journal of surgical pathology*, vol. 40, no. 2, pp. 244–252, 2016.

57. S. B. Edge, D. R. Byrd, C. C. Compton, A. G. Fritz, F. L. Greene, A. Trotti, *et al.*, *AJCC cancer staging manual*, vol. 649. Springer New York, 2010.

58. J. M. Caster, A. D. Falchook, L. H. Hendrix, and R. C. Chen, "Risk of pathologic upgrading or locally advanced disease in early prostate cancer patients based on biopsy gleason score and psa: A population-based study of modern patients," *Int J Radiat Oncol Biol Phys*, vol. 92, pp. 244–51, Jun 2015.

59. V. Ravery, "The significance of recurrent psa after radical prostatectomy: benign versus malignant sources.," in *Seminars in urologic oncology*, vol. 17, pp. 127–129, 1999.

60. A. A. Caire, L. Sun, O. Ode, D. A. Stackhouse, K. Maloney, C. Donatucci, V. Mouraviev, T. J. Polascik, C. N. Robertson, D. M. Albala, *et al.*, "Delayed prostate-specific antigen recurrence after radical prostatectomy: how to identify and what are their clinical outcomes?," *Urology*, vol. 74, no. 3, pp. 643–647, 2009.

61. S. A. Boorjian, R. H. Thompson, M. K. Tollefson, L. J. Rangel, E. J. Bergstralh, M. L. Blute, and R. J. Karnes, "Long-term risk of clinical progression after biochemical recurrence following radical prostatectomy: the impact of time from surgery to recurrence," *European urology*, vol. 59, no. 6, pp. 893–899, 2011.

62. R. Kalluri, "Emt: when epithelial cells decide to become mesenchymal-like cells," *J Clin Invest*, vol. 119, pp. 1417–9, Jun 2009.

63. M. I. Khan, A. Hamid, V. M. Adhami, R. K. Lall, and H. Mukhtar, "Role of epithelial mesenchymal transition in prostate tumorigenesis," *Current pharmaceutical design*, vol. 21, no. 10, p. 1240, 2015.

64. K. Gravdal, O. J. Halvorsen, S. A. Haukaas, and L. A. Akslen, "A switch from e-cadherin to n-cadherin expression indicates epithelial to mesenchymal transition and is of strong and independent importance for the progress of prostate cancer," *Clin Cancer Res*, vol. 13, pp. 7003–11, Dec 2007.

65. J. Heuberger and W. Birchmeier, "Interplay of cadherin-mediated cell adhesion and canonical wnt signaling," *Cold Spring Harb Perspect Biol*, vol. 2, p. a002915, Feb 2010.

66. T. S. Gujral, M. Chan, L. Peshkin, P. K. Sorger, M. W. Kirschner, and G. MacBeath, "A noncanonical frizzled2 pathway regulates epithelial-mesenchymal transition and metastasis," *Cell*, vol. 159, pp. 844–56, Nov 2014.

67. P. Polakis, "Wnt signaling in cancer," *Cold Spring Harb Perspect Biol*, vol. 4, May 2012.

68. C. Nüsslein-Volhard and E. Wieschaus, "Mutations affecting segment number and polarity in drosophila," *Nature*, vol. 287, pp. 795–801, Oct 1980.

69. R. Nusse and H. E. Varmus, "Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome," *Cell*, vol. 31, pp. 99–109, Nov 1982.

70. F. Rijsewijk, M. Schuermann, E. Wagenaar, P. Parren, D. Weigel, and R. Nusse, "The drosophila homolog of the mouse mammary oncogene int-1 is identical to the segment polarity gene wingless," *Cell*, vol. 50, pp. 649–57, Aug 1987.

71. C. Y. Logan and R. Nusse, "The wnt signaling pathway in development and disease," *Annu Rev Cell Dev Biol*, vol. 20, pp. 781–810, 2004.

72. S. M. Powell, N. Zilz, Y. Beazer-Barclay, T. M. Bryan, S. R. Hamilton, S. N. Thibodeau, B. Vogelstein, and K. W. Kinzler, "Apc mutations occur early during colorectal tumorigenesis," *Nature*, vol. 359, pp. 235–7, Sep 1992.

73. V. Korinek, N. Barker, P. J. Morin, D. van Wichen, R. de Weger, K. W. Kinzler, B. Vogelstein, and H. Clevers, "Constitutive transcriptional activation by a beta-catenin-tcf complex in apc-/- colon carcinoma," *Science*, vol. 275, pp. 1784–7, Mar 1997.

74. I. Nishisho, Y. Nakamura, Y. Miyoshi, Y. Miki, H. Ando, A. Horii, K. Koyama, J. Utsunomiya, S. Baba, and P. Hedge, "Mutations of chromosome 5q21 genes in fap and colorectal cancer patients," *Science*, vol. 253, pp. 665–9, Aug 1991.

75. M. Watanabe, H. Kakiuchi, H. Kato, T. Shiraishi, R. Yatani, T. Sugimura, and M. Nagao, "Apc gene mutations in human prostate cancer," *Jpn J Clin Oncol*, vol. 26, pp. 77–81, Apr 1996.

76. I. Bisson and D. M. Prowse, "Wnt signaling regulates self-renewal and differentiation of prostate cancer cells with stem cell characteristics," *Cell Res*, vol. 19, pp. 683–97, Jun 2009.

77. M. Verras, J. Brown, X. Li, R. Nusse, and Z. Sun, "Wnt3a growth factor induces androgen receptor-mediated transcription and enhances cell growth in human prostate cancer cells," *Cancer Res*, vol. 64, pp. 8860–6, Dec 2004.

78. X. Wan, J. Liu, J.-F. Lu, V. Tzelepi, J. Yang, M. W. Starbuck, L. Diao, J. Wang, E. Efstathiou, E. S. Vazquez, P. Troncoso, S. N. Maity, and N. M. Navone, "Activation of β-catenin signaling in androgen receptor-negative prostate cancer cells," *Clin Cancer Res*, vol. 18, pp. 726–36, Feb 2012.

79. G. Chen, N. Shukeir, A. Potti, K. Sircar, A. Aprikian, D. Goltzman, and S. A. Rabbani, "Up-regulation of wnt-1 and beta-catenin production in patients with advanced metastatic prostate carcinoma: potential pathogenetic and prognostic implications," *Cancer*, vol. 101, pp. 1345–56, Sep 2004.

80. T. A. Bismar, P. A. Humphrey, D. J. Grignon, and H. L. Wang, "Expression of beta-catenin in prostatic adenocarcinomas: a comparison with colorectal adenocarcinomas," *Am J Clin Pathol*, vol. 121, pp. 557–63, Apr 2004.

81. Q. Wang, A. J. Symes, C. A. Kane, A. Freeman, J. Nariculam, P. Munson, C. Thrasivoulou, J. R. W. Masters, and A. Ahmed, "A novel role for wnt/ca2+ signaling in actin cytoskeleton remodeling and cell motility in prostate cancer," *PLoS One*, vol. 5, no. 5, p. e10456, 2010.

82. H. Yamamoto, N. Oue, A. Sato, Y. Hasegawa, H. Yamamoto, A. Matsubara, W. Yasui, and A. Kikuchi, "Wnt5a signaling is involved in the aggressiveness of prostate cancer and expression of metalloproteinase," *Oncogene*, vol. 29, pp. 2036–46, Apr 2010.

References

83. A. S. Syed Khaja, L. Helczynski, A. Edsjö, R. Ehrnström, A. Lindgren, D. Ulmert, T. Andersson, and A. Bjartell, "Elevated level of wnt5a protein in localized prostate cancer tissue is associated with better outcome," *PLoS One*, vol. 6, no. 10, p. e26539, 2011.

84. S. Thiele, A. Göbel, T. D. Rachner, S. Fuessel, M. Froehner, M. H. Muders, G. B. Baretton, R. Bernhardt, F. Jakob, C. C. Glüer, M. Bornhäuser, M. Rauner, and L. C. Hofbauer, "Wnt5a has anti-prostate cancer effects in vitro and reduces tumor growth in the skeleton in vivo," *J Bone Miner Res*, vol. 30, pp. 471–80, Mar 2015.

85. A. S. S. Khaja, L. Egevad, L. Helczynski, P. Wiklund, T. Andersson, and A. Bjartell, "Emphasizing the role of wnt5a protein expression to predict favorable outcome after radical prostatectomy in patients with low-grade prostate cancer," *Cancer Med*, vol. 1, pp. 96–104, Aug 2012.

86. E. Sandsmark, A. F. Hansen, K. M. Selnæs, H. Bertilsson, A. M. Bofin, A. J. Wright, T. Viset, E. Richardsen, F. Drabløs, T. F. Bathen, M.-B. Tessem, and M. B. Rye, "A novel non-canonical wnt signature for prostate cancer aggressiveness," *Oncotarget*, vol. 8, pp. 9572–9586, Feb 2017.

87. B. Hoang, M. Moos, S. Vukicevic, and F. P. Luyten, "Primary structure and tissue distribution of frzb, a novel protein related to drosophila frizzled, suggest a role in skeletal morphogenesis," *Journal of Biological Chemistry*, vol. 271, no. 42, pp. 26131–26137, 1996.

88. P. Bovolenta, P. Esteve, J. M. Ruiz, E. Cisneros, and J. Lopez-Rios, "Beyond wnt inhibition: new functions of secreted frizzled-related proteins in development and disease," *J Cell Sci*, vol. 121, no. 6, pp. 737–746, 2008.

89. A. Üren, F. Reichsman, V. Anest, W. G. Taylor, K. Muraiso, D. P. Bottaro, S. Cumberledge, and J. S. Rubin, "Secreted frizzled-related protein-1 binds directly to wingless and is a biphasic modulator of wnt signaling," *Journal of Biological Chemistry*, vol. 275, no. 6, pp. 4374–4382, 2000.

90. S. E. Jones and C. Jomary, "Secreted frizzled-related proteins: searching for relationships and patterns," *Bioessays*, vol. 24, no. 9, pp. 811–820, 2002.

91. S. Pohl, R. Scott, F. Arfuso, V. Perumal, and A. Dharmarajan, "Secreted frizzled-related protein 4 and its implications in cancer and apoptosis," *Tumor Biology*, vol. 36, no. 1, pp. 143–152, 2015.

92. J.-H. Luo, Y. P. Yu, K. Cieply, F. Lin, P. Deflavia, R. Dhir, S. Finkelstein, G. Michalopoulos, and M. Becich, "Gene expression analysis of prostate cancers," *Molecular carcinogenesis*, vol. 33, no. 1, pp. 25–35, 2002.

93. C. Wissmann, P. J. Wild, S. Kaiser, S. Roepcke, R. Stoehr, M. Woenckhaus, G. Kristiansen, J.-C. Hsieh, F. Hofstaedter, A. Hartmann, *et al.*, "Wif1, a component of the wnt pathway, is down-regulated in prostate, breast, lung, and bladder cancer," *The Journal of pathology*, vol. 201, no. 2, pp. 204–212, 2003.

94. M. M. Mortensen, S. Høyer, A.-S. Lynnerup, T. F. Ørntoft, K. D. Sørensen, M. Borre, and L. Dyrskjøt, "Expression profiling of prostate cancer tissue delineates genes associated with recurrence after prostatectomy," *Scientific reports*, vol. 5, 2015.

95. L. G. Horvath, S. M. Henshall, J. G. Kench, D. N. Saunders, C.-S. Lee, D. Golovsky, P. C. Brenner, G. F. O'Neill, R. Kooner, P. D. Stricker, *et al.*, "Membranous expression of secreted frizzled-related protein 4 predicts for good prognosis in localized prostate cancer and inhibits pc3 cellular proliferation in vitro," *Clinical cancer research*, vol. 10, no. 2, pp. 615–625, 2004.

96. L. G. Horvath, J. E. Lelliott, J. G. Kench, C. Lee, E. D. Williams, D. N. Saunders, J. J. Grygiel, R. L. Sutherland, S. M. Henshall, *et al.*, "Secreted frizzled-related protein 4 inhibits proliferation and metastatic potential in prostate cancer," *The Prostate*, vol. 67, no. 10, pp. 1081–1090, 2007.

97. S. A. Tomlins, D. R. Rhodes, S. Perner, S. M. Dhanasekaran, R. Mehra, X.-W. Sun, S. Varambally, X. Cao, J. Tchinda, R. Kuefer, C. Lee, J. E. Montie, R. B. Shah, K. J. Pienta, M. A. Rubin, and A. M. Chinnaiyan, "Recurrent fusion of tmprss2 and ets transcription factor genes in prostate cancer," *Science*, vol. 310, pp. 644–8, Oct 2005.

98. B. Lin, C. Ferguson, J. T. White, S. Wang, R. Vessella, L. D. True, L. Hood, and P. S. Nelson, "Prostate-localized and androgen-regulated expression of the membrane-bound serine protease tmprss2," *Cancer Res*, vol. 59, pp. 4180–4, Sep 1999.

99. M. Miettinen, Z.-F. Wang, A. Paetau, S.-H. Tan, A. Dobi, S. Srivastava, and I. Sesterhenn, "Erg transcription factor as an immunohistochemical marker for vascular endothelial tumors and prostatic carcinoma," *Am J Surg Pathol*, vol. 35, pp. 432–41, Mar 2011.

100. I. G. Maroulakou and D. B. Bowe, "Expression and function of ets transcription factors in mammalian development: a regulatory network," *Oncogene*, vol. 19, pp. 6432–42, Dec 2000.

101. F. McLaughlin, V. J. Ludbrook, J. Cox, I. von Carlowitz, S. Brown, and A. M. Randi, "Combined genomic and antisense analysis reveals that the transcription factor erg is implicated in endothelial cell differentiation," *Blood*, vol. 98, pp. 3332–9, Dec 2001.

102. P. Adamo and M. R. Ladomery, "The oncogene erg: a key factor in prostate cancer," *Oncogene*, Apr 2015.

103. S. Gupta, K. Iljin, H. Sara, J. P. Mpindi, T. Mirtti, P. Vainio, J. Rantala, K. Alanen, M. Nees, and O. Kallioniemi, "Fzd4 as a mediator of erg oncogene-induced wnt signaling and epithelial-to-mesenchymal transition in human prostate cancer cells," *Cancer Res*, vol. 70, pp. 6735–45, Sep 2010.

104. E. K. Markert, H. Mizuno, A. Vazquez, and A. J. Levine, "Molecular classification of prostate cancer using curated expression signatures," *Proc Natl Acad Sci U S A*, vol. 108, pp. 21276–81, Dec 2011.

105. M. B. Rye, H. Bertilsson, F. Drabløs, A. Angelsen, T. F. Bathen, and M.-B. Tessem, "Gene signatures esc, myc and erg-fusion are early markers of a potentially dangerous subtype of prostate cancer," *BMC Med Genomics*, vol. 7, p. 50, 2014.

106. R. K. Nam, L. Sugar, Z. Wang, W. Yang, R. Kitching, L. H. Klotz, V. Venkateswaran, S. A. Narod, and A. Seth, "Expression of tmprss2:erg gene fusion in prostate cancer cells is an important prognostic factor for cancer progression," *Cancer Biol Ther*, vol. 6, pp. 40–5, Jan 2007.

107. R. K. Nam, L. Sugar, W. Yang, S. Srivastava, L. H. Klotz, L.-Y. Yang, A. Stanimirovic, E. Encioiu, M. Neill, D. A. Loblaw, J. Trachtenberg, S. A. Narod, and A. Seth, "Expression of the tmprss2:erg fusion gene predicts cancer recurrence after surgery for localised prostate cancer," *Br J Cancer*, vol. 97, pp. 1690–5, Dec 2007.

108. G. Attard, J. Clark, L. Ambroisine, G. Fisher, G. Kovacs, P. Flohr, D. Berney, C. S. Foster, A. Fletcher, W. L. Gerald, H. Moller, V. Reuter, J. S. De Bono, P. Scardino, J. Cuzick, C. S. Cooper, and Transatlantic Prostate Group, "Duplication of the fusion of tmprss2 to erg sequences identifies fatal human prostate cancer," *Oncogene*, vol. 27, pp. 253–63, Jan 2008.

109. A. Pettersson, R. E. Graff, S. R. Bauer, M. J. Pitt, R. T. Lis, E. C. Stack, N. E. Martin, L. Kunz, K. L. Penney, A. H. Ligon, *et al.*, "The tmprss2: Erg rearrangement, erg expression, and prostate cancer outcomes: a cohort study and meta-analysis," *Cancer Epidemiology Biomarkers & Prevention*, vol. 21, no. 9, pp. 1497–1509, 2012.

110. S. Meller, H.-A. Meyer, B. Bethan, D. Dietrich, S. G. Maldonado, M. Lein, M. Montani, R. Reszka, P. Schatz, E. Peter, *et al.*, "Integration of tissue metabolomics, transcriptomics and immunohistochemistry reveals erg-and gleason score-specific metabolomic alterations in prostate cancer," *Oncotarget*, vol. 7, no. 2, p. 1421, 2016.

111. P. Massoner, K. G. Kugler, K. Unterberger, R. Kuner, L. A. Mueller, M. Fälth, G. Schäfer, C. Seifarth, S. Ecker, I. Verdorfer, *et al.*, "Characterization of transcriptional changes in erg rearrangement-positive prostate cancer identifies the regulation of metabolic sensors such as neuropeptide y," *PloS one*, vol. 8, no. 2, p. e55207, 2013.

112. D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: the next generation," *Cell*, vol. 144, pp. 646–74, Mar 2011.

113. R. A. Cairns, I. S. Harris, and T. W. Mak, "Regulation of cancer cell metabolism," *Nature Reviews Cancer*, vol. 11, no. 2, pp. 85–95, 2011.

References

114. R. A. Cairns and T. W. Mak, "The current state of cancer metabolism," *Nature Reviews Cancer*, vol. 16, no. 10, pp. 613–614, 2016.

115. N. E. Warburg O, Posener K, *Über den Stoffwechsel der Carcinomzelle*, vol. 152, 309-344. Biochem. Zeitschr., 1926.

116. O. WARBURG, "On the origin of cancer cells," *Science*, vol. 123, pp. 309–14, Feb 1956.

117. J.-Q. Chen and J. Russo, "Dysregulation of glucose transport, glycolysis, tca cycle and glutaminolysis by oncogenes and tumor suppressors in cancer cells," *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, vol. 1826, no. 2, pp. 370–384, 2012.

118. A. Kinnaird, S. Zhao, K. E. Wellen, and E. D. Michelakis, "Metabolic control of epigenetics in cancer," *Nature Reviews Cancer*, 2016.

119. L. C. Costello and R. B. Franklin, "A comprehensive review of the role of zinc in normal prostate function and metabolism; and its implications in prostate cancer," *Archives of biochemistry and biophysics*, vol. 611, pp. 100–112, 2016.

120. A. R. Lima, M. de Lourdes Bastos, M. Carvalho, and P. G. de Pinho, "Biomarker discovery in human prostate cancer: an update in metabolomics studies," *Translational Oncology*, vol. 9, no. 4, pp. 357–370, 2016.

121. X. Wu, G. Daniels, P. Lee, and M. E. Monaco, "Lipid metabolism in prostate cancer," *American journal of clinical and experimental urology*, vol. 2, no. 2, p. 111, 2014.

122. M.-B. Tessem, M. G. Swanson, K. R. Keshari, M. J. Albers, D. Joun, Z. L. Tabatabai, J. P. Simko, K. Shinohara, S. J. Nelson, D. B. Vigneron, *et al.*, "Evaluation of lactate and alanine as metabolic biomarkers of prostate cancer using 1h hr-mas spectroscopy of biopsy tissues," *Magnetic resonance in medicine*, vol. 60, no. 3, pp. 510–516, 2008.

123. K. Glunde, M. A. Jacobs, and Z. M. Bhujwalla, "Choline metabolism in cancer: implications for diagnosis and therapy," *Expert review of molecular diagnostics*, vol. 6, no. 6, pp. 821–829, 2006.

124. H. Bertilsson, M.-B. Tessem, A. Flatberg, T. Viset, I. Gribbestad, A. Angelsen, and J. Halgunset, "Changes in gene transcription underlying the aberrant citrate and choline metabolism in human prostate cancer samples," *Clin Cancer Res*, vol. 18, pp. 3261–9, Jun 2012.

125. H. M. Awwad, J. Geisel, and R. Obeid, "The role of choline in prostate cancer," *Clinical biochemistry*, vol. 45, no. 18, pp. 1548–1553, 2012.

126. G. F. Giskeødegård, H. Bertilsson, K. M. Selnæs, A. J. Wright, T. F. Bathen, T. Viset, J. Halgunset, A. Angelsen, I. S. Gribbestad, and M.-B. Tessem, "Spermine and citrate as metabolic biomarkers for assessing prostate cancer aggressiveness," *PLoS One*, vol. 8, no. 4, p. e62375, 2013.

127. E. A. Decelle and L. L. Cheng, "High-resolution magic angle spinning 1h mrs in prostate cancer," *NMR Biomed*, vol. 27, pp. 90–9, Jan 2014.

128. S. Shah, W. J. Carriveau, J. Li, S. L. Campbell, P. K. Kopinski, H.-W. Lim, N. Daurio, S. Trefely, K.-J. Won, D. C. Wallace, *et al.*, "Targeting acly sensitizes castration-resistant prostate cancer cells to ar antagonism by impinging on an acly-ampk-ar feedback mechanism," *Oncotarget*, vol. 7, no. 28, p. 43713, 2016.

129. J. Kurhanewicz, D. B. Vigneron, S. J. Nelson, H. Hricak, J. M. MacDonald, B. Konety, and P. Narayan, "Citrate as an in vivo marker to discriminate prostate cancer from benign prostatic hyperplasia and normal prostate peripheral zone: detection via localized proton spectroscopy," *Urology*, vol. 45, pp. 459–66, Mar 1995.

130. S. A. Reinsberg, G. S. Payne, S. F. Riches, S. Ashley, J. M. Brewster, V. A. Morgan, and N. M. Desouza, "Combined use of diffusion-weighted mri and 1h mr spectroscopy to increase accuracy in prostate cancer detection," *American Journal of Roentgenology*, vol. 188, no. 1, pp. 91–98, 2007.

131. M. G. Swanson, A. S. Zektzer, Z. L. Tabatabai, J. Simko, S. Jarso, K. R. Keshari, L. Schmitt, P. R. Carroll, K. Shinohara, D. B. Vigneron, and J. Kurhanewicz, "Quantitative analysis of prostate metabolites using 1h hr-mas spectroscopy," *Magn Reson Med*, vol. 55, pp. 1257–64, Jun 2006.

132. R. Dittrich, J. Kurth, E. A. Decelle, E. M. DeFeo, M. Taupitz, S. Wu, C.-L. Wu, W. S. McDougal, and L. L. Cheng, "Assessing prostate cancer growth with citrate measured by intact tissue proton magnetic resonance spectroscopy," *Prostate Cancer Prostatic Dis*, vol. 15, pp. 278–82, Sep 2012.

133. J. J. A. van Asten, V. Cuijpers, C. Hulsbergen-van de Kaa, C. Soede-Huijbregts, J. A. Witjes, A. Verhofstad, and A. Heerschap, "High resolution magic angle spinning nmr spectroscopy for metabolic assessment of cancer presence and gleason score in human prostate needle biopsies," *MAGMA*, vol. 21, pp. 435–42, Nov 2008.

134. S. L. Nowotarski, P. M. Woster, and R. A. Casero, Jr, "Polyamines and cancer: implications for chemotherapy and chemoprevention," *Expert Rev Mol Med*, vol. 15, p. e3, 2013.

135. A. M. Betts, I. Waite, D. E. Neal, and C. N. Robson, "Androgen regulation of ornithine decarboxylase in human prostatic cells identified using differential display," *FEBS Lett*, vol. 405, pp. 328–32, Apr 1997.

136. H. E. Fjösne, M. A. Ostensen, H. Haarstad, and A. Sunde, "Androgen regulation of polyamine synthesis in seminal vesicle and in different lobes of the rat prostate," *Prostate*, vol. 17, no. 1, pp. 1–11, 1990.

137. M. Auvinen, A. Paasinen, L. C. Andersson, and E. Höltta, "Ornithine decarboxylase activity is critical for cell transformation," *Nature*, vol. 360, pp. 355–8, Nov 1992.

138. R. R. Mohan, A. Challa, S. Gupta, D. G. Bostwick, N. Ahmad, R. Agarwal, S. R. Marengo, S. B. Amini, F. Paras, G. T. MacLennan, M. I. Resnick, and H. Mukhtar, "Overexpression of ornithine decarboxylase in prostate cancer and prostatic fluid in humans," *Clin Cancer Res*, vol. 5, pp. 143–7, Jan 1999.

139. S. Bettuzzi, P. Davalli, S. Astancolle, C. Carani, B. Madeo, A. Tampieri, A. Corti, B. Saverio, D. Pierpaola, A. Serenella, C. Cesare, M. Bruno, T. Auro, and C. Arnaldo, "Tumor progression is accompanied by significant changes in the levels of expression of polyamine metabolism regulatory genes and clusterin (sulfated glycoprotein 2) in human prostate cancer specimens," *Cancer Res*, vol. 60, pp. 28–34, Jan 2000.

140. R. C. Smith, M. S. Litwin, Y. Lu, and B. R. Zetter, "Identification of an endogenous inhibitor of prostatic carcinoma cell growth," *Nat Med*, vol. 1, pp. 1040–5, Oct 1995.

141. A. C. Goodwin, S. Jadallah, A. Toubaji, K. Lecksell, J. L. Hicks, J. Kowalski, G. S. Bova, A. M. De Marzo, G. J. Netto, and R. A. Casero, Jr, "Increased spermine oxidase expression in human prostate cancer and prostatic intraepithelial neoplasia tissues," *Prostate*, vol. 68, pp. 766–72, May 2008.

142. M. van der Graaf, R. G. Schipper, G. O. Oosterhof, J. A. Schalken, A. A. Verhofstad, and A. Heerschap, "Proton mr spectroscopy of prostatic tissue focused on the detection of spermine, a possible biomarker of malignant behavior in prostate cancer," *MAGMA*, vol. 10, pp. 153–9, Jul 2000.

143. L. L. Cheng, C.-l. Wu, M. R. Smith, and R. G. Gonzalez, "Non-destructive quantitation of spermine in human prostate tissue samples using hrmas 1h nmr spectroscopy at 9.4 t," *FEBS letters*, vol. 494, no. 1-2, pp. 112–116, 2001.

144. A. Maxeiner, C. B. Adkins, Y. Zhang, M. Taupitz, E. F. Halpern, W. S. McDougal, C.-L. Wu, and L. L. Cheng, "Retrospective analysis of prostate cancer recurrence potential with tissue metabolomic profiles," *The Prostate*, vol. 70, no. 7, pp. 710–717, 2010.

145. V. E. Velculescu, L. Zhang, W. Zhou, J. Vogelstein, M. A. Basrai, D. E. Bassett, Jr, P. Hieter, B. Vogelstein, and K. W. Kinzler, "Characterization of the yeast transcriptome," *Cell*, vol. 88, pp. 243–51, Jan 1997.

146. E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin ..., J. Szustakowki, and International Human Genome Sequencing Consortium, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, pp. 860–921, Feb 2001.

147. J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton ..., and X. Zhu, "The sequence of the human genome," *Science*, vol. 291, pp. 1304–51, Feb 2001.

References

148.   A. L. Tarca, R. Romero, and S. Draghici, "Analysis of microarray experiments of gene expression profiling," *American journal of obstetrics and gynecology*, vol. 195, no. 2, pp. 373–388, 2006.

149.   C. M. Perou, T. Sørlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, *et al.*, "Molecular portraits of human breast tumours," *Nature*, vol. 406, no. 6797, pp. 747–752, 2000.

150.   T. Sørlie, C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. Van De Rijn, S. S. Jeffrey, *et al.*, "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications," *Proceedings of the National Academy of Sciences*, vol. 98, no. 19, pp. 10869–10874, 2001.

151.   D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, C. Ladd, P. Tamayo, A. A. Renshaw, A. V. D'Amico, J. P. Richie, E. S. Lander, M. Loda, P. W. Kantoff, T. R. Golub, and W. R. Sellers, "Gene expression correlates of clinical prostate cancer behavior," *Cancer Cell*, vol. 1, pp. 203–9, Mar 2002.

152.   J. Cuzick, G. P. Swanson, G. Fisher, A. R. Brothman, D. M. Berney, J. E. Reid, D. Mesher, V. Speights, E. Stankiewicz, C. S. Foster, *et al.*, "Prognostic value of an rna expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study," *The lancet oncology*, vol. 12, no. 3, pp. 245–255, 2011.

153.   E. A. Klein, M. R. Cooperberg, C. Magi-Galluzzi, J. P. Simko, S. M. Falzarano, T. Maddala, J. M. Chan, J. Li, J. E. Cowan, A. C. Tsiatis, *et al.*, "A 17-gene assay to predict prostate cancer aggressiveness in the context of gleason grade heterogeneity, tumor multifocality, and biopsy undersampling," *European urology*, vol. 66, no. 3, pp. 550–560, 2014.

154.   N. Erho, A. Crisan, I. A. Vergara, A. P. Mitra, M. Ghadessi, C. Buerki, E. J. Bergstralh, T. Kollmeyer, S. Fink, Z. Haddad, *et al.*, "Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy," *PloS one*, vol. 8, no. 6, p. e66855, 2013.

155.   A. Schroeder, O. Mueller, S. Stocker, R. Salowsky, M. Leiber, M. Gassmann, S. Lightfoot, W. Menzel, M. Granzow, and T. Ragg, "The rin: an rna integrity number for assigning integrity values to rna measurements," *BMC Mol Biol*, vol. 7, p. 3, 2006.

156.   S. C. Schuster, "Next-generation sequencing transforms today's biology," *Nature methods*, vol. 5, no. 1, p. 16, 2008.

157.   Z. Wang, M. Gerstein, and M. Snyder, "Rna-seq: a revolutionary tool for transcriptomics," *Nature reviews genetics*, vol. 10, no. 1, pp. 57–63, 2009.

158.   F. Ozsolak and P. M. Milos, "Rna sequencing: advances, challenges and opportunities," *Nature reviews genetics*, vol. 12, no. 2, pp. 87–98, 2011.

159.   S. A. Byron, K. R. Van Keuren-Jensen, D. M. Engelthaler, J. D. Carpten, and D. W. Craig, "Translating rna sequencing into clinical diagnostics: opportunities and challenges," *Nature Reviews Genetics*, 2016.

160.   A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, *et al.*, "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.

161.   D. A. Barbie, P. Tamayo, J. S. Boehm, S. Y. Kim, S. E. Moody, I. F. Dunn, A. C. Schinzel, P. Sandy, E. Meylan, C. Scholl, *et al.*, "Systematic rna interference reveals that oncogenic kras-driven cancers require tbk1," *Nature*, vol. 462, no. 7269, pp. 108–112, 2009.

162.   J. Ramos-Vara and M. Miller, "When tissue antigens and antibodies get along: revisiting the technical aspects of immunohistochemistry—the red, brown, and blue technique," *Veterinary pathology*, vol. 51, no. 1, pp. 42–87, 2014.

163. A. H. Coons, H. J. Creech, and R. N. Jones, "Immunological properties of an antibody containing a fluorescent group.," *Proceedings of the Society for Experimental Biology and Medicine*, vol. 47, no. 2, pp. 200–202, 1941.

164. J. Ramos-Vara, "Technical aspects of immunohistochemistry," *Veterinary Pathology Online*, vol. 42, no. 4, pp. 405–426, 2005.

165. M. Kobayashi, T. Honma, Y. Matsuda, Y. Suzuki, R. Narisawa, Y. Ajioka, and H. Asakura, "Nuclear translocation of beta-catenin in colorectal cancer," *British journal of cancer*, vol. 82, no. 10, p. 1689, 2000.

166. P. R. Langer-Safer, M. Levine, and D. C. Ward, "Immunological method for mapping genes on drosophila polytene chromosomes," *Proceedings of the National Academy of Sciences*, vol. 79, no. 14, pp. 4381–4385, 1982.

167. J. M. Levsky and R. H. Singer, "Fluorescence in situ hybridization: past, present and future," *Journal of cell science*, vol. 116, no. 14, pp. 2833–2838, 2003.

168. J. L. Griffin and J. P. Shockcor, "Metabolic profiles of cancer cells," *Nat Rev Cancer*, vol. 4, pp. 551–61, Jul 2004.

169. G. A. N. Gowda, S. Zhang, H. Gu, V. Asiago, N. Shanaiah, and D. Raftery, "Metabolomics-based methods for early disease diagnostics," *Expert Rev Mol Diagn*, vol. 8, pp. 617–33, Sep 2008.

170. I. I. Rabi, "Space quantization in a gyrating magnetic field," *Physical Review*, vol. 51, no. 8, p. 652, 1937.

171. E. M. Purcell, H. Torrey, and R. V. Pound, "Resonance absorption by nuclear magnetic moments in a solid," *Physical review*, vol. 69, no. 1-2, p. 37, 1946.

172. F. Bloch, "Nuclear induction," *Physical review*, vol. 70, no. 7-8, p. 460, 1946.

173. E. Andrew, A. Bradbury, and R. Eades, "Nuclear magnetic resonance spectra from a crystal rotated at high speed," *Nature*, vol. 182, p. 1659, Dec 1958.

174. I. Lowe, "Free induction decays of rotating solids," *Phys. Rev. Lett.*, vol. 2, April 1959.

175. L. L. Cheng, C. L. Lean, A. Bogdanova, S. C. Wright, J. L. Ackerman, T. J. Brady, and L. Garrido, "Enhanced resolution of proton nmr spectra of malignant lymph nodes using magic-angle spinning," *Magnetic resonance in medicine*, vol. 36, no. 5, pp. 653–658, 1996.

176. S. Moestue, B. Sitter, T. Frost Bathen, M.-B. Tessem, and I. Susann Gribbestad, "Hr mas mr spectroscopy in metabolic characterization of cancer," *Current topics in medicinal chemistry*, vol. 11, no. 1, pp. 2–26, 2011.

177. H. Bertilsson, A. Angelsen, T. Viset, H. Skogseth, M.-B. Tessem, and J. Halgunset, "A new method to provide a fresh frozen prostate slice suitable for gene expression study and mr spectroscopy," *The Prostate*, vol. 71, no. 5, pp. 461–469, 2011.

178. J. T. Bjerrum and Bjerrum, *Metabonomics*. Springer, 2015.

179. G. F. Giskeødegård, M. D. Cao, and T. F. Bathen, "High-resolution magic-angle-spinning nmr spectroscopy of intact tissue," *Metabonomics: Methods and Protocols*, pp. 37–50, 2015.

180. T. R. Brown, B. M. Kincaid, and K. Ugurbil, "Nmr chemical shift imaging in three dimensions," *Proc Natl Acad Sci U S A*, vol. 79, pp. 3523–6, Jun 1982.

181. M. van der Graaf, "In vivo magnetic resonance spectroscopy: basic methodology and clinical applications," *Eur Biophys J*, vol. 39, pp. 527–40, Mar 2010.

182. T. Kobus, A. J. Wright, T. W. J. Scheenen, and A. Heerschap, "Mapping of prostate cancer by 1h mrsi," *NMR Biomed*, vol. 27, pp. 39–52, Jan 2014.

183. S. W. Provencher, "Estimation of metabolite concentrations from localized in vivo proton nmr spectra," *Magn Reson Med*, vol. 30, pp. 672–9, Dec 1993.

184. S. W. Provencher, "Automatic quantitation of localized in vivo 1h spectra with lcmodel," *NMR Biomed*, vol. 14, pp. 260–4, Jun 2001.

185. K. S. Opstad, B. A. Bell, J. R. Griffiths, and F. A. Howe, "Toward accurate quantification of metabolites, lipids, and macromolecules in hrmas spectra of human brain tumor biopsies using lcmodel," *Magn Reson Med*, vol. 60, pp. 1237–42, Nov 2008.

References

186. O. N. Keene, "The log transformation is special," *Statistics in medicine*, vol. 14, no. 8, pp. 811–819, 1995.

187. N. M. Laird and J. H. Ware, "Random-effects models for longitudinal data," *Biometrics*, pp. 963–974, 1982.

188. J. Jiang, *Linear and generalized linear mixed models and their applications*. Springer Science & Business Media, 2007.

189. S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1, pp. 37–52, 1987.

190. S. Wold, P. Geladi, K. Esbensen, and J. Öhman, "Multi-way principal components-and pls-analysis," *Journal of chemometrics*, vol. 1, no. 1, pp. 41–56, 1987.

191. J. A. Westerhuis, H. C. Hoefsloot, S. Smit, D. J. Vis, A. K. Smilde, E. J. van Velzen, J. P. van Duijnhoven, and F. A. van Dorsten, "Assessment of plsda cross validation," *Metabolomics*, vol. 4, no. 1, pp. 81–89, 2008.

192. B. Rosner, *Fundamentals of Biostatistics*. Brooks/Cole Cengage Learning, 7th ed., 2011.

193. E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American statistical association*, vol. 53, no. 282, pp. 457–481, 1958.

194. N. Mantel, "Evaluation of survival data and two new rank order statistics arising in its consideration.," *Cancer chemotherapy reports. Part 1*, vol. 50, no. 3, pp. 163–170, 1966.

195. D. Cox, "Regression models and life tables," *Journal of the Royal Statistical Society*, vol. 34, pp. 187–220, 1972.

196. M. Egger, G. D. Smith, and A. N. Phillips, "Meta-analysis: principles and procedures.," *BMJ: British Medical Journal*, vol. 315, no. 7121, p. 1533, 1997.

197. J. Cohen, "Statistical power analysis for the behavioral sciences lawrence earlbaum associates," *Hillsdale, NJ*, pp. 20–26, 1988.

198. S. S. Sawilowsky, "New effect size rules of thumb," *Journal of Modern Applied Statistical Methods*, vol. 8, no. 26, 2009.

199. M. R. Crager, "Generalizing the standardized hazard ratio to multivariate proportional hazards regression, with an application to clinical˜ genomic studies," *Journal of Applied Statistics*, vol. 39, no. 2, pp. 399–417, 2012.

200. Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 289–300, 1995.

201. C. G. A. R. Network *et al.*, "The molecular taxonomy of primary prostate cancer," *Cell*, vol. 163, no. 4, pp. 1011–1025, 2015.

202. H. Ross-Adams, A. Lamb, M. Dunning, S. Halim, J. Lindberg, C. Massie, L. Egevad, R. Russell, A. Ramos-Montoya, S. Vowler, *et al.*, "Integration of copy number and transcriptomics provides risk stratification in prostate cancer: a discovery and validation cohort study," *EBioMedicine*, vol. 2, no. 9, pp. 1133–1144, 2015.

203. Y. Wang, X.-Q. Xia, Z. Jia, A. Sawyers, H. Yao, J. Wang-Rodriquez, D. Mercola, and M. McClelland, "In silico estimates of tissue components in surgical samples based on expression profiling data," *Cancer research*, vol. 70, no. 16, pp. 6448–6455, 2010.

204. Z. Jia, Y. Wang, A. Sawyers, H. Yao, F. Rahmatpanah, X.-Q. Xia, Q. Xu, R. Pio, T. Turan, J. A. Koziol, *et al.*, "Diagnosis of prostate cancer using differentially expressed genes in stroma," *Cancer research*, vol. 71, no. 7, pp. 2476–2487, 2011.

205. X. Chen, S. Xu, M. McClelland, F. Rahmatpanah, A. Sawyers, Z. Jia, and D. Mercola, "An accurate prostate cancer prognosticator using a seven-gene signature plus gleason score and taking cell type heterogeneity into account," *PloS one*, vol. 7, no. 9, p. e45178, 2012.

206. A. Sboner, F. Demichelis, S. Calza, Y. Pawitan, S. R. Setlur, Y. Hoshida, S. Perner, H.-O. Adami, K. Fall, L. A. Mucci, *et al.*, "Molecular sampling of prostate cancer: a dilemma for predicting disease progression," *BMC medical genomics*, vol. 3, no. 1, p. 8, 2010.

207. B. S. Taylor, N. Schultz, H. Hieronymus, A. Gopalan, Y. Xiao, B. S. Carver, V. K. Arora, P. Kaushik, E. Cerami, B. Reva, *et al.*, "Integrative genomic profiling of human prostate cancer," *Cancer cell*, vol. 18, no. 1, pp. 11–22, 2010.

208. S. G. Zhao, J. R. Evans, V. Kothari, G. Sun, A. Larm, V. Mondine, E. M. Schaeffer, A. E. Ross, E. A. Klein, R. B. Den, *et al.*, "The landscape of prognostic outlier genes in high-risk prostate cancer," *Clinical Cancer Research*, vol. 22, no. 7, pp. 1777–1786, 2016.

209. H. Bertilsson, *Prostate Cancer- Translational Research. Optimizing tissue sampling suitable for histopathologic, transcriptomic and metabolic profiling [PhD thesis].* PhD thesis, Trondheim: Norwegian University of Science and Technology, SBN 978–82–471–3786–4 (electronic ver.), 2012.

210. D. L. Langer, T. H. van der Kwast, A. J. Evans, A. Plotkin, J. Trachtenberg, B. C. Wilson, and M. A. Haider, "Prostate tissue composition and mr measurements: Investigating the relationships between adc, t2, k trans, ve, and corresponding histologic features 1," *Radiology*, vol. 255, no. 2, pp. 485–494, 2010.

211. K. M. Selnæs, I. S. Gribbestad, H. Bertilsson, A. Wright, A. Angelsen, A. Heerschap, and M.-B. Tessem, "Spatially matched in vivo and ex vivo mr metabolic profiles of prostate cancer–investigation of a correlation with gleason score," *NMR in biomedicine*, vol. 26, no. 5, pp. 600–606, 2013.

212. T. W. Scheenen, S. W. Heijmink, S. A. Roell, C. A. Hulsbergen-Van de Kaa, B. C. Knipscheer, J. A. Witjes, J. O. Barentsz, and A. Heerschap, "Three-dimensional proton mr spectroscopy of human prostate at 3 t without endorectal coil: Feasibility 1," *Radiology*, vol. 245, no. 2, pp. 507–516, 2007.

213. M.-B. Tessem, H. Bertilsson, A. Angelsen, T. F. Bathen, F. Drabløs, and M. B. Rye, "A balanced tissue composition reveals new metabolic and gene expression markers in prostate cancer," *PloS one*, vol. 11, no. 4, p. e0153727, 2016.

214. S. A. Tomlins, R. Mehra, D. R. Rhodes, X. Cao, L. Wang, S. M. Dhanasekaran, S. Kalyana-Sundaram, J. T. Wei, M. A. Rubin, K. J. Pienta, *et al.*, "Integrative molecular concept modeling of prostate cancer progression," *Nature genetics*, vol. 39, no. 1, pp. 41–51, 2007.

215. S. A. Tomlins, B. Laxman, S. Varambally, X. Cao, J. Yu, B. E. Helgeson, Q. Cao, J. R. Prensner, M. A. Rubin, R. B. Shah, *et al.*, "Role of the tmprss2-erg gene fusion in prostate cancer," *Neoplasia*, vol. 10, no. 2, pp. 177–IN9, 2008.

216. S. R. Setlur, K. D. Mertz, Y. Hoshida, F. Demichelis, M. Lupien, S. Perner, A. Sboner, Y. Pawitan, O. Andrén, L. A. Johnson, *et al.*, "Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer," *Journal of the National Cancer Institute*, vol. 100, no. 11, pp. 815–825, 2008.

217. K. Iljin, M. Wolf, H. Edgren, S. Gupta, S. Kilpinen, R. I. Skotheim, M. Peltola, F. Smit, G. Verhaegh, J. Schalken, *et al.*, "Tmprss2 fusions with oncogenic ets factors in prostate cancer involve unbalanced genomic rearrangements and are associated with hdac1 and epigenetic reprogramming," *Cancer research*, vol. 66, no. 21, pp. 10242–10246, 2006.

218. R. M. Kypta and J. Waxman, "Wnt/β-catenin signalling in prostate cancer," *Nature Reviews Urology*, vol. 9, no. 8, pp. 418–428, 2012.

219. M. Verras and Z. Sun, "Roles and regulation of wnt signaling and β-catenin in prostate cancer," *Cancer letters*, vol. 237, no. 1, pp. 22–32, 2006.

220. S. Majid, S. Saini, and R. Dahiya, "Wnt signaling pathways in urological cancers: past decades and still growing," *Mol Cancer*, vol. 11, no. 7, 2012.

221. A. F. Hansen, E. Sandsmark, M. B. Rye, A. J. Wright, H. Bertilsson, E. Richardsen, T. Viset, A. M. Bofin, A. Angelsen, K. M. Selnæs, *et al.*, "Presence of tmprss2-erg is associated with alterations of the metabolic profile in human prostate cancer," *Oncotarget*, vol. 7, no. 27, p. 42071, 2016.

References

222. P. A. Kupelian, J. Katcher, H. S. Levin, and E. A. Klein, "State t1–2 prostate cancer: a multivariate analysis of factors affecting biochemical and clinical failures after radical prostatectomy," *International Journal of Radiation Oncology* Biology* Physics*, vol. 37, no. 5, pp. 1043–1052, 1997.

223. A. V. D'amico, R. Whittington, S. B. Malkowicz, D. Schultz, K. Blank, G. A. Broderick, J. E. Tomaszewski, A. A. Renshaw, I. Kaplan, C. J. Beard, *et al.*, "Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer," *Jama*, vol. 280, no. 11, pp. 969–974, 1998.

224. K. S. Opstad, B. A. Bell, J. R. Griffiths, and F. A. Howe, "An assessment of the effects of sample ischaemia and spinning time on the metabolic profile of brain tumour biopsy specimens as determined by high-resolution magic angle spinning 1h nmr," *NMR in Biomedicine*, vol. 21, no. 10, pp. 1138–1147, 2008.

225. T. H. Haukaas, S. A. Moestue, R. Vettukattil, B. Sitter, S. Lamichhane, R. Segura, G. F. Giskeødegård, and T. F. Bathen, "impact of freezing delay time on tissue samples for metabolomic studies," *Frontiers in oncology*, vol. 6, 2016.

226. P. Micke, M. Ohshima, S. Tahmasebpoor, Z.-P. Ren, A. Östman, F. Pontén, and J. Botling, "Biobanking of fresh frozen tissue: Rna is stable in nonfixed surgical specimens," *Laboratory Investigation*, vol. 86, no. 2, pp. 202–211, 2006.

227. U. R. Chandran, R. Dhir, C. Ma, G. Michalopoulos, M. Becich, and J. Gilbertson, "Differences in gene expression in prostate cancer, normal appearing prostate tissue adjacent to cancer and prostate tissue from cancer free organ donors," *BMC cancer*, vol. 5, no. 1, p. 45, 2005.

228. M. C. Risk, B. S. Knudsen, I. Coleman, R. F. Dumpit, A. R. Kristal, N. LeMeur, R. C. Gentleman, L. D. True, P. S. Nelson, and D. W. Lin, "Differential gene expression in benign prostate epithelium of men with and without prostate cancer: evidence for a prostate cancer field effect," *Clinical Cancer Research*, vol. 16, no. 22, pp. 5414–5423, 2010.

229. S. Weis, I. Llenos, J. Dulay, M. Elashoff, F. Martinez-Murillo, and C. Miller, "Quality control for microarray analysis of human brain samples: the impact of postmortem factors, rna characteristics, and histopathology," *Journal of neuroscience methods*, vol. 165, no. 2, pp. 198–209, 2007.

230. S. Imbeaud, E. Graudens, V. Boulanger, X. Barlet, P. Zaborski, E. Eveno, O. Mueller, A. Schroeder, and C. Auffray, "Towards standardization of rna quality assessment using user-independent classifiers of microcapillary electrophoresis traces," *Nucleic acids research*, vol. 33, no. 6, pp. e56–e56, 2005.

231. T. Casneuf, Y. Van de Peer, and W. Huber, "In situ analysis of cross-hybridisation on microarrays and the inference of expression correlation," *BMC bioinformatics*, vol. 8, no. 1, p. 461, 2007.

232. S. Tomiuk and K. Hofmann, "Microarray probe selection strategies," *Briefings in bioinformatics*, vol. 2, no. 4, pp. 329–340, 2001.

233. S. Draghici, P. Khatri, A. C. Eklund, and Z. Szallasi, "Reliability and reproducibility issues in dna microarray measurements," *TRENDS in Genetics*, vol. 22, no. 2, pp. 101–109, 2006.

234. K. J. Mantione, R. M. Kream, H. Kuzelova, R. Ptacek, J. Raboch, J. M. Samuel, and G. B. Stefano, "Comparing bioinformatic gene expression profiling methods: Microarray and rna-seq," *Medical science monitor basic research*, vol. 20, p. 138, 2014.

235. J. L. Gregg, K. E. Brown, E. M. Mintz, H. Piontkivska, and G. C. Fraizer, "Analysis of gene expression in prostate cancer epithelial and interstitial stromal cells using laser capture microdissection," *BMC cancer*, vol. 10, no. 1, p. 165, 2010.

236. L. A. Field, B. Deyarmin, C. D. Shriver, D. L. Ellsworth, and R. E. Ellsworth, "Laser microdissection for gene expression profiling," *Laser Capture Microdissection: Methods and Protocols*, pp. 17–45, 2011.

237. N. Crosetto, M. Bienko, and A. Van Oudenaarden, "Spatially resolved transcriptomics and beyond," *Nature Reviews Genetics*, vol. 16, no. 1, pp. 57–66, 2015.

238. P. L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, *et al.*, "Visualization and analysis of gene expression in tissue sections by spatial transcriptomics," *Science*, vol. 353, no. 6294, pp. 78–82, 2016.

239. T. Maier, M. Güell, and L. Serrano, "Correlation of mrna and protein in complex biological samples," *FEBS letters*, vol. 583, no. 24, pp. 3966–3973, 2009.

240. A. B. Sachs, P. Sarnow, and M. W. Hentze, "Starting at the beginning, middle, and end: translation initiation in eukaryotes," *Cell*, vol. 89, no. 6, pp. 831–838, 1997.

241. M. Hochstrasser, "Ubiquitin, proteasomes, and the regulation of intracellular protein degradation," *Current opinion in cell biology*, vol. 7, no. 2, pp. 215–223, 1995.

242. D. Barford, A. K. Das, and M.-P. Egloff, "The structure and mechanism of protein phosphatases: insights into catalysis and regulation," *Annual review of biophysics and biomolecular structure*, vol. 27, no. 1, pp. 133–164, 1998.

243. A. Görg, W. Weiss, and M. J. Dunn, "Current two-dimensional electrophoresis technology for proteomics," *Proteomics*, vol. 4, no. 12, pp. 3665–3685, 2004.

244. R. Aebersold and M. Mann, "Mass spectrometry-based proteomics," *Nature*, vol. 422, no. 6928, pp. 198–207, 2003.

245. N. Barker and M. van den Born, "Detection of beta-catenin localization by immunohistochemistry," in *Wnt Signaling*, pp. 91–98, Springer, 2008.

246. R. Walker, "Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment i," *Histopathology*, vol. 49, no. 4, pp. 406–410, 2006.

247. C. Taylor and R. Levenson, "Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment ii," *Histopathology*, vol. 49, no. 4, pp. 411–424, 2006.

248. J. Kononen, L. Bubendorf, A. Kallionimeni, M. Bärlund, P. Schraml, S. Leighton, J. Torhorst, M. J. Mihatsch, G. Sauter, and O.-P. Kallionimeni, "Tissue microarrays for high-throughput molecular profiling of tumor specimens," *Nature medicine*, vol. 4, no. 7, pp. 844–847, 1998.

249. E. C. Stack, C. Wang, K. A. Roman, and C. C. Hoyt, "Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of tyramide signal amplification, multispectral imaging and multiplex analysis," *Methods*, vol. 70, no. 1, pp. 46–58, 2014.

250. B. Sitter, T. F. Bathen, M.-B. Tessem, and I. S. Gribbestad, "High-resolution magic angle spinning (hr mas) mr spectroscopy in metabolic characterization of human cancer," *Progress in nuclear magnetic resonance spectroscopy*, vol. 54, no. 3, pp. 239–254, 2009.

251. N. Tayari, A. Heerschap, T. W. Scheenen, and T. Kobus, "In vivo mr spectroscopic imaging of the prostate, from application to interpretation," *Analytical Biochemistry*, 2017.

252. J. L. Taylor, C.-L. Wu, D. Cory, R. G. Gonzalez, A. Bielecki, and L. L. Cheng, "High-resolution magic angle spinning proton nmr analysis of human prostate tissue with slow spinning rates," *Magnetic resonance in medicine*, vol. 50, no. 3, pp. 627–632, 2003.

253. T. F. Bathen, B. Sitter, T. E. Sjøbakk, M.-B. Tessem, and I. S. Gribbestad, "Magnetic resonance metabolomics of intact tissue: a biotechnological tool in cancer diagnostics and treatment evaluation," *Cancer research*, vol. 70, no. 17, pp. 6692–6696, 2010.

254. K. Dettmer, P. A. Aronov, and B. D. Hammock, "Mass spectrometry-based metabolomics," *Mass spectrometry reviews*, vol. 26, no. 1, pp. 51–78, 2007.

255. A. Alonso, S. Marsal, and A. Julià, "Analytical methods in untargeted metabolomics: state of the art in 2015," *Frontiers in bioengineering and biotechnology*, vol. 3, p. 23, 2015.

References

256. S. G. Villas-Bôas, S. Mas, M. Åkesson, J. Smedsgaard, and J. Nielsen, "Mass spectrometry in metabolome analysis," *Mass spectrometry reviews*, vol. 24, no. 5, pp. 613–646, 2005.

257. Y. Fujimura and D. Miura, "Maldi mass spectrometry imaging for visualizing in situ metabolism of endogenous metabolites and dietary phytochemicals," *Metabolites*, vol. 4, no. 2, pp. 319–346, 2014.

258. D. A. Pirman, E. Efuet, X.-P. Ding, Y. Pan, L. Tan, S. M. Fischer, R. N. DuBois, and P. Yang, "Changes in cancer cell metabolism revealed by direct sample analysis with maldi mass spectrometry," *PloS one*, vol. 8, no. 4, p. e61379, 2013.

259. G. Braadland, Peder R .and Giskeodegaard, E. Sandsmark, H. Bertilsson, L. R. Euceda, A. F. Hansen, I. J. Guldvik, K. M. Selnaes, H. H. Grytli, B. Katz, T. F. Svindal, Aud amd Bathen, L. M. Eri, S. Nygård, V. Berge, K. A. Tasken, and M.-B. Tessem, "Ex vivo metabolic fingerprinting identifies biomarkers predictive of prostate cancer recurrence following radical prostatectomy." Sumbitted: British Journal of Cancer 2017.

260. J. N. Anastas and R. T. Moon, "Wnt signalling pathways as therapeutic targets in cancer," *Nature Reviews Cancer*, vol. 13, no. 1, pp. 11–26, 2013.

261. W. Lu, H. N. Tinsley, A. Keeton, Z. Qu, G. A. Piazza, and Y. Li, "Suppression of wnt/β-catenin signaling inhibits prostate cancer cell proliferation," *European journal of pharmacology*, vol. 602, no. 1, pp. 8–14, 2009.

262. D. Grandy, J. Shan, X. Zhang, S. Rao, S. Akunuru, H. Li, Y. Zhang, I. Alpatov, X. A. Zhang, R. A. Lang, *et al.*, "Discovery and characterization of a small molecule inhibitor of the pdz domain of dishevelled," *Journal of Biological Chemistry*, vol. 284, no. 24, pp. 16256–16263, 2009.

263. M. Irfan Maqsood, M. M. Matin, A. R. Bahrami, and M. M. Ghasroldasht, "Immortality of cell lines: challenges and advantages of establishment," *Cell biology international*, vol. 37, no. 10, pp. 1038–1045, 2013.

264. S. McDonald and A. Silver, "The opposing roles of wnt-5a in cancer," *British journal of cancer*, vol. 101, no. 2, pp. 209–214, 2009.

265. N. Zhu, L. Qin, Z. Luo, Q. Guo, L. Yang, and D. Liao, "Challenging role of wnt5a and its signaling pathway in cancer metastasis (review)," *Experimental and therapeutic medicine*, vol. 8, no. 1, pp. 3–8, 2014.

266. L. Topol, X. Jiang, H. Choi, L. Garrett-Beal, P. J. Carolan, and Y. Yang, "Wnt-5a inhibits the canonical wnt pathway by promoting gsk-3–independent β-catenin degradation," *The Journal of cell biology*, vol. 162, no. 5, pp. 899–908, 2003.

267. H. Yuzugullu, K. Benhaj, N. Ozturk, S. Senturk, E. Celik, A. Toylu, N. Tasdemir, M. Yilmaz, E. Erdal, K. C. Akcali, *et al.*, "Canonical wnt signaling is antagonized by noncanonical wnt5a in hepatocellular carcinoma cells," *Mol Cancer*, vol. 8, no. 90, 2009.

268. S. Wang, M. Krinks, K. Lin, F. P. Luyten, and M. Moos, "Frzb, a secreted protein expressed in the spemann organizer, binds and inhibits wnt-8," *Cell*, vol. 88, no. 6, pp. 757–766, 1997.

269. P. García-Tobilla, S. R. Solórzano, I. Salido-Guadarrama, V. González-Covarrubias, G. Morales-Montor, C. E. Díaz-Otañez, and M. Rodríguez-Dorantes, "Sfrp1 repression in prostate cancer is triggered by two different epigenetic mechanisms," *Gene*, vol. 593, no. 2, pp. 292–301, 2016.

270. A. S. Perry, G. O'hurley, O. A. Raheem, K. Brennan, S. Wong, A. O'grady, A.-M. Kennedy, L. Marignol, T. M. Murphy, L. Sullivan, *et al.*, "Gene expression and epigenetic discovery screen reveal methylation of sfrp2 in prostate cancer," *International journal of cancer*, vol. 132, no. 8, pp. 1771–1780, 2013.

271. T. Shi, Y. Gao, W.-j. Qian, R. Leach, I. Thompson, S.-I. Quek, W. Ellis, E. Vitello, and A. Liu, "S&t-10 prostate cancer diagnosis by multiple secreted protein biomarkers in voided urine," *The Journal of Urology*, vol. 195, no. 4, p. e312, 2016.

272. D. Ricci, S. Thomas, M. Gormley, Y. Rajpurohit, and M. Schaffer, "Whole blood based mrna markers for predicting prostate cancer and methods of detecting the same," Mar. 11 2016. US Patent App. 15/067,716.

273. L. C. Costello and R. B. Franklin, "Novel role of zinc in the regulation of prostate citrate metabolism and its implications in prostate cancer," *Prostate*, vol. 35, pp. 285–96, Jun 1998.

274. L. C. Costello and R. B. Franklin, "The clinical relevance of the metabolism of prostate cancer; zinc and tumor suppression: connecting the dots," *Molecular cancer*, vol. 5, no. 1, p. 17, 2006.

275. S. Mandal, A. Mandal, H. E. Johansson, A. V. Orjalo, and M. H. Park, "Depletion of cellular polyamines, spermidine and spermine, causes a total arrest in translation and growth in mammalian cells," *Proceedings of the National Academy of Sciences*, vol. 110, no. 6, pp. 2169–2174, 2013.

276. F. Gibellini and T. K. Smith, "The kennedy pathway—de novo synthesis of phosphatidylethanolamine and phosphatidylcholine," *IUBMB life*, vol. 62, no. 6, pp. 414–428, 2010.

277. M. G. Swanson, K. R. Keshari, Z. L. Tabatabai, J. P. Simko, K. Shinohara, P. R. Carroll, A. S. Zektzer, and J. Kurhanewicz, "Quantification of choline-and ethanolamine-containing metabolites in human prostate tissues using 1h hr-mas total correlation spectroscopy," *Magnetic Resonance in Medicine*, vol. 60, no. 1, pp. 33–40, 2008.

Paper I

Research Paper

Presence of TMPRSS2-ERG is associated with alterations of the metabolic profile in human prostate cancer

Ailin Falkmo Hansen[1], Elise Sandsmark[1], Morten Beck Rye[2,3,4], Alan J. Wright[5], Helena Bertilsson[2,4], Elin Richardsen[6], Trond Viset[7], Anna M. Bofin[8], Anders Angelsen[1], Kirsten M. Selnæs[1], Tone Frost Bathen[1], May-Britt Tessem[1]

[1]Department of Circulation and Medical Imaging, Faculty of Medicine, NTNU, Norwegian University of Science and Technology, Trondheim, Norway

[2]Department of Cancer Research and Molecular Medicine, Faculty of Medicine, NTNU, Norwegian University of Science and Technology, Trondheim, Norway

[3]St. Olavs Hospital, Trondheim, Norway

[4]Department of Urology, St. Olavs Hospital, Trondheim, Norway

[5]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom

[6]Department of Medical Biology, UiT - The Arctic University of Norway, Tromsø, Norway

[7]Department of Pathology and Medical Genetics, St. Olavs Hospital, Trondheim, Norway

[8]Department of Laboratory Medicine, Children's and Women's Health, Faculty of Medicine, NTNU, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence to: May-Britt Tessem, **email:** may-britt.tessem@ntnu.no

ABSTRACT

TMPRSS2-ERG has been proposed to be a prognostic marker for prostate cancer. The aim of this study was to identify changes in metabolism, genes and biochemical recurrence related to TMPRSS2-ERG by using an integrated approach, combining metabolomics, transcriptomics, histopathology and clinical data in a cohort of 129 human prostate samples (41 patients). Metabolic analyses revealed lower concentrations of citrate and spermine comparing ERG_{high} to ERG_{low} samples, suggesting an increased cancer aggressiveness of ERG_{high} compared to ERG_{low}. These results could be validated in a separate cohort, consisting of 40 samples (40 patients), and magnetic resonance spectroscopy imaging (MRSI) indicated an *in vivo* translational potential. Alterations of gene expression levels associated with key enzymes in the metabolism of citrate and polyamines were in consistence with the metabolic results. Furthermore, the metabolic alterations between ERG_{high} and ERG_{low} were more pronounced in low Gleason samples than in high Gleason samples, suggesting it as a potential tool for risk stratification. However, no significant difference in biochemical recurrence was detected, although a trend towards significance was detected for low Gleason samples. Using an integrated approach, this study suggests TMPRSS2-ERG as a potential risk stratification tool for inclusion of active surveillance patients.

INTRODUCTION

The genetic fusion between the erythroblast transformation-specific (*ETS*) transcriptional factor ETS-related gene (*ERG*) and the androgen-responsive promotor transmembrane protease, serine 2 (*TMPRSS2*) [1] is suggested to be a major mechanism driving prostate carcinogenesis. The TMPRSS2-ERG gene fusion is the most common gene rearrangement in prostate cancer [2], with a reported prevalence of 15-78% [3]. Presence of the gene fusion is the main reason for overexpression of *ERG* which is further associated with epithelial-to-mesenchymal potential, cell invasion and cell proliferation [4].

From the initial discovery in 2005 [5], the TMPRSS2-ERG gene fusion has been linked to clinical outcome parameters such as early onset of prostate cancer [6], negative outcome in watchful waiting patients

[7–9] and a higher risk of disease progression in active surveillance patients [10]. However, considering the prognostic value of TMPRSS2-ERG in prostatectomy patients, most studies find no association to outcome after surgery [6, 11–13]. In a meta-analysis of 5,074 prostatectomy specimens, there were no associations between the presence of TMPRSS2-ERG and biochemical recurrence or lethal disease [14]. Although the clinical significance of TMPRSS2-ERG is yet to be proven, presence of the fusion gene is a key genomic event specific for prostate cancer that may be of importance for risk assessment or treatment stratification of prostate cancer patients.

Metabolic markers may be indicative of aggressive disease and provide diagnostic and therapeutic information for improved characterization and stratification of prostate cancer patients. Lower levels of citrate and spermine have previously been linked to higher Gleason grade and more aggressive prostate cancer [15]. Citrate and spermine, including choline and creatine are metabolites detectable by *in vivo* patient magnetic resonance spectroscopy imaging (MRSI), which imply a potential for transferring biomarkers to a clinical setting [16]. A recent study revealed ERG-specific metabolic alterations, particularly connected to fatty acid oxidation [17] and an earlier study found increased glucose uptake to be related to the metabolic sensor neuropeptide gamma (*NPY*) in ERG rearrangement positive prostate cancer [18]. Apart from these two studies, the relationship between cancer metabolism and TMPRSS2-ERG remains unexplored.

The integration of transcriptomic data with metabolomics and histopathology is a promising tool for gaining important molecular information, in order to understand states and pathways of disease. In this study, we used prostatectomy tissue samples obtained through a standardized harvesting protocol [19] where metabolic and gene expression data are collected after histopathology evaluation [20] in order to integrate data from transcriptomics, metabolomics and histopathology. Prostate tissue samples were analyzed by HR-MAS (high resolution magic angle spinning) MRS (magnetic resonance spectroscopy), followed by detection of the fusion gene using gene expression microarray measurements for the main cohort, and fluorescence *in situ* hybridization (FISH) for an independent validation cohort. HR-MAS is a non-destructive method, which permits gene expression analysis and histology to be performed on the exact same tissue sample, providing an excellent basis for correlating metabolic findings with concordant alterations in the transcriptome. The main objective of this study was to combine these techniques to investigate presence of the TMPRSS2-ERG gene fusion in two cohorts of human prostate cancer tissue and to identify its association to metabolism and biochemical recurrence.

RESULTS AND DISCUSSION

The presence of TMPRSS2-ERG or expressing high ERG levels was in our prostate cancer patient cohorts associated with metabolic alterations and concordant changes of gene expression levels related to key metabolic genes. In two independent patient cohorts, we observed a decrease in concentrations of citrate and spermine in fusion positive and ERG_{high} patients, indicating increased aggressiveness according to previous findings on prostate cancer metabolism [15]. In addition, this relationship was significant within low Gleason samples which propose an early patient stratification possibility based on the fusion status and metabolic biomarkers.

Presence of TMPRSS2-ERG/high ERG status

A 2 mm transversal prostate tissue slice was collected from 41 patients and from each slice several samples (median: 3, range: 1 to 6 per slice, depending on tumor size) were collected from cancerous and adjacent benign areas, in total 95 cancer and 34 benign samples, and termed the main cohort. Among the cancer samples, 34 of 95 (35.8%) were classified as ERG_{high}, and were expected to possess the TMPRSS2-ERG fusion gene, while 30 (31.6%) and 31 (32.6%) were classified as ERG_{low} and $ERG_{intermediate}$, respectively. In addition, 34 (26.4%) of the 129 samples in the cohort were classified as benign samples. The proportions harboring the fusion gene are in the lower range of the reported prevalence of 15-78% [3].

Generally, samples obtained from the same prostate, were all placed in the same ERG group or the adjacent ERG group. However, out of the 41 patients, 6 (14.6%) patients had samples belonging to all three ERG groups (Supplementary Table S1), which is in consistence with previously reports of ERG interfocal heterogeneity [21, 22]. Three patients had no cancer samples, leaving 38 patients as the main focus of this study. In order to validate our results, a second cohort of 90 prostate cancer patients was included, consisting of one needle biopsy sample per patient obtained after radical prostatectomy. Only 40 of the needle biopsies contained cancer and were included in the present study. In the validation cohort, 7 out of 40 patients, (17.5%) were fusion positive, while 33 out of 40 (82.5%) were fusion negative. The lower prevalence of TMPRSS2-ERG in the validation cohort may be due to a lower amount of tumor in the samples (median cancer content 40% and 70% in the validation and main cohort, respectively) and sampling only one sample per patient may fail to detect presence of TMPRSS2-ERG present in other parts of the prostate. Sample characteristics of both cohorts are presented in Table 1.

Table 1: Clinical characteristics for samples in the main- and validation cohort

	Main cohort			Validation cohort	
	ERG$_{low}$	ERG$_{intermediate}$	ERG$_{high}$	TMPRSS2-ERG negative	TMPRSS2-ERG positive
Prevalence	30	31	34	33	7
Gleason score of tissue samples					
0	0	0	0	0	0
6	7	8	9	5	1
3 + 4	5	8	8	11	3
4 + 3	4	7	9	9	0
8	8	2	5	3	2
9	6	6	3	2	1
10	0	0	0	2	0
Not evaluated	-	-	-	1	-
Cancer content (%) mean (range)	59 (10 to 90)	66 (20 to 90)	64 (20 to 85)	38 (5 to 80)	36 (5 to 80)
Stroma content (%) mean (range)	28 (5 to 50)	23 (0 to 70)	26 (10 to 50)	39 (20 to 70)	45 (10 to 65)
Benign epithelial content (%) mean (range)	13 (0 to 50)	11 (0 to 30)	10 (0 to 40)	26 (0 to 40)	26 (10 to 30)
Luminal space (%) mean (range)	9 (0 to 32)	6 (0 to 30)	8 (0 to 21)	4 (0 to 14)	5 (0 to 13)

Metabolic alterations associated with TMPRSS2-ERG/high ERG status

Unsupervised multivariate analysis of the metabolic profiles of the main cohort revealed a trend of clustering with respect to the three different ERG groups and the benign samples (Figure 1A). Significant trends across increasing ERG groups (cancer samples) were detected for the levels of citrate, spermine, putrescine, ethanolamine, glucose, glycine, phosphocholine and phosphoethanolamine (Figure 1B and Table 2). In normal prostate cells citrate is accumulated, while in prostate cancer, citrate is decreased or depleted [23]. Additionally, the normal prostate cells have one of the highest concentrations of polyamines in the body [24], and the polyamines are important for a variety of functions within the cell such as e.g. apoptosis, cell proliferation and differentiation [25, 26]. Decreasing levels of citrate, spermine and putrescine with increasing ERG status, suggested increased aggressiveness [15] of higher ERG status groups compared to lower ERG status groups. The increased levels of ethanolamine, phosphocholine and phosphoethanolamine further suggest an increased aggressiveness of the higher ERG status, as increased concentrations of choline-associated metabolites have been reported in prostate cancer, and are important

in proliferation as structural components of cellular membranes [27, 28]. Glycine may also be important considering previous findings in breast cancer, suggesting it to be a marker of lower survival rates [29].

Comparable results were found building a partial least squares discriminant analysis (PLS-DA) model based on the metabolic profiles where ERG$_{high}$ was separated from ERG$_{low}$ with an accuracy of 77% (sensitivity: 79%, specificity: 74%), $p < 0.001$ (Figure 1C). This proves that the metabolic profiles of samples which are expected to possess the fusion gene are well separated from those most likely not to harbor the gene rearrangement. Further, the loading plot for the latent variable 1 (LV1) (Figure 1D), explaining which metabolites that are important for the separation along LV1, showed decreased levels of citrate and polyamine levels in ERG$_{high}$ compared to ERG$_{low}$, while levels of choline-containing compounds were higher, supporting the hypothesis of a more aggressive phenotype of fusion positive prostate cancers.

Among the 23 quantified metabolites in the main cohort, the concentrations of citrate, spermine, putrescine and glucose were significantly decreased in ERG$_{high}$ samples compared to ERG$_{low}$, while the concentrations of glycine were significantly increased (Supplementary Table S2). However, after multiple testing corrections, only citrate and spermine were significant (Figure 2A and 2B, Table 2). In addition,

our study revealed similar metabolic levels between ERG_{low} and benign samples (Supplementary Table S3), possibly confounded by effects of tissue heterogeneity [30], especially differences in stromal content between cancer and benign samples. Similar metabolic levels of citrate and spermine have previously been found comparing low Gleason grade and benign samples [15]. Despite the low prevalence of the fusion gene in the validation cohort, significantly decreased concentrations of citrate and spermine were detected in fusion positive samples. However, these differences were not
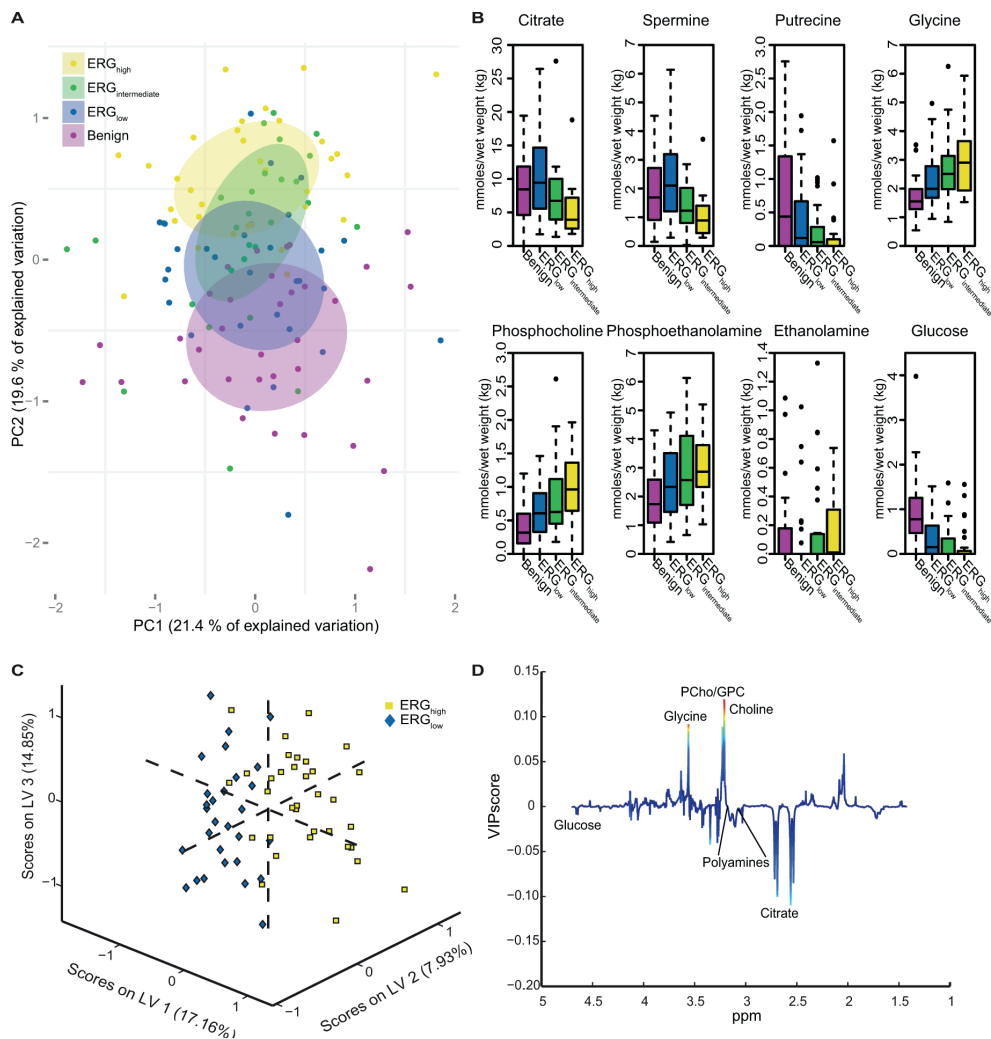
Figure 1: Multivariate analysis of spectral data and absolute quantification reveals metabolic differences between ERG groups. (**A**) Principal component analysis (PCA) reveals a trend in the distribution of the metabolic clusters of metabolic profile from benign samples (purple) across ERGlow (blue) and ERGintermediate (green) to ERGhigh (yellow). (**B**) Absolute quantification of 23 metabolites showed significant trend across cancer samples, from ERGlow, (blue) through ERGintermediate (green) to ERGhigh (yellow) for eight of the metabolites. Increasing trends were found for glycine, phosphocholine, phosphoethanolamine, and ethanolamine. Decreasing trends were found for citrate, spermine, putrescine and glucose. Benign (purple) samples are shown for comparisons. (**C**) A partial least squares discriminant analysis (PLS-DA) model was able to separate ERGhigh (yellow) and ERGlow (blue) with an accuracy of 77 %, $p < 0.001$. (**D**) Loadings plot of latent variable 1 (LV1) indicate lower levels of citrate and the polyamines and higher levels of choline-containing metabolites comparing ERGhigh to ERGlow. The loadings are colored according to the variable importance in the projection (VIP) scores.

Table 2: Differences in levels of quantified metabolites in the main cohort comparing ERG_high samples with ERG_low samples

Metabolites	ERG_low $n = 30$ Concentrations mmoles/kg wet weight, median (IQR)	ERG_intermediate $n = 31$ Concentrations mmoles/kg wet weight, median (IQR)	ERG_high $n = 34$ Concentrations mmoles/kg wet weight, median (IQR)	ERG_high vs ERG_low (p-values)	ERG_high vs ERG_low (adjusted p-values)	p-trend
Citrate	9.44 (5.56 to 14.68)	6.74 (3.94 to 10.34)	3.91 (2.59 to 7.20)	< 0.001	< 0.001	< 0.001
Ethm	0 (0 to 0)	0 (0 to 0.15)	0.01 (0 to 0.31)	0.490	0.663	0.043
Glucose	0.15 (0.00 to 0.63)	0.00 (0.00 to 0.43)	0.00 (0.00 to 0.07)	0.008	0.061	0.007
Glycine	1.99 (1.68 to 2.78)	2.51 (1.98 to 3.18)	2.90 (1.93 to 3.65)	0.023	0.115	0.008
PCh	0.61 (0.33 to 0.91)	0.63 (0.43 to 1.17)	0.96 (0.64 to 1.36)	0.067	0.248	0.005
PE	2.33 (1.46 to 3.51)	2.57 (1.67 to 4.14)	2.86 (2.33 to 3.79)	0.087	0.248	0.032
Putrescine	0.12 (0 to 0.67)	0.06 (0 to 0.30)	0 (0 to 0.10)	0.025	0.115	0.003
Spermine	2.10 (1.20 to 3.19)	1.23 (0.79 to 2.02)	0.89 (0.45 to 1.40)	< 0.001	< 0.001	< 0.001

Ethm: Ethanolamine, PCh: Phosphocholine, PE: Phosphoethanolamine, IQR: Interquartile range. Adjusted p-values are adjusted using Benjamini-Hochberg correction for multiple testing.

significant after corrections for multiple testing, possibly due to the relatively small patient cohort, variations in the amount of cancer tissue between samples and the low number of fusion positive samples in this cohort (Figure 2C and 2D, Supplementary Table S4). A recent study [17] supports our metabolic findings by presenting significantly increased levels of glycerophosphoethanolamine, glycine, isoleucine, leucine and glutamate between ERG positive and ERG negative patients, and significantly decreased levels of myo-inositol, creatine, citrate, glucose, spermine and putrescine. However, we were not able to reveal any metabolic changes related to glycerophosphoethanolamine, isoleucine, leucine, glutamate and myo-inositol suggested by Meller et al. [17].

Targeted analyses of key metabolic pathways

Due to the observed citrate and spermine changes, we performed targeted analyses of genes related to the polyamine pathway and citrate. We also investigated metabolic pathways connected to glycine and glucose metabolism, as TMPRSS2-ERG has been suggested to be linked to increased glucose uptake [17, 18].

The polyamine pathway

Expression of polyamine pathway genes were found to be increased in ERG_high samples compared to ERG_low, where spermidine synthase (SRM) and spermidine N(1)-acetyltransferase (SAT1) displayed the highest significance (Figure 3A and Supplementary Table S5). Especially, the strong upregulation of SAT1 leads to a rapid depletion of cellular spermidine and spermine [31], which is in

agreement with the low concentrations of spermine observed in ERG_high compared to ERG_low. In addition, overexpression of SRM without concordant upregulation of ornithine decarboxylase (ODC1) will lead to reduced levels of putrescine which was observed in the present study. ODC1 overexpression is reported frequently among prostate cancer patients [32, 33], where it mediates the conversion of ornithine to putrescine which is the rate-limiting enzyme of the polyamine pathway. However, this does not seem to be the main mode of regulation in ERG_high versus ERG_low samples in our cohort, where changes in SAT1 and SRM seem to be the main drivers of altered polyamine metabolism.

Citrate and fatty acid synthesis

In the present study, we found that a significantly decreased expression of ACO2 in the TCA cycle (Figure 3B and Supplementary Table S5) is linked to a phenotype characterized by low levels of citrate in ERG_high tissue samples. Franklin and Costello [34] suggested that normal prostate epithelial cells are citrate-producing, but become citrate-oxidizing following transformation to malignant cells, and that ACO2 is the key enzyme for this transformation. Decreased expression of ACO2 have been linked to increased citrate secretion [35], causing higher levels of citrate which can be redirected to the cytosol, contributing to restore acetyl-CoA and oxaloacetate pools. Our results indicate that citrate is shunted out to the cytosol where it may be used for de novo synthesis of fatty acids to meet the high demands for building blocks for biosynthesis in cancer, as we observed an increased expression of the
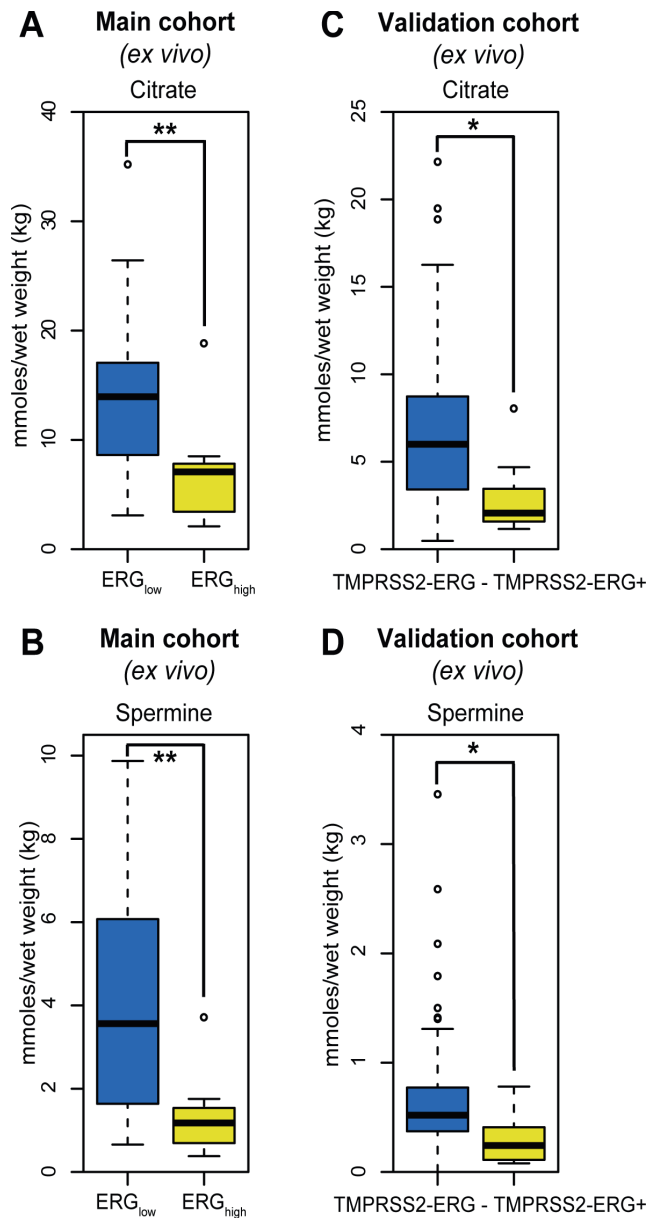
Figure 2: Box-plots for citrate and spermine comparing ERGhigh and ERGlow samples in the main cohort (*ex vivo*) and fusion positive and fusion negative patients in the validation cohort. (**A**) Decreased levels of citrate were found comparing ERGhigh to ERGlow samples in the main cohort, $p < 0.001$, (**B**) Decreased levels of spermine were found comparing ERGhigh to ERGlow samples in the main cohort, $p < 0.001$, (**C**) Decreased levels of citrate were found comparing fusion positive to fusion negative patients in the validation cohort, $p = 0.013$, (**D**) Decreased levels of spermine were found comparing fusion positive to fusion negative patients in the validation cohort, $p = 0.021$.

key lipogenic enzymes acetyl-CoA carboxylase alpha (*ACACA*) and fatty acid synthase (*FASN*) in ERG$_{high}$ tissue samples. Additionally, a higher expression of long-chain acyl-CoA synthetase3 (*ASCL3*) was detected, which is previously suggested to cause lipid accumulation [36]. High expression of *FASN*, have been found increased in several types of cancers, including prostate cancer and is strongly correlated with malignant transformation and poor prognosis [37, 38]. Increased fatty acid synthesis is suggested to be a key feature of prostate cancer suggesting aggressiveness of disease [38], and the increased lipogenic profile of ERG$_{high}$ samples supports the hypothesis of an increased aggressiveness with presence of TMPRSS2-ERG.

Glucose, glycine and pentose phosphate pathway

A significant reduction of glucose was found prior to correction for multiple testing, comparing ERG$_{high}$ with ERG$_{low}$. We detected a differential expression of NPY, in line with results from a previous study [18], comparing ERG$_{high}$ and ERG$_{low}$ samples, where lower levels of glucose were connected to a phenotype with a higher expression of *NPY*. Our results indicate that ERG$_{high}$ samples have lower glucose levels or are rapidly consuming glucose and thus lowering the detectable glucose levels. Moreover, there was not an increased concentration of lactate in ERG$_{high}$ compared to ERG$_{low}$, and both increased and decreased expression of key enzymes within glycolysis and the TCA were detected (Figure 3B and Supplementary Table S5). However, a

highly significant increased expression of oxoglutarate dehydrogenase-like (*OGDHL*), a key control point in the TCA, was found in ERG$_{high}$ compared to ERG$_{low}$. Interestingly, the increased expression of pyruvate kinase (*PKM2*) may slow glycolysis and redirect carbohydrate intermediates to e.g. the pentose phosphate pathway (PPP). This is supported by overexpression of key enzymes both in the oxidative and the reductive part of the PPP, specifically the expression of 6-phosphogluconolactonase (*PGLS*), transketolase (*TKT*), and ribokinase (*RBKS*) (Supplementary Table S5). Collectively, these results suggest that glucose may be shunted into the PPP among ERG$_{high}$ samples. The PPP provides nucleotide precursors and helps regenerate NADPH which is important for maintaining the redox state and for supporting the synthesis of fatty acids for cancer cells [39].

When investigating the most central genes associated with the metabolism of glycine we did not find any possible explanation for the increased levels of glycine among ERG$_{high}$ compared to ERG$_{low}$ samples (Supplementary Table S5).

Risk stratification based on the presence of TMPRSS2-ERG

Risk stratification for choice of treatment in low grade prostate cancer is currently a challenge. We therefore investigated the possibility to stratify patients according to the presence of fusion/ERG
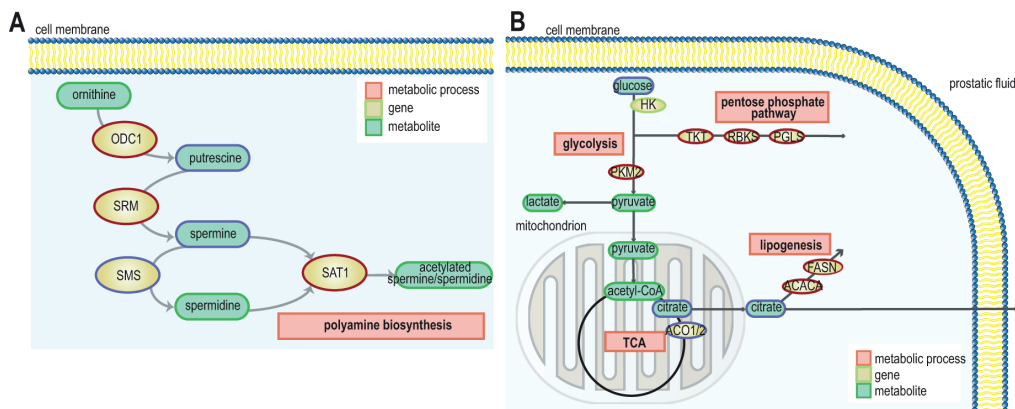
Figure 3: Schematic representation of pathways and gene expression levels of associated key enzymes altered due to presence of the fusion gene (**A**) the polyamine pathway, gene/protein names: ODC1: ornithine decarboxylase 1, SRM: spermidine synthase, SMS: spermine synthase, SAT1: spermidine/spermine N1-acetyltransferase 1 (blue = down-regulation, red = up-regulation) and (**B**) TCA cycle, fatty acid synthesis and pentose phosphate pathway. ACO1/2: aconitase 1/2, ACACA: acetyl-CoA carboxylase alpha, FASN: fatty acid synthase, HK: hexokinase, PKM2: pyruvate kinase, PGLS: 6-phosphogluconolactonase, RBKS: ribokinase, and TKT: transketolase (blue = down-regulation, red = up-regulation).

status within low Gleason samples (Gleason score $\leq 3 + 4$). We detected pronounced differences both in metabolism and gene expression levels between ERG_{high} and ERG_{low}, restricted to low Gleason samples (Supplementary Table S8). A significant decrease in the concentrations of citrate, spermine, putrescine and glycerophosphoethanolamine were detected in ERG_{high} samples compared to ERG_{low} samples, and a significant increase were found in glutamine and glycine after multiple testing corrections (Supplementary Table S8). However, restricting our analyses to high Gleason samples (Gleason score $\geq 4 + 3$), only significantly decreased concentrations of citrate and spermine were observed comparing ERG_{high} and ERG_{low} (Supplementary Table S9). Expression levels of key enzymes in the metabolism of the polyamines, glucose and fatty acid displayed higher significance levels when restricting the analyses to low Gleason samples compared to high Gleason samples (Supplementary Table S6 and S7).

As both the metabolic and the gene expression levels were more pronounced in the low Gleason group, presence of the fusion gene may serve as a tool for risk- or treatment stratification of low Gleason patients. In high Gleason samples, we generally observed less significant metabolic and transcriptomic alterations due to ERG status. High Gleason score has been linked to genomic instability and multiple genetic alterations [40, 41]. As the high Gleason samples are heterogeneous, the transcriptomic- and metabolic differences between ERG_{high} and ERG_{low} may possibly be masked by the effect of other genetic alterations present among these samples.

To increase the understanding of metabolism associated with the presence of the fusion gene, INMEX and ssGSEA analyses were performed, and indicated several metabolic pathway differences between ERG_{high} and ERG_{low} (Figure 4A) including glutathione (including polyamines), glycolysis, and additionally purine and pyrimidine, which are important precursors for nucleotides (Supplementary Table S10). In concordance with our findings in metabolic concentrations, both INMEX and GSEA showed more significant differences when the analyses were restricted to low Gleason samples (Figure 4B) than to high Gleason samples (Figure 4C). These results are presented in Supplementary Tables S11–S18.

In conclusion, metabolic alterations in the presence of the fusion gene are more pronounced in the low grade compared to aggressive cancer, and may be suggested as a possible risk stratification tool for low Gleason prostate cancer patients. Metabolism suggests a more aggressive phenotype connected to presence of the fusion gene. However, further studies on prognostics and validation are needed. Due to the small number of samples in the validation cohort, metabolic differences between low- and high Gleason samples could not be validated by this cohort.

Biochemical recurrence and ERG status

Prognostics and biochemical recurrence connected to presence of the fusion gene have previously shown varying results [14]. At a median follow-up of 6.5 years in our study (range 1.8 to 8.3 years), 10 (33.3%) of the 30 patients with follow-up data had experienced biochemical recurrence (prostate-specific antigen (PSA) ≥ 0.2 ng/ml). No significant difference was observed in biochemical recurrence between ERG_{high} and ERG_{low} patients in the main cohort (Figure 4D–4F), which is in agreement with other studies on radical prostatectomy cohorts [6, 11–14]. However, there was a trend towards significance when restricting to the low Gleason patients, $p = 0.205$ (Figure 4E). Due to the low number of included patients, the current study may not have the sufficient statistical power to reveal significant differences in rate of biochemical recurrence between ERG_{high} and ERG_{low}.

Translational potential by *in vivo* patient magnetic resonance spectroscopy imaging

The potential of transferring biomarkers and knowledge to 3T and 7T *in vivo* patient MRSI [16], makes HR-MAS MRS on prostate tissue samples attractive for basic research. A subset of the patients in the main cohort (9 patients, 21 samples) had data from *in vivo* MRSI acquired prior to surgery. The *in vivo* spectroscopy voxels were spatially matched to the HRMAS tissue samples [16]. The *in vivo* citrate/creatine ratio from spatially matched voxels was decreased with borderline significance in ERG_{high} compared to ERG_{low}, $p = 0.083$ (Figure 5A), while choline/creatine and spermine/creatine ratios were not significant, $p = 0.667$ and $p = 0.158$, respectively (Figure 5B). However, in the low Gleason group (5 patients, 11 samples), the citrate/creatine ratio was significantly decreased, $p < 0.001$ (Figure 5C) and in addition, the levels of choline/creatine ratio was significantly increased ($p = 0.041$), while the spermine/creatine ratio was borderline decreased ($p = 0.094$) in ERG_{high} compared to ERG_{low} (Figure 5D). Within high Gleason samples (4 patients, 10 samples), only a decreased choline/creatine ratio was significantly detected ($p = 0.018$), comparing ERG_{high} to ERG_{low}. Alterations in MRSI *in vivo* measured citrate, choline and spermine levels may offer a possibility for stratification of low risk prostate cancer patients without the need of biopsies, and a possibility to enroll patients into active surveillance programs, with non-invasive MRSI monitoring.

Concluding remarks

This study presents a distinct metabolic profile with concordant alterations of gene expression levels of key metabolic enzymes in prostate tissue samples with the presence of TMPRSS2-ERG. The metabolic profile was especially connected to the metabolism of polyamines and citrate, but also glycolysis and fatty acid metabolism. Our results indicate that TMPRSS2-ERG differentiates towards a phenotype that is associated with characteristics of an aggressive phenotype of prostate cancer. Additionally, the observed metabolic alterations can be translated to *in vivo* patient MRSI.

Figure 4: Single sample GSEA (ssGSEA) show alterations of several key metabolic pathways and biochemical recurrence for ERGhigh and ERGlow and Kaplan-Maier plots for biochemical recurrence. ssGSEA results of key metabolic pathways for the main cohort (**A**) the low Gleason samples (**B**) and the low Gleason samples within the main cohort (**C**) where red indicates up-regulation of the given gene set and blue down-regulation of a given gene set. Pathways with significantly different ($p < 0.05$) ssGSEA-values comparing ERGhigh and ERGlow while adjusting for multiple samples per patient, are indicated in bold. Kaplan-Maier plots for ERGhigh and ERGlow in the main cohort (**D**), low Gleason patients (**E**) and high Gleason patients (**F**).

Figure 5: Box-plots for citrate and spermine comparing ERGhigh and ERGlow samples in the main cohort (*in vivo*).
(**A**) No significant difference in levels of citrate/creatine in the main cohort (*in vivo* MRSI) between ERGhigh and ERGlow, $p = 0.083$, (**B**) Decreased levels of citrate/creatine were found comparing ERGhigh to ERGlow in the main cohort (*in vivo* MRSI), restricting to low Gleason samples, $p < 0.001$, (**C**) No significant difference in levels of spermine/creatine between ERGhigh and ERGlow in the main cohort (*in vivo* MRSI), $p = 0.158$, (**D**) No significant difference in levels of spermine/creatine were found comparing ERGhigh and ERGlow in the main cohort (*in vivo* MRSI), restricting to low Gleason samples, $p = 0.094$.

MATERIALS AND METHODS

Samples and patient cohorts

In the main cohort, a 2 mm transversal prostate tissue slice was collected from 48 prostate cancer patients after radical prostatectomy at St.Olavs Hospital, Trondheim, between 2007 and February 2010, with no previous treatment for prostate cancer, using a highly standardized harvesting method thoroughly described by Bertilsson et al. [15, 19]. From each tissue slice, several samples (average: 6, range: 4 to 11 per slice, depending on tumor size) were collected from cancerous and adjacent benign areas. In total 362 samples were extracted for RNA and acceptable RNA integrity number (RIN) scores were obtained from 354 samples. Samples with a high Gleason score, large extent of cancer and high quality RNA were prioritized. Seven patients were excluded either due to lack of cancer in the extracted samples (2 slices) or lack of quality of the samples in the microarray analyses (5 slices), and 4 samples were excluded due to low HR- MAS spectral quality. In total 95 cancer and 34 benign samples from 41 patients (median: 3, range: 1 to 6 per slice) were collected.

For validation of the results, a second cohort of 90 prostate cancer patients was included, consisting of one needle biopsy sample per patient obtained after radical prostatectomy. The samples were selected from a large biobank (~1000 patients, ~2000 samples) collected from prostate cancer patients after radical prostatectomy at St.Olavs Hospital, Trondheim, between 2007 and February 2010. The samples were selected from patients with highest tumor volume in order to collect tissue with high cancer content. The patients had not received any treatment for prostate cancer prior to sampling. Only 40 of the needle biopsies contained cancer and were included in the present study. The two cohorts were independent, i.e. no patient belonged to both cohorts. Sample characteristics for the main cohort and the validation cohort are given in Table 1. Both cohorts are approved by the Regional Committee of Medical and Health Research Ethics (REC), Central Norway, and all patients gave written, informed consent.

HR-MAS MRS

^{1}H HR-MAS MRS analysis was performed as previously described [15, 19]. Acquired spectral data were exponential Fourier transformed (line broadening 0.3 Hz), baseline- and phase-corrected using Topspin 3.2 (Bruker Biospin, Germany). Samples in the validation cohort were of equivalent weight (mean weight: 12.3 mg, range 6.7 to 21.9 mg) to samples in the main cohort (mean weight: 12.7 mg, range: 3.0 to 21.9 mg). Samples were hematoxylin-and eosin stained (HE, main cohort) or hematoxylin-eosin-saffron stained (HES, validation cohort) due to different routines in staining protocols at different times. HE/HES stained sections were used for histopathological evaluation of Gleason grading and assessment of cancer-, benign epithelial-, and stromal content. Two pathologists (TV and ER) evaluated the sections from the main cohort and an interrater agreement (κ) of 0.66, indicating substantial agreement, was found for distinguishing the samples into benign, low Gleason (Gleason score $\leq$ 3 + 4) and high Gleason (Gleason score $\geq$ 4 + 3). The first reading (TV) was used for grading in this study due to a slight degradation of the cryosections from the initial reading to the second reading. The validation cohort sections were evaluated by one pathologist (TV).

Definition of ERG groups and combining transcriptomics and metabolomics data

Gene expression profiles from the main cohort were obtained as previously described by Bertilsson et al. [19, 20]. The microarray data has previously been published in Array Expression with access number: E-MTAB-1041. The gene expression data were log2 transformed and quantile normalized [20]. Gene Set Enrichment Analysis (GSEA) scores were calculated for detection of specific enrichment of the ERG-fusion gene set based on prostate cancer related gene sets [42] as previously described by Rye et al. [43]. GSEA focuses on gene sets, i.e. groups of genes that share common biological function, chromosomal location, or regulation and in order to detect pathway changes more sensitively [44]. Based on the overall ERG GSEA score the samples were classified as ERG_{high} if the score were increased two-fold compared to the mean ERG GSEA of the cancer samples. The rest of the cancer samples were equally divided into groups of ERG_{low} and $ERG_{intermediate}$. The ERG_{high} samples were defined as possessing the highest probability for being fusion positive, while the ERG_{low} samples were defined as having the lowest probability for being fusion positive. Due to uncertainties of the fusion status of the $ERG_{intermediate}$ group, most of the differential analyses of metabolite and gene expression levels were performed comparing the ERG_{high} and ERG_{low} groups. Classification of samples per patient to the individual ERG groups is presented in Supplementary Table S1.

To link transcriptomic and metabolomics data connected to ERG status, we used two approaches; 1) integrative meta-analysis of expression data (INMEX) where lists of genes and metabolites (individually analyzed) are combined and significant genes and metabolites are mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [45], and where enrichment and topology analysis identify important pathways [46], 2) single sample GSEA (ssGSEA) [44] calculates separate enrichment scores for each pairing of a sample and gene set, which represents the degree of up- or down-regulation of a gene [47]. Enrichment of

KEGG gene set collections in the Molecular Signatures Database (Broad Institute, version 5.0) were performed using the GSEA software (Broad Institute, version 2.0.14) [44, 48]. ssGSEA analyses were performed using ssGSEAprojection [47], and results from the 38 most relevant metabolic pathways were visualized using HeatMapViewer (version 13), using GenePattern (version 3.9.4) [49].

Fluorescence *in situ* hybridization

The TMPRSS2-ERG status of samples in the validation cohort was determined by using a break-apart assay with a triple-labeled color commercial probe (Kreatech Diagnostics, The Netherlands). The probe detects the deletion between TMPRSS2 and ERG at 21q22. The FISH assay was carried out on 4 μm formalin-fixed, paraffin-embedded tissue sections after deparaffinization which were then pretreated using a commercial tissue section kit for paraffin-embedded tissue (Histology FISH Accessory Kit, Dako). The probe mix was applied and denatured at 80°C for 5 minutes before hybridization at 37°C overnight using a Dako hybridizer. The slides were counterstained with DAPI (4',6-diamidino-2-phenylindole) from the Histology FISH Accessory kit. Results were visualized using a 100x oil immersion objective on a Nikon Eclipse 90i fluorescent microscope (Nikon Corp., Japan) equipped with appropriate filters. For each sample, 25 non-overlapping nuclei in cancer areas were evaluated for deletion of the TMPRSS2 (21q22) gene region associated with TMPRSS2-ERG. In order to compensate for nuclear truncation, the cut-off level for an informative result was defined as loss of the TMPRSS2 (21q22) gene region at least 80% of tumor cell nuclei.

Luminal space measurements

Cryosections from the main cohort and the paraffin-embedded sections from the validation cohort were digitalized with 40x magnification and the luminal spaces were identified using a color-based segmentation (positive pixel count algorithm in ImageScope v8.0, Aperio Technologies) as described by Langer et al. [50].

Quantification of metabolites

Individual metabolites in the HR-MAS spectra were quantified using LCModel [51] based on a basis set containing 23 metabolites generated using NMRSIM (Bruker BioSpin, Germany) as previously described by Giskeødegård et al. [15]. Similarly, a separate basis set of 25 metabolites was built for the validation cohort, adding glutathione and ascorbate to the basis set as improvements of the previous basis set. In both cohorts, metabolites were quantified according to known amounts of formate and reported as mmoles/kg wet weight.

In vivo magnetic resonance spectroscopy imaging

As part of a previous published study [16], 9 patients (24 samples) in the main cohort had *in vivo* MRSI metabolic data from patients from spatially matched voxels to the tissue sampling sites. Due to the low number of samples, the samples were divided into two equal groups: ERG$_{high}$ for samples with ERG score higher than the median of the cancer samples, and ERG$_{low}$ for the samples with ERG score lower than median. Three samples were excluded due to low spectral quality of the associated MRSI spectrum. Details regarding e.g. acquisition and quantification of *in vivo* metabolite levels have previously been described in Selnæs et al. [16].

Statistical analysis

The HR-MAS MRS spectra were baseline corrected and peak aligned using icoshift [52] in MATLAB r2013a (The Mathworks, Inc., USA). Contamination signals from ethanol (3.65-3.69 ppm) were removed before normalization by probabilistic quotient normalization (PQN) [53]. Principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were performed on the Carr-Purcell-Meiboom-Gill (CPMG) spectra between 1.46 and 4.66 ppm. Data were centered prior to analysis. To avoid overfitting, PLS-DA models were validated through a 5-fold random subset cross-validation, and repeated 10 times. The number of latent variables was chosen based on the first local minima of cross-validated classification error for PLS-DA. Permutation testing was performed to assess the significance of the multivariate models ($n = 1000$). PCA and PLS-DA models were built using mixOmics in R [54] and PLS_toolbox 7.8.2 (Eigenvector Research, Inc., USA) in MATLAB, respectively.

In the main cohort, comparisons of quantified metabolites and gene expression levels, including metabolite levels from *in vivo* MRSI were performed by using linear mixed models in Stata 13 (StataCorp, USA), accounting for the effect of several samples originating from the same patient. Gene expression levels for metabolic enzymes were mainly chosen according to their proximity and influence of the quantified metabolites found within KEGG pathways. For the polyamine pathway, genes previous reported to be central in polyamine metabolism, provided as the basis for the analyses [26]. A total of 63 genes were included in the study, and are listed as part of Supplementary Tables S5–S7. Comparisons of gene expression levels were performed using the most significant probe if several probes for the same gene were available. Adjusted linear mixed models were built by including the relative amount of stroma, benign epithelia, cancer tissue and luminal spaces as continuous covariates, in order to minimize the possible confounding effects of tissue heterogeneity.

Adjusted models for gene expression data are presented in Supplementary Tables S5–S7, while adjusted models for metabolic data are presented in Supplementary Tables S2, S8 and S9. Test for trends of metabolite levels over ERG groups were performed using the nptrend function in Stata. In the validation cohort, comparisons of metabolite levels between TMPRSS2-ERG positive and negative samples were compared using Student t-test. P-values less than 0.05 were considered significant and q-values less than 0.05 were considered significant after corrections for multiple testing

Correlations between individual metabolites and tissue composition and relative luminal space were examined using Pearson's correlation, and correlations are presented in Supplementary Table S19. Corrections for multiple testing were performed by Benjamini-Hochberg correction. Multiple testing corrections were performed individually for the main- and the validation cohort, accounting for the number of comparisons for the metabolic data and the gene expression data individually. Prior to statistical analysis, all metabolite concentrations were transformed in order to obtain normal-distributed data or residuals. Type of transformation performed was based on visual inspections of resulting QQ-plots and histograms of the transformed data. Metabolic data were in general square-root transformed, except lactate which was log-transformed and glycine which was transformed by 1 divided by the square-root.

Differences in rates of biochemical recurrence (PSA ≥ 0.2 ng/ml) after prostatectomy were estimated with the Kaplan-Meier method and compared using the log rank statistics and the Cox proportional hazards regression model. Patients were classified as ERG_{high} if they had one or more samples within ERG_{high}. Patients were followed from date of surgery until last measured PSA or death. Time to event was calculated as the time in months between date of surgery and date of PSA-blood collection indicating biochemical recurrence or date of last follow-up blood collection. In total, 30 patients classified as ERG_{high} or ERG_{low} were included in the analysis, while patients with only benign or $ERG_{intermediate}$ samples were excluded from the analysis.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

None.

FINANCIAL SUPPORT

REFERENCES

1. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, Sboner A, Esgueva R, Pflueger D, Sougnez C, Onofrio R, Carter SL, Park K, et al. The genomic complexity of primary human prostate cancer. Nature. 2011; 470:214–220.

2. Esgueva R, Perner S, LaFargue CJ, Scheble V, Stephan C, Lein M, Fritzsche FR, Dietel M, Kristiansen G, Rubin MA. Prevalence of TMPRSS2-ERG and SLC45A3-ERG gene fusions in a large prostatectomy cohort. Mod Pathol. 2010; 23:539–546.

3. Tomlins SA, Bjartell A, Chinnaiyan AM, Jenster G, Nam RK, Rubin MA, Schalken JA. ETS Gene Fusions in Prostate Cancer: From Discovery to Daily Clinical Practice. Eur Urol. 2009; 56:275–286.

4. FitzGerald L, Agalliu I, Johnson K, Miller M, Kwon E, Hurtado-Coll A, Fazli L, Rajput A, Gleave M, Cox M, Ostrander E, Stanford J, Huntsman D. Association of TMPRSS2-ERG gene fusion with clinical characteristics and outcomes: results from a population-based study of prostate cancer. BMC Cancer. 2008; 8:1–10.

5. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun X-W, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, et al. Recurrent Fusion of TMPRSS2 and ETS Transcription Factor Genes in Prostate Cancer. Science. 2005; 310:644–648.

6. Steurer S, Mayer PS, Adam M, Krohn A, Koop C, Ospina-Klinck D, Tehrani AA, Simon R, Tennstedt P, Graefen M, Wittmer C, Brors B, Plass C, et al. TMPRSS2-ERG Fusions Are Strongly Linked to Young Patient Age in Low-grade Prostate Cancer. Eur Urol. 2014; 66:978–981.

7. Demichelis F, Fall K, Perner S, Andren O, Schmidt F, Setlur SR, Hoshida Y, Mosquera JM, Pawitan Y, Lee C, Adami HO, Mucci LA, Kantoff PW, et al. TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. Oncogene. 2007; 26:4596–4599.

8. Hägglöf C, Hammarsten P, Strömvall K, Egevad L, Josefsson A, Stattin P, Granfors T, Bergh A. TMPRSS2-ERG Expression Predicts Prostate Cancer Survival and Associates with Stromal Biomarkers. PLoS ONE. 2014; 9:e86824.

9. Qi M, Yang X, Zhang F, Lin T, Sun X, Li Y, Yuan H, Ren Y, Zhang J, Qin X, Han B. ERG Rearrangement Is Associated with Prostate Cancer-Related Death in Chinese Prostate Cancer Patients. PLoS ONE. 2014; 9:e84959.

10. Berg KD, Vainer B, Thomsen FB, Røder MA, Gerds TA, Toft BG, Brasso K, Iversen P. ERG Protein Expression in Diagnostic Specimens Is Associated with Increased Risk of Progression During Active Surveillance for Prostate Cancer. Eur Urol. 2014; 66:851–860.

11. Gopalan A, Leversha MA, Satagopan JM, Zhou Q, Al-Ahmadie HA, Fine SW, Eastham JA, Scardino PT, Scher HI, Tickoo SK, Reuter VE, Gerald WL. TMPRSS2-ERG Gene Fusion Is Not Associated with Outcome in Patients Treated by Prostatectomy. Cancer Res. 2009; 69:1400–1406.

12. Hoogland AM, Jenster G, van Weerden WM, Trapman J, van der Kwast T, Roobol MJ, Schroder FH, Wildhagen MF and van Leenders GJLH. ERG immunohistochemistry is not predictive for PSA recurrence, local recurrence or overall survival after radical prostatectomy for prostate cancer. Mod Pathol. 2012; 25:471–479.

13. Minner S, Enodien M, Sirma H, Luebke AM, Krohn A, Mayer PS, Simon R, Tennstedt P, Müller J, Scholz L, Brase JC, Liu AY, Schlüter H, et al. ERG Status Is Unrelated to PSA Recurrence in Radically Operated Prostate Cancer in the Absence of Antihormonal Therapy. Clin Cancer Res. 2011; 17:5878–5888.

14. Pettersson A, Graff RE, Bauer SR, Pitt MJ, Lis RT, Stack EC, Martin NE, Kunz L, Penney KL, Ligon AH, Suppan C, Flavin R, Sesso HD, et al. The TMPRSS2:ERG Rearrangement, ERG Expression, and Prostate Cancer Outcomes: A Cohort Study and Meta-analysis. Cancer Epidemiol. Biomarkers Prev. 2012; 21:1497–1509.

15. Giskeødegård GF, Bertilsson H, Selnæs KM, Wright AJ, Bathen TF, Viset T, Halgunset J, Angelsen A, Gribbestad IS, Tessem MB. Spermine and Citrate as Metabolic Biomarkers for Assessing Prostate Cancer Aggressiveness. PLoS ONE. 2013; 8:e62375.

16. Selnæs KM, Gribbestad IS, Bertilsson H, Wright A, Angelsen A, Heerschap A, Tessem M-B. Spatially matched *in vivo* and *ex vivo* MR metabolic profiles of prostate cancer – investigation of a correlation with Gleason score. NMR Biomed. 2013; 26:600–606.

17. Meller S, Meyer H-A, Bethan B, Dietrich D, Maldonado SG, Lein M, Montani M, Reszka R, Schatz P, Peter E, Stephan C, Jung K, Kamlage B, et al. Integration of tissue metabolomics, transcriptomics and immunohistochemistry reveals ERG- and gleason score- specific metabolomic alterations in prostate cancer. Oncotarget. 2015; 7:1421–1438. doi: 10.18632/oncotarget.6370.

18. Massoner P, Kugler KG, Unterberger K, Kuner R, Mueller LAJ, Fälth M, Schäfer G, Seifarth C, Ecker S, Verdorfer I, Graber A, Sültmann H, Klocker H. Characterization of Transcriptional Changes in ERG Rearrangement-Positive Prostate Cancer Identifies the Regulation of Metabolic Sensors Such as Neuropeptide Y. PLoS ONE. 2013; 8:e55207.

19. Bertilsson H, Angelsen A, Viset T, Skogseth H, Tessem MB, Halgunset J. A new method to provide a fresh frozen prostate slice suitable for gene expression study and MR spectroscopy. Prostate. 2011; 71:461–469.

20. Bertilsson H, Tessem M-B, Flatberg A, Viset T, Gribbestad I, Angelsen A, Halgunset J. Changes in Gene Transcription Underlying the Aberrant Citrate and Choline Metabolism in Human Prostate Cancer Samples. Clin. Cancer Res. 2012; 18:3261–3269.

21. Mertz KD, Horcic M, Hailemariam S, D'Antonio A, Dirnhofer S, Hartmann A, Agaimy A, Eppenberger-Castori S, Obermann E, Cathomas G, Bubendorf L. Heterogeneity of ERG expression in core needle biopsies of patients with early prostate cancer. Hum Pathol. 2013; 44:2727–2735.

22. Minner S, Gartner M, Freudenthaler F, Bauer M, Kluth M, Salomon G, Heinzer H, Graefen M, Bokemeyer C, Simon R, Sauter G, Schlomm T, Wilczak W. Marked heterogeneity of ERG expression in large primary prostate cancers. Mod Pathol. 2013; 26:106–116.

23. Swanson MG, Zektzer AS, Tabatabai ZL, Simko J, Jarso S, Keshari KR, Schmitt L, Carroll PR, Shinohara K, Vigneron DB, Kurhanewicz J. Quantitative analysis of prostate metabolites using 1H HR-MAS spectroscopy. Magn Reson Med. 2006; 55:1257–1264.

24. Trock BJ. Application of Metabolomics to Prostate Cancer. Urol Oncol. 2011; 29:572–581.

25. Moinard C, Cynober L, de Bandt JP. Polyamines: metabolism and implications in human diseases. Clin Nutr. 2005; 24:184–197.

26. Thomas T, Thomas TJ. Polyamine metabolism and cancer. J Cell Mol Med. 2003; 7:113–126.

27. Ridgway ND. The role of phosphatidylcholine and choline metabolites to cell proliferation and survival. Crit Rev Biochem Mol Biol. 2013; 48:20–38.

28. Bathen TF, Sitter B, Sjøbakk TE, Tessem MB, Gribbestad IS. Magnetic Resonance Metabolomics of Intact Tissue: A Biotechnological Tool in Cancer Diagnostics and Treatment Evaluation. Cancer Res. 2010; 70:6692–6696.

29. Giskeødegård GF, Lundgren S, Sitter B, Fjøsne HE, Postma G, Buydens LMC, Gribbestad IS, Bathen TF. Lactate and glycine-potential MR biomarkers of prognosis in estrogen receptor-positive breast cancers. NMR Biomed. 2012; 25:1271–1279.

30. Tessem M-B, Bertilsson H, Angelsen A, Bathen TF, Drabløs F, Rye MB. A Balanced Tissue Composition Reveals New Metabolic and Gene Expression Markers in Prostate Cancer. PLoS ONE. 2016; 11:e0153727.

31. Mandal S, Mandal A, Johansson HE, Orjalo AV, Park MH. Depletion of cellular polyamines, spermidine and spermine, causes a total arrest in translation and growth in mammalian cells. PNAS. 2013; 110:2169–2174.

32. Mohan RR, Challa A, Gupta S, Bostwick DG, Ahmad N, Agarwal R, Marengo SR, Amini SB, Paras F, MacLennan GT, Resnick MI, Mukhtar H. Overexpression of Ornithine Decarboxylase in Prostate Cancer and Prostatic Fluid in Humans. Clin Cancer Res. 1999; 5:143–147.

33. Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM. Meta-Analysis of Microarrays: Interstudy Validation of Gene Expression Profiles Reveals Pathway Dysregulation in Prostate Cancer. Cancer Res. 2002; 62:4427–4433.

34. Costello LC, Franklin RB. Bioenergetic theory of prostate malignancy. Prostate. 1994; 25:162–166.

35. Juang HH. Modulation of mitochondrial aconitase on the bioenergy of human prostate carcinoma cells. Mol Genet Metab. 2004; 81:244–252.

36. Chang YS, Tsai CT, Huangfu CA, Huang WY, Lei HY, Lin CF, Su IJ, Chang WT, Wu PH, Chen YT, Hung JH, Young KC, Lai MD. ACSL3 and GSK-3β are essential for lipid upregulation induced by endoplasmic reticulum stress in liver cells. J Cell. Biochem. 2011; 112:881–893.

37. Menendez JA, Lupu R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. Nat Rev Cancer. 2007; 7:763–777.

38. Wu X, Daniels G, Lee P, Monaco ME. Lipid metabolism in prostate cancer. Am J Clin Exp Urol. 2014; 2:111–120.

39. Patra KC, Hay N. The pentose phosphate pathway and cancer. Trends Biochem Sci. 2014; 39:347–354.

40. Fraser M, Berlin A, Bristow RG and van der Kwast T. Genomic, pathological, and clinical heterogeneity as drivers of personalized medicine in prostate cancer. Urol Oncol Semin Ori. 2015; 33:85–94.

41. Tapia-Laliena MA, Korzeniewski N, Hohenfellner M, Duensing S. High-risk prostate cancer: A disease of genomic instability. Urol Oncol Semin Ori. 2014; 32:1101–1107.

42. Markert EK, Mizuno H, Vazquez A, Levine AJ. Molecular classification of prostate cancer using curated expression signatures. PNAS. 2011; 108:21276–21281.

43. Rye M, Bertilsson H, Drablos F, Angelsen A, Bathen T, Tessem M-B. Gene signatures ESC, MYC and ERG-fusion are early markers of a potentially dangerous subtype of prostate cancer. BMC Med Genomics. 2014; 7:50.

44. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS. 2005; 102:15545–15550.

45. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Research. 2014; 42:199–205.

46. Xia J, Fjell CD, Mayer ML, Pena OM, Wishart DS, Hancock REW. INMEX—a web-based tool for integrative meta-analysis of expression data. Nucleic Acids Resa. 2013; 41:63–70.

47. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, Frohling S, Chan EM, Sos ML, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009; 462:108–112.

48. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003; 34:267–273.

49. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. Nat Genet. 2006; 38:500–501.

50. Langer DL, van der Kwast TH, Evans AJ, Plotkin A, Trachtenberg J, Wilson BC, Haider MA. Prostate Tissue Composition and MR Measurements: Investigating the Relationships between ADC, T2, Ktrans, ve, and Corresponding Histologic Features. Radiology. 2010; 255:485–494.

51. Provencher SW. Estimation of metabolite concentrations from localized *in vivo* proton NMR spectra. Magn Reson Med. 1993; 30:672–679.

52. Savorani F, Tomasi G, Engelsen SB. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. Magn Reson Med. 2010; 202:190–202.

53. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics. Anal Chem. 2006; 78:4281–4290.

54. Lê Cao K-A, González I, Déjean S. integrOmics: an R package to unravel relationships between two omics datasets. Bioinformatics. 2009; 25:2855–2856.

Presence of TMPRSS2-ERG is associated with alterations of the metabolic profile in human prostate cancer

Supplementary Materials

Supplementary Table S1: Sample distribution per patient in the main cohort classified as ERG_{low}, $ERG_{intermediate}$ and ERG_{high}

Patient	Benign	ERG_{low}	$ERG_{intermediate}$	ERG_{high}	Total
1	1	0	0	2	3
2	1	0	3	1	5
3	1	0	0	2	3
4	1	0	1	1	3
5	1	0	1	1	3
6	0	1	1	2	4
7	1	0	0	2	3
8	0	0	0	4	4
9	0	2	0	0	2
10	1	3	1	0	5
11	1	1	2	1	5
12	1	1	2	0	4
13	1	0	1	0	2
14	0	0	2	0	2
15	1	1	0	1	3
16	2	0	1	2	5
17	1	0	0	1	2
18	1	0	1	2	4
19	0	0	0	3	3
20	1	3	0	0	4
21	1	2	0	0	3
22	1	0	1	1	3
23	1	1	2	0	4
24	0	0	0	2	2
25	1	0	0	0	1
26	1	1	0	0	2
27	1	1	0	0	2
28	1	0	0	0	1
29	1	1	1	0	3
30	2	2	0	0	4
31	1	1	0	0	2
32	1	1	0	0	2

33	1	0	0	0	1
34	0	1	2	0	3
35	0	1	1	3	5
36	2	1	0	0	3
37	1	2	2	1	6
38	0	0	2	1	3
39	1	1	1	1	4
40	1	2	0	0	3
41	0	0	3	0	3
Total	34	30	31	34	129

Supplementary Table S2: Differences in levels of quantified metabolites in the main cohort comparing ERG$_{high}$ samples to ERG$_{low}$ samples

Metabolites	Benign n = 34 Concentrations mmoles/kg wet weight, median (IQR)	ERG$_{low}$ n = 30 Concentrations mmoles/kg wet weight, median (IQR)	ERG$_{intermediate}$ n = 31 Concentrations mmoles/kg wet weight, median (IQR)	ERG$_{high}$ n = 34 Concentrations mmoles/kg wet weight, median (IQR)	ERG$_{high}$ vs ERG$_{low}$ p-values	ERG$_{high}$ vs ERG$_{low}$ p-values*	p-trend**	Corrected for stromal content	Corrected for cancer content	Corrected for benign epithelial content	Corrected for luminal space content
Alanine	1.74 (1.32 to 2.09)	2.18 (1.75 to 2.79)	2.16 (1.65 to 2.57)	2.48 (1.78 to 2.85)	0.711	0.866	0.611	0.876	0.929	0.744	0.666
Choline	0.45 (0.28 to 0.61)	0.99 (0.59 to 1.43)	1.01 (0.65 to 1.68)	1.23 (0.89 to 1.51)	0.126	0.29	0.116	0.096	0.102	0.043	0.038
Citrate	8.45 (4.62 to 11.85)	9.44 (5.56 to 14.68)	6.74 (3.94 to 10.34)	3.91 (2.59 to 7.20)	<0.001	<0.001	<0.001	5.60E-06	2.20E-05	6.10E-05	5.80E-05
Creatine	2.51 (1.97 to 3.06)	2.27 (1.65 to 2.70)	2.10 (1.82 to 2.55)	2.00 (1.64 to 2.54)	0.953	1	0.546	0.913	0.817	0.666	0.729
Ethm	0 (0 to 0.18)	0 (0 to 0)	0 (0 to 0.15)	0.01 (0 to 0.31)	0.49	0.663	0.043	0.421	0.352	0.238	0.326
Glucose	0.78 (0.47 to 1.25)	0.15 (0.00 to 0.63)	0.00 (0.00 to 0.43)	0.00 (0.00 to 0.07)	0.008	0.061	0.007	0.012	0.021	0.014	6.40E-03
Glutamate	2.73 (2.42 to 3.53)	4.60 (4.02 to 6.11)	4.75 (3.55 to 6.88)	5.98 (4.26 to 7.79)	0.139	0.291	0.101	0.133	0.244	0.116	0.049
Glutamine	2.05 (1.58 to 2.35)	2.57 (2.27 to 3.41)	2.82 (2.14 to 3.50)	3.05 (2.61 to 3.70)	0.422	0.647	0.12	0.485	0.281	0.413	0.3
Glycine	1.55 (1.28 to 1.98)	1.99 (1.68 to 2.78)	2.51 (1.98 to 3.18)	2.90 (1.93 to 3.65)	0.023	0.115	0.008	5.70E-03	1.00E-02	6.10E-03	4.00E-03
GPC	0.41 (0.25 to 0.50)	0.76 (0.49 to 1.28)	0.73 (0.43 to 1.41)	0.90 (0.60 to 1.17)	1	1	0.865	0.72	0.563	0.823	0.653
GPE	0.23 (0 to 0.43)	0.17 (0 to 0.55)	0.03 (0 to 0.51)	0 (0 to 0.72)	0.999	1	0.94	0.913	0.862	0.909	0.831
Isoleucine	0.09 (0 to 0.12)	0.17 (0.03 to 0.24)	0.15 (0.06 to 0.29)	0.19 (0.11 to 0.31)	0.097	0.248	0.138	0.024	0.014	0.022	0.05
Lactate	13.93 (10.35 to 16.54)	21.92 (16.52 to 25.63)	20.52 (13.61 to 26.41)	17.34 (15.69 to 22.01)	0.34	0.602	0.166	0.148	0.134	0.178	0.174
Leucine	0.24 (0.16 to 0.36)	0.38 (0.30 to 0.60)	0.46 (0.32 to 0.70)	0.51 (0.37 to 0.70)	0.715	0.866	0.168	0.746	0.788	0.598	0.668
Myo-inositol	9.06 (8.01 to 10.67)	10.03 (6.92 to 11.67)	9.13 (7.16 to 12.66)	8.97 (7.23 to 10.56)	0.49	0.663	0.294	0.373	0.431	0.454	0.475
PCh	0.32 (0.16 to 0.60)	0.61 (0.33 to 0.91)	0.63 (0.43 to 1.17)	0.96 (0.64 to 1.36)	0.067	0.248	0.005	0.017	0.027	0.026	0.018
PE	1.73 (1.09 to 2.59)	2.33 (1.46 to 3.51)	2.57 (1.67 to 4.14)	2.86 (2.33 to 3.79)	0.087	0.248	0.032	0.02	0.029	0.042	0.045
Putrescine	0.44 (0 to 1.34)	0.12 (0 to 0.67)	0.06 (0 to 0.30)	0 (0 to 0.10)	0.025	0.115	0.003	2.20E-03	4.00E-03	7.70E-03	6.40E-03
Scyllo-inositol	0.33 (0.25 to 0.56)	0.39 (0.31 to 0.59)	0.50 (0.37 to 0.63)	0.47 (0.37 to 0.61)	0.087	0.248	0.438	4.50E-03	5.50E-03	0.029	0.067
Spermine	1.69 (0.90 to 2.71)	2.10 (1.20 to 3.19)	1.23 (0.79 to 2.02)	0.89 (0.45 to 1.40)	<0.001	<0.001	<0.001	7.20E-06	3.10E-05	9.90E-05	7.00E-05
Succinate	0.37 (0.30 to 0.47)	0.60 (0.43 to 0.74)	0.62 (0.44 to 0.89)	0.63 (0.49 to 0.88)	0.412	0.647	0.446	0.401	0.433	0.247	0.181
Taurine	5.74 (3.98 to 6.27)	5.39 (3.90 to 6.96)	4.84 (3.48 to 7.09)	4.26 (3.28 to 5.54)	0.168	0.322	0.097	0.117	0.167	0.14	0.071
Valine	0.21 (0.18 to 0.29)	0.35 (0.27 to 0.50)	0.39 (0.22 to 0.49)	0.39 (0.26 to 0.53)	0.972	1	0.666	0.975	0.983	0.847	0.999

Ethm: Ethanolamine, GPC: Glycerophosphocholine, GPE: Glycerophosphoethanolamine, PCh: Phosphocholine, PE: Phosphoethanolamine.

*Benjamini-Hochberg adjusted values, **trend across increasing ERG groups (not including benign samples).

Supplementary Table S3: Differences in levels of quantified metabolites in the main cohort, comparing all ERG groups to benign samples and ERG$_{low}$ to ERG$_{intermediate}$ and ERG$_{intermediate}$ to ERG$_{high}$

Metabolites	Benign vs ERG$_{low}$	Benign vs ERG$_{intermediate}$	Benign vs ERG$_{high}$	ERG$_{low}$ vs ERG$_{int}$	ERG$_{int}$ vs ERG$_{high}$
	p-value	*p*-value	*p*-value	*p*-value	*p*-value
Alanine	2.80E–03	0.044	5.30E–03	0.728	0.779
Choline	9.70E–06	1.70E–07	8.70E–09	0.035	0.687
Citrate	0.334	0.099	1.30E–04	0.015	0.19
Creatine	0.692	0.995	0.342	0.282	0.518
Ethm	0.161	0.476	0.739	0.356	0.633
Glucose	5.40E–06	4.50E–11	1.40E–14	0.094	0.374
Glutamate	4.30E–09	7.00E–10	4.50E–13	0.15	0.239
Glutamine	6.80E–04	1.90E–03	1.30E–06	0.332	0.381
Glycine	6.30E–03	1.60E–04	2.10E–07	0.091	0.39
GPC	1.10E–09	2.20E–07	1.50E–06	0.32	0.316
GPE	0.986	0.604	0.84	0.601	0.945
Isoleucine	0.109	0.015	3.60E–04	0.254	0.177
Lactate	8.70E–08	6.00E–06	1.00E–04	0.881	0.386
Leucine	2.50E–04	2.80E–07	1.60E–05	0.719	0.847
Myo–inositol	0.286	0.222	0.751	0.927	0.327
PCh	9.00E–03	1.80E–06	1.00E–08	0.077	0.273
PE	1.60E–03	4.40E–06	2.90E–08	0.238	0.501
Putrescine	0.261	7.80E–03	5.70E–05	0.03	0.939
Scyllo–inositol	0.372	0.608	0.688	0.963	0.492
Spermine	0.118	0.201	3.80E–05	0.035	0.086
Succinate	0.016	0.019	4.70E–06	0.243	0.838
Taurine	0.796	0.489	0.104	0.827	0.035
Valine	0.065	0.152	0.02	0.764	0.359

Ethm: Ethanolamine, GPC: Glycerophosphocholine, GPE: Glycerophosphoethanolamine.
GSH: Glutathione, PCh:Phosphocholine, PE: Phosphoethanolamine.

Supplementary Table S4: Differences in levels of quantified metabolites in the validation cohort comparing TMPRSS2–ERG positive and negative patients

Metabolites	TMPRSS2:ERG negative	TMPRSS2:ERG positive	*p*-value	*p*-value*
	n = 33	*n* = 7		
	Concentrations mmoles/ kg wet weight (IQR)	Concentrations mmoles/kg wet weight (IQR)		
Alanine	2.51 (2.07 to 2.88)	2.13 (1.59 to 2.51)	0.202	0.721
Ascorbate	0.06 (0.00 to 0.06)	0.08 (0.00 to 0.22)	0.487	0.84
Choline	0.96 (0.79 to 1.19)	0.79 (0.40 to 0.99)	0.148	0.617
Citrate	6.74 (3.88 to 8.74)	3.05 (1.25 to 4.69)	0.013	0.263
Creatine	3.01 (2.61 to 3.51)	2.69 (2.14 to 3.66)	0.293	0.775
Ethm	0.22 (0.00 to 0.36)	0.00 (0.00 to 0.00)	0.088	0.55
Glucose	0.39 (0.00 to 0.77)	0.55 (0.00 to 1.56)	0.805	0.957
Glutamate	4.08 (3.12 to 4.62)	3.94 (3.45 to 4.64)	0.832	0.957
Glutamine	2.09 (1.61 to 2.47)	2.09 (1.78 to 2.55)	0.919	0.957
Glycine	2.37 (1.92 to 2.86)	2.41 (1.54 to 2.64)	0.857	0.957
GPC	0.90 (0.52 to 1.14)	0.85 (0.48 to 1.21)	0.905	0.957
GPE	0.13 (0.00 to 0.16)	0.18 (0.00 to 0.07)	0.099	0.55
GSH	0.84 (0.74 to 1.04)	0.86 (0.64 to 1.13)	0.898	0.957
Isoleucine	0.17 (0.09 to 0.25)	0.13 (0.08 to 0.16)	0.413	0.794
Lactate	16.01 (12.74 to 18.12)	15.92 (12.34 to 20.40)	0.959	0.959
Leucine	0.36 (0.22 to 0.47)	0.40 (0.27 to 0.38)	0.571	0.892
Myo–inositol	8.82 (7.31 to 10.05)	8.10 (7.09 to 8.90)	0.504	0.84
PCh	0.62 (0.31 to 0.81)	0.85 (0.63 to 0.98)	0.11	0.55
PE	2.34 (1.82 to 2.78)	2.68 (2.18 to 2.80)	0.282	0.775
Putrescine	0.17 (0.00 to 0.27)	0.06 (0.00 to 0.00)	0.348	0.791
Scyllo–inositol	0.60 (0.38 to 0.78)	0.48 (0.34 to 0.47)	0.381	0.794
Spermine	0.69 (0.38 to 0.81)	0.31 (0.10 to 0.52)	**0.021**	0.263
Succinate	0.54 (0.40 to 0.69)	0.46 (0.37 to 0.57)	0.31	0.775
Taurine	5.81 (4.83 to 7.06)	5.66 (5.16 to 6.80)	0.869	0.957
Valine	0.48 (0.41 to 0.50)	0.46 (0.39 to 0.49)	0.81	0.957

Ethm: Ethanolamine, GPC: Glycerophosphocholine, GPE: Glycerophosphoethanolamine, GSH: Glutathione, PCh: Phosphocholine, PE: Phosphoethanolamine *Benjamini-Hochberg corrected for multiple testing.

Supplementary Table S5: Differences in expression levels of key metabolic genes comparing ERG$_{high}$ and ERG$_{low}$ in the main cohort, including estimated means and adjusted for benign epithelia, stroma, cancer and luminal space

Gene	Estimated/predicted means med 95 % CI		*p*-value	*p*-value*	Adjusted for benign epithelial content	Adjusted for stromal content	Adjusted for cancer content	Adjusted for luminal space content
	ERG$_{low}$	ERG$_{high}$						
ACACA	9.81 (9.67, 9.94)	10.10 (9.97, 10.24)	9.80E–04	0.003	2.50E–03	1.50E–03	3.10E–03	4.70E–04
ACO1	9.71 (9.61, 9.82)	9.40 (9.30, 9.51)	2.50E–06	6.09E–05	6.30E–06	6.00E–06	1.40E–05	2.60E–06
ACO2	8.06 (7.97, 8.15)	7.93 (7.85, 8.02)	0.036	0.065	0.031	0.013	0.014	0.046
ACSL1			0.683	0.683	0.715	0.593	0.606	0.64
ACSL3	10.78 (10.59, 10.97)	11.07 (10.88, 11.26)	5.80E–03	0.014	1.40E–03	4.80E–03	1.50E–03	8.80E–03
ACSL4	7.08 (6.95, 7.22)	6.82 (6.69, 6.95)	3.80E–03	0.010	6.90E–03	6.20E–03	0.012	4.00E–03
ACSL5	6.20 (5.85, 6.55)	6.67 (6.33, 7.01)	0.045	0.073	0.021	0.035	0.018	0.045
ACSL6			0.264	0.320	0.312	0.209	0.247	0.171
AGXT1			0.155	0.208	0.193	0.215	0.255	0.139
ALDOA	9.33 (9.15, 9.50)	8.79 (8.62, 8.96)	1.40E–05	1.89E–04	2.40E–05	2.20E–05	4.80E–05	1.90E–05
ALDOB	6.03 (5.85, 6.20)	6.37 (6.20, 6.55)	3.40E–03	0.009	3.40E–03	6.50E–03	8.40E–03	9.70E–04
ALDOC			0.38	0.420	0.43	0.546	0.636	0.311
AMD1	6.93 (6.73, 7.13)	7.28 (7.08, 7.48)	3.40E–03	0.009	1.70E–03	4.50E–03	2.80E–03	2.70E–03
CHKA	8.10 (7.99, 8.21)	8.42 (8.32, 8.53)	1.60E–05	1.89E–04	2.70E–05	2.00E–05	3.10E–05	2.50E–05
CS			0.148	0.203	0.175	0.144	0.161	0.12
DLD	7.83 (7.71, 7.94)	7.59 (7.47, 7.71)	3.40E–04	0.001	6.10E–04	4.60E–04	7.70E–04	8.40E–04
DLST			0.132	0.189	0.115	0.074	0.054	0.14
ENO1	11.39 (11.29, 11.49)	11.64 (11.55, 11.74)	8.10E–05	4.78E–04	1.70E–04	1.20E–04	2.10E–04	9.90E–05
FASN	11.62 (11.36, 11.87)	11.98 (11.72, 12.23)	0.015	0.030	0.014	0.017	0.015	6.80E–03
FH			0.549	0.567	0.58	0.365	0.361	0.853
G6PD	6.95 (6.86, 7.04)	7.08 (6.99, 7.17)	0.033	0.061	0.037	0.044	0.053	0.028
GAPDH			0.285	0.339	0.33	0.392	0.446	0.181
GPI			0.652	0.663	0.75	0.818	0.906	0.559
HK1			0.06	0.095	0.057	0.033	0.033	0.063
HK2			0.241	0.298	0.053	0.121	0.025	0.344
IDH1			0.324	0.378	0.415	0.323	0.394	0.157
IDH2			0.191	0.241	0.164	0.231	0.215	0.15
IDH3A			0.338	0.380	0.322	0.281	0.253	0.404
IDH3B	8.92 (8.84, 8.99)	8.74 (8.67, 8.82)	3.70E–04	0.001	4.40E–04	5.70E–04	6.90E–04	4.10E–04
IDH3G			0.119	0.179	0.179	0.17	0.279	0.081
LDHA			0.178	0.229	0.223	0.258	0.315	0.239
LDHB	9.41 (9.13, 9.68)	8.77 (8.50, 9.05)	1.90E–04	9.21E–04	4.00E–04	2.80E–04	5.20E–04	3.00E–06
MDH1	10.33 (10.22, 10.45)	10.19 (10.07, 10.30)	0.041	0.068	0.036	0.082	0.079	7.30E–03
MDH2	11.93 (11.86, 12.00)	12.10 (12.03, 12.16)	4.00E–05	3.60E–04	2.60E–05	6.90E–05	6.10E–05	2.80E–05

NPY	10.93 (10.18, 11.68)	13.24 (12.50, 13.98)	2.90E–06	6.09E–05	1.60E–06	3.30E–06	2.10E–06	6.10E–06
OAZ			0.17	0.223	0.177	0.174	0.091	0.587
ODC	10.75 (10.49, 11.01)	11.12 (10.85, 11.38)	0.012	0.024	0.013	0.032	0.04	7.50E–03
OGDH			0.075	0.115	0.056	0.092	0.079	0.071
OGDHL	6.13 (5.94, 6.32)	7.48 (7.29, 7.66)	7.70E–29	4.85E–27	2.50E–33	2.10E–28	2.80E–30	5.90E–28
PDHA	8.55 (8.47, 8.62)	8.69 (8.62, 8.76)	7.90E–03	0.018	0.011	9.70E–03	0.014	8.90E–03
PDHB			0.448	0.487	0.479	0.384	0.418	0.424
PGD	8.41 (8.29, 8.53)	8.66 (8.54, 8.77)	2.80E–04	0.001	4.30E–04	6.20E–04	1.20E–03	5.70E–04
PGK1	10.36 (10.23, 10.50)	10.04 (9.91, 10.17)	7.70E–05	4.78E–04	1.00E–04	6.10E–05	9.50E–05	3.40E–05
PGLS	8.69 (8.58, 8.79)	8.94 (8.84, 9.04)	9.10E–05	4.78E–04	1.80E–04	2.20E–04	4.40E–04	1.70E–05
PGLS	8.69 (8.59, 8.79)	8.94 (8.84, 9.04)	9.10E–05	4.78E–04	1.80E–04	2.20E–04	4.40E–04	1.70E–05
PGM1	9.56 (9.40, 9.73)	9.19 (9.03, 9.36)	1.10E–03	0.003	1.80E–03	1.70E–03	3.40E–03	1.20E–03
PKM2	9.20 (9.09, 9.31)	9.36 (9.26, 9.47)	0.027	0.052	0.027	0.028	0.027	0.025
RBKS	6.82 (6.73, 6.91)	7.05 (6.97, 7.14)	2.50E–04	0.001	6.10E–04	3.50E–04	8.10E–04	3.20E–04
RPE			0.335	0.380	0.332	0.244	0.24	0.443
RPIA	7.44 (7.30, 7.58)	7.62 (7.49, 7.76)	0.04	0.068	0.051	0.06	0.075	4.40E–03
SAT1	12.27 (12.12, 12.43)	12.64 (12.49, 12.80)	4.00E–04	0.001	1.30E–04	2.40E–04	9.80E–05	1.30E–03
SDHA	8.67 (8.59, 8.74)	8.55 (8.47, 8.62)	0.012	0.024	0.013	0.016	0.017	0.015
SDHB			0.548	0.567	0.492	0.528	0.584	0.609
SDHC			0.132	0.189	0.139	0.287	0.161	0.216
SDHD	7.59 (7.44, 7.75)	7.33 (7.17, 7.48)	0.011	0.024	0.012	0.013	0.311	6.50E–03
SHMT1	7.01 (6.9, 7.2)	7.31 (7.16, 7.46)	0.003	0.009	0.003	0.001	0.002	2.00E–03
SMOX	7.15 (6.86, 7.43)	7.75 (7.47, 8.03)	1.80E–05	1.89E–04	3.00E–05	2.50E–05	4.00E–05	1.10E–06
SMS	9.73 (9.46, 10.00)	9.37 (9.10, 9.64)	0.041	0.068	0.019	0.011	4.60E–03	0.067
SRM	8.06 (7.94, 8.19)	8.30 (8.17, 8.42)	6.20E–03	0.014	0.011	9.00E–03	0.015	5.20E–03
SUCLG			0.148	0.203	0.165	0.137	0.142	0.198
TALDO1			0.506	0.540	0.637	0.461	0.552	0.473
TKT	9.46 (9.31, 9.61)	9.86 (9.71, 10.01)	4.80E–05	3.78E–04	1.10E–04	1.00E–04	2.50E–04	6.20E–06
TPI1	9.69 (9.52, 9.86)	9.30 (9.14, 9.47)	4.50E–04	0.001	7.60E–04	7.00E–04	1.10E–03	7.00E–04

ACACA: Acetyl-CoA carboxylase alpha, ACO1: Aconitase 1, ACO2: Aconitase 2, ACSL1: Acyl-CoA synthetase 1, ACSL3: Acyl-CoA synthetase 3, ACSL4: Acyl-CoA synthetase 4,ACSL5: Acyl-CoA synthetase 5, ACSL6: Acyl-CoA synthetase 6, AGXT1: Alanine-glyoxylate aminotransferase 1, ALDOA: Aldolase A, ALDOB: Aldolase B, ALDOC: Aldolase C,AMD1: Adenosylmethionine decarboxylase, CHKA: Choline kinase alpha, CS: Citrate synthase, DLD: Dihydrolipoamide dehydrogenase, DLST: Dihydrolipoamides–succinyltransferase, ENO1: Enolase 1, FASN: Fatty acid synthase, FH: Fumarate hydratase, G6PD: Glucose-6-phosphate dehydrogenase, GAPDH: Glyceraldehyde-3-phosphatedehydrogenase, GPI: Glucose-6-phosphate isomerase, HK1: Hexokinase 1, HK2: Hexokinase 2, IDH1: Isocitrate dehydrogenase 1, IDH2: Isocitrate dehydrogenase 2, IDH3A:Isocitrate dehydrogenase 3A, IDH3B: Isocitrate dehydrogenase 3B, IDH3G: Isocitrate dehydrogenase 3G, MDH1: Malate dehydrogenase 1, MDH2: Malate dehydrogenase 2, NPY:Neuropeptide Y, OAZ: Ornithine decarboxylase antizyme, ODC: Ornithine decarboxylase, OGDH: Oxoglutarate (alpha–ketoglutarate) dehydrogenase, OGDHL: Oxoglutarate dehydrogenase-like, PDHA: Pyruvate dehydrogenase alpha, PDHB: Pyruvate dehydrogenase beta, PGK1: Phosphoglycerate kinase, PGD: Phosphogluconate dehydrogenase, PGLS: 6-phosphogluconolactonase, PGM1: Phosphoglucomutase, PKM2: Pyruvate kinase, RBKS: Ribokinase, RPE: Ribulose-5-phosphate-3-epimerase, RPIA: Ribose 5-phosphate isomerase, SAT1: Spermidine/spermine-N1-acetyltransferase 1, SDHA: Succinate dehydrogenase complex, subunit A, SDHB: Succinate dehydrogenase complex, subunit B, SDHD: Succinate dehydrogenase complex, subunit D, SHMT1: Serine hydroxymethyltransferase 1, SMOX: Spermine oxidase, SMS: Spermine synthase, SRM: Spermidine synthase, SUCLG: Succinyl-CoA ligase, TALDO1: Transaldolase 1, TKT: Transketolase, TPI1: Triosephosphate isomerase 1. Since the gene expression values were log2 transformed, a difference in expression by one unit corresponds to a twofold mean change in probe intensities. *Benjamini–Hochberg corrected.

Supplementary Table S6: Differences in expression levels of key metabolic genes comparing ERG_{high} and ERG_{low} in low Gleason (Gleason $\leq 3 + 4$) samples in the main cohort, including estimated means and adjusted for benign epithelia, stroma, cancer and luminal space

Gene	Estimated/predicted means med 95 % CI		*p*-value	*p*-value*	Adjusted for benign epithelial content	Adjusted for stromal content	Adjusted for cancer content	Adjusted for luminal space content
	ERG_{low}	ERG_{high}						
ACACA	9.81 (9.64, 9.98)	10.19 (10.04, 10.33)	2.90E–04	0.001	0.247	0.55	0.355	0.22
ACO1	9.70 (9.58, 9.83)	9.47 (9.36, 9.58)	4.20E–03	0.010	0.016	0.026	8.90E–03	0.041
ACO2	8.12 (8.01, 8.23)	7.91 (7.81, 8.00)	2.80E–03	0.007	0.016	7.60E–04	4.40E–03	0.016
ACSL1	9.30 (8.99, 9.60)	8.86 (8.59, 9.13)	0.017	0.029	0.026	0.072	0.078	0.028
ACSL3			0.294	0.331	1.06E–01	0.062	6.00E–02	8.70E–02
ACSL4	7.18 (7.01, 7.35)	6.76 (6.62, 6.91)	1.80E–04	7.56E–04	0.552	0.647	0.615	0.617
ACSL5			0.248	0.284	6.40E–05	0.029	1.90E–04	2.10E–04
ACSL6			0.826	0.853	0.14	0.136	0.22	0.488
AGXT1			0.951	0.957	3.70E–07	4.40E–06	6.30E–07	4.30E–07
ALDOA	9.39 (9.14, 9.63)	8.74 (8.53, 8.94)	5.80E–05	3.32E–04	8.90E–03	0.022	0.015	0.015
ALDOB	6.06 (5.82, 6.30)	6.45 (6.25, 6.65)	0.014	0.025	2.00E–04	2.80E–04	5.20E–05	2.00E–05
ALDOC	7.43384 (7.29, 7.58)	7.05 (6.92, 7.19)	3.30E–06	2.97E–05	0.526	0.563	0.62	0.497
AMD1	6.93 (6.66, 7.21)	7.35 (7.11, 7.59)	0.013	0.024	3.70E–05	3.70E–05	2.40E–05	4.70E–05
CHKA			0.129	0.159	4.30E–03	6.20E–05	6.30E–05	1.10E–04
CS			0.638	0.681	4.10E–03	9.70E–03	5.50E–03	5.50E–03
DLD	7.89 (7.73, 8.04)	7.55 (7.41, 7.70)	2.00E–06	2.10E–05	1.40E–04	4.70E–04	1.70E–04	8.80E–05
DLST			0.321	0.355	1.10E–04	2.40E–13	3.70E–04	2.50E–11
ENO1	11.29 (11.15, 11.43)	11.63 (11.51, 11.74)	2.60E–04	0.001	1.50E–06	1.50E–11	8.70E–03	1.50E–06
FASN			0.066	0.099	1.30E–04	2.10E–03	2.90E–04	2.50E–04
FH			0.957	0.957	2.10E–03	3.20E–03	3.50E–03	3.50E–03
G6PD			0.111	0.143	0.767	0.319	0.564	0.942
GAPDH	11.36 (11.22, 11.51)	11.59 (11.47, 11.72)	7.00E–03	0.015	7.90E–04	6.30E–07	3.40E–03	2.60E–08
GPI			0.406	0.441	0.15	0.02	0.049	0.06
HK1	5.99 (5.86, 6.12)	5.78 (5.67, 5.90)	0.015	0.026	0.03	4.00E–03	4.00E–03	4.40E–04
HK2			0.077	0.108	2.30E–03	5.40E–03	6.40E–03	1.30E–03
IDH1	9.06 (8.76, 9.37)	8.64 (8.37, 8.91)	0.02	0.033	0.011	0.011	0.012	6.90E–03
IDH2	9.03 (8.87, 9.18)	9.23 (9.09, 9.36)	0.038	0.058	0.081	0.144	0.104	0.108
IDH3A			0.184	0.223	3.50E–03	6.10E–06	1.50E–03	3.40E–05
IDH3B	8.91 (8.82, 9.01)	8.75 (8.67, 8.84)	9.80E–03	0.019	0.017	0.011	0.091	2.10E–03
IDH3G	8.43 (8.31, 8.54)	8.74 (8.63, 8.84)	2.60E–08	4.10E–07	0.032	7.20E–03	0.019	0.017
LDHA			0.072	0.105	3.60E–04	0.073	0.027	0.045
LDHB	9.44 (9.05, 9.83)	8.83 (8.49, 9.17)	0.012	0.023	0.039	0.14	0.227	0.028
MDH1	10.34 (10.18, 10.50)	10.15 (10.01, 10.29)	0.038	0.058	0.227	0.023	0.077	0.068
MDH2	11.89 (11.80, 11.99)	12.14 (12.06, 12.22)	4.10E–05	2.58E–04	3.60E–05	2.30E–04	4.50E–05	6.00E–05
NPY	10.41 (9.34, 11.48)	13.24 (12.33, 14.15)	6.40E–05	3.36E–04	4.70E–04	8.20E–07	3.00E–03	3.70E–06
OAZ	6.57 (6.33, 6.81)	6.38 (6.15, 6.62)	3.10E–03	0.008	0.123	0.029	0.026	0.08
ODC	10.68 (10.28, 11.09)	11.22 (10.86, 11.58)	0.028	0.045	9.10E–23	5.50E–23	2.50E–21	1.80E–24
OGDH	8.35 (8.24, 8.47)	8.54 (8.44, 8.64)	8.60E–03	0.017	0.062	0.055	0.044	0.07
OGDHL	5.95 (5.70, 6.21)	7.62 (7.41, 7.84)	3.40E–24	2.14E–22	4.80E–04	9.40E–04	2.40E–03	6.30E–05
PDHA			0.096	0.129	4.60E–07	2.50E–12	4.60E–07	4.70E–13
PDHB			0.107	0.140	2.20E–04	1.30E–03	4.40E–04	4.40E–04
PGD	8.38 (8.23, 8.52)	8.70 (8.59, 8.82)	3.30E–04	0.001	3.60E–04	3.40E–04	2.40E–04	3.00E–04

PGK1	10.36 (10.15, 10.57)	9.97 (9.79, 10.16)	4.40E–03	0.010	0.133	0.164	0.107	0.186
PGLS	8.69 (8.54, 8.84)	8.97 (8.84, 9.10)	2.10E–03	0.006	0.017	0.019	0.045	0.015
PGLS	8.69 (8.54, 8.84)	8.97 (8.84, 9.10)	2.10E–03	0.006	0.37	0.974	0.732	0.407
PGM1	9.64 (9.41, 9.88)	9.15 (8.95, 9.35)	1.80E–03	0.005	0.123	0.211	0.12	0.1
PKM2	9.06 (8.95, 9.18)	9.43 (9.32, 9.54)	2.90E–13	9.14E–12	4.10E–03	0.01	7.30E–03	3.80E–03
RBKS	6.73 (6.60, 6.86)	7.05 (6.94, 7.16)	3.80E–04	0.001	0.062	0.087	0.079	0.063
RPE			0.755	0.793	5.90E–04	1.90E–04	3.10E–03	3.90E–05
RPIA			0.241	0.281	9.50E–04	1.90E–03	3.60E–03	2.00E–03
SAT1	12.05 (11.82, 12.27)	12.64 (12.45, 12.83)	9.10E–05	4.10E–04	0.123	0.22	0.16	0.116
SDHA	8.70 (8.63, 8.78)	8.53 (8.46, 8.60)	3.30E–05	2.31E–04	0.966	0.901	0.92	0.955
SDHB			0.206	0.245	9.80E–03	0.148	0.25	0.012
SDHC	7.75 (7.53, 7.96)	7.30 (7.11, 7.49)	9.00E–04	0.003	0.951	0.703	0.716	0.957
SDHD	7.72 (7.52, 7.92)	7.19 (7.01, 7.37)	8.90E–06	7.01E–05	0.814	0.931	0.873	0.915
SHMT1			0.077	0.108	2.30E–03	5.40E–03	6.40E–03	1.30E–03
SMOX	7.21 (6.92, 7.51)	8.08 (7.80, 8.36)	1.40E–11	2.94E–10	3.10E–03	6.50E–04	1.80E–03	2.90E–03
SMS			0.094	0.129	1.10E–03	5.00E–03	1.70E–03	2.10E–03
SRM	8.06 (7.90, 8.21)	8.35 (8.22, 8.47)	5.00E–03	0.011	9.50E–03	0.017	0.011	0.01
SUCLG	9.62 (9.49, 9.75)	9.32 (9.20, 9.44)	7.40E–05	3.59E–04	0.012	0.014	0.011	0.017
TALDO1			0.122	0.154	0.221	0.383	0.4	0.228
TKT	9.30 (9.07, 9.52)	10.00 (9.80, 10.19)	8.10E–07	1.02E–05	0.1	0.36	0.177	0.447
TPI1	9.73 (9.50, 9.95)	9.29 (9.10, 9.48)	3.30E–03	0.008	0.088	0.019	0.092	0.01

ACACA: Acetyl-CoA carboxylase alpha, ACO1: Aconitase 1, ACO2: Aconitase 2, ACSL1: Acyl-CoA synthetase 1, ACSL3: Acyl-CoA synthetase 3, ACSL4: Acyl-CoA synthetase 4,ACSL5: Acyl-CoA synthetase 5, ACSL6: Acyl-CoA synthetase 6, AGXT1: Alanine-glyoxylate aminotransferase 1, ALDOA: Aldolase A, ALDOB: Aldolase B, ALDOC: Aldolase C,AMD1: Adenosylmethionine decarboxylase, CHKA: Choline kinase alpha, CS: Citrate synthase, DLD: Dihydrolipoamide dehydrogenase, DLST: Dihydrolipoamides-succinyltransferase, ENO1: Enolase 1, FASN: Fatty acid synthase, FH: Fumarate hydratase, G6PD: Glucose-6-phosphate dehydrogenase, GAPDH: Glyceraldehyde-3-phosphatedehydrogenase, GPI: Glucose-6-phosphate isomerase, HK1: Hexokinase 1, HK2: Hexokinase 2, IDH1: Isocitrate dehydrogenase 1, IDH2: Isocitrate dehydrogenase 2, IDH3A:Isocitrate dehydrogenase 3A, IDH3B: Isocitrate dehydrogenase 3B, IDH3G: Isocitrate dehydrogenase 3G, MDH1: Malate dehydrogenase 1, MDH2: Malate dehydrogenase 2, NPY:Neuropeptide Y, OAZ: Ornithine decarboxylase antizyme, ODC: Ornithine decarboxylase, OGDH: Oxoglutarate (alpha-ketoglutarate) dehydrogenase, OGDHL: Oxoglutarate dehydrogenase-like, PDHA: Pyruvate dehydrogenase alpha, PDHB: Pyruvate dehydrogenase beta, PGK1: Phosphoglycerate kinase, PGD: Phosphogluconate dehydrogenase, PGLS: 6-phosphogluconolactonase, PGM1: Phosphoglucomutase, PKM2: Pyruvate kinase, RBKS: Ribokinase, RPE: Ribulose-5-phosphate-3-epimerase, RPIA: Ribose 5-phosphate isomerase, SAT1: Spermidine/spermine-N1-acetyltransferase 1, SDHA: Succinate dehydrogenase complex, subunit A, SDHB: Succinate dehydrogenase complex, subunit B, SDHD: Succinate dehydrogenase complex, subunit D, SHMT1: Serine hydroxymethyltransferase 1, SMOX: Spermine oxidase, SMS: Spermine synthase, SRM: Spermidine synthase, SUCLG: Succinyl-CoA ligase, TALDO1: Transaldolase 1, TKT: Transketolase, TPI1: Triosephosphate isomerase 1. Since the gene expression values were log2 transformed, a difference inexpression by one unit corresponds to a twofold mean change in probe intensities. *Benjamini-Hochberg corrected.

Supplementary Table S7: Differences in expression levels of key metabolic genes comparing ERG$_{high}$ and ERGl$_{ow}$ in high Gleason (Gleason $\geq 4 + 3$) samples in the main cohort, including estimated means and adjusted for benign epithelia, stroma, cancer and luminal space

Gene	Estimated/predicted means med 95 % CI		*p*-value	*p*-value*	Adjusted for benign epithelial content	Adjusted for stromal content	Adjusted for cancer content	Adjusted for luminal space content
	ERG$_{low}$	ERG$_{high}$						
ACACA			0.054	0.144	0.741	0.55	0.794	0.431
ACO1	9.68 (9.52, 9.84)	9.37 (9.21, 9.52)	1.10E–03	0.023	0.402	0.699	0.395	0.708
ACO2			0.253	0.349	0.247	0.01	0.047	5.00E–03
ACSL1			0.444	0.519	0.665	0.418	0.488	0.386
ACSL3	10.55 (10.32, 10.78)	10.89 (10.66, 11.11)	0.017	0.082	0.048	0.094	0.107	0.036
ACSL4			0.201	0.288	0.247	0.01	0.047	5.00E–03
ACSL5			0.055	0.144	0.277	0.395	0.365	0.363
ACSL6			0.099	0.164	0.94	0.837	0.769	0.964
AGXT1			0.062	0.147	0.143	0.06	0.16	0.998
ALDOA	9.27(9.03, 9.51)	8.88 (8.63, 9.12	0.022	0.089	0.273	0.25	0.238	0.175
ALDOB			0.079	0.158	0.133	0.633	0.204	0.829
ALDOC			0.382	0.491	0.167	0.191	0.17	0.158
AMD1			0.085	0.158	0.932	0.738	0.632	0.95
CHKA	8.96 (8.83, 9.09)	9.44 (9.30, 9.57)	6.60E–07	2.08E–05	0.015	1.20E–03	2.90E–03	2.10E–03
CS			0.255	0.349	0.444	0.063	0.432	0.055
DLD			0.088	0.158	7.00E–03	1.20E–03	6.60E–04	0.025
DLST	6.89 (6.70, 7.07)	7.28 (7.09, 7.46)	1.50E–03	0.024	0.539	0.433	0.57	0.44
ENO1	11.47 (11.35, 11.59)	11.66 (11.53, 11.78)	0.024	0.089	5.80E–04	7.20E–04	1.70E–04	1.30E–04
FASN			0.115	0.186	1.30E–10	1.20E–10	4.60E–11	1.10E–10
FH			0.596	0.636	0.891	0.174	0.383	0.123
G6PD			0.194	0.284	0.468	0.123	0.309	0.228
GAPDH			0.531	0.589	0.035	0.018	0.013	0.01
GPI			0.533	0.589	9.40E–03	6.80E–03	0.011	1.60E–03
HK1			0.394	0.496	0.18	0.081	0.143	7.60E–03
HK2			0.453	0.519	0.774	0.896	0.771	0.931
IDH1			0.43	0.519	0.062	0.188	0.155	0.239
IDH2			0.063	0.147	0.167	0.448	0.27	0.483
IDH3A			0.093	0.160	8.40E–03	3.30E–03	2.90E–03	3.30E–03
IDH3B	8.92 (8.81, 9.03)	8.70 (8.59, 8.81)	4.30E–03	0.039	1.10E–03	0.057	0.011	0.042
IDH3G			0.592	0.636	0.012	0.064	0.014	0.102
LDHA			0.307	0.403	0.733	0.03	0.118	0.041
LDHB	9.39 (9.02, 9.75)	8.77 (8.40, 9.13)	0.011	0.063	0.034	0.092	0.205	6.20E–03
MDH1			0.903	0.918	0.308	0.083	0.179	0.029
MDH2			0.077	0.158	0.023	0.081	0.04	0.13
NPY	11.48 (10.49, 12.48)	13.35 (12.36, 14.34)	3.80E–03	0.039	0.888	0.156	0.544	0.285
OAZ	5.30 (5.12, 5.48)	4.92 (4.75, 5.10)	3.20E–03	0.039	0.012	0.018	0.014	0.016
ODC			0.084	0.158	0.106	0.011	0.03	1.10E–03

OGDH			0.805	0.831	0.964	0.798	0.755	0.626
OGDHL	6.33 (6.06, 6.60)	7.32 (7.06, 7.59)	1.70E–10	1.07E–08	0.747	0.053	0.135	0.038
PDHA	5.84 (5.74, 5.95)	5.66 (5.55, 5.77)	0.019	0.086	3.20E–03	6.80E–05	1.90E–04	1.60E–04
PDHB			0.951	0.951	0.09	0.033	0.062	0.024
PGD			0.142	0.224	0.069	0.02	0.016	0.012
PGK1	8.21 (8.05, 8.37)	8.01 (7.85, 8.17)	0.032	0.099	0.489	0.595	0.444	0.91
PGLS	8.68 (8.54, 8.81)	8.90 (8.77, 9.03)	1.00E–02	0.063	0.14	0.056	0.079	0.119
PGLS	8.68 (8.54, 8.81)	8.90 (8.77, 9.03)	1.00E–02	0.063	0.195	0.013	0.079	0.034
PGM1			0.094	0.160	0.085	0.097	0.162	0.039
PKM2			0.445	0.519	0.088	0.026	0.042	0.042
RBKS	6.89 (6.78, 7.00)	7.06 (6.94, 7.17)	0.037	0.106	0.717	0.625	0.752	0.347
RPE			0.194	0.284	0.013	0.037	0.023	4.20E–03
RPIA			0.08	0.158	0.678	0.518	0.692	0.314
SAT1	12.38 (12.18, 12.58)	12.66 (12.46, 12.85)	0.033	0.099	0.761	0.133	0.418	0.252
SDHA			0.061	0.147	0.496	0.48	0.921	0.452
SDHB			0.414	0.511	0.036	0.026	0.018	0.082
SDHC	6.29 (6.15, 6.44)	6.06 (5.91, 6.21)	0.028	0.098	0.21	0.025	0.055	0.021
SDHD			0.295	0.395	0.756	0.426	0.468	0.372
SHMT1	6.98 (6.76, 7.21)	7.32 (7.10, 7.55)	0.023	0.089	0.142	0.083	0.106	0.164
SMOX	6.87 (6.47, 7.27)	7.45 (7.06, 7.84)	9.50E–03	0.063	0.167	0.095	0.104	0.079
SMS			0.078	0.158	3.00E–03	8.40E–03	9.00E–03	2.40E–03
SRM			0.166	0.255	0.918	0.946	0.996	0.718
SUCLG	8.90 (8.77, 9.03)	9.11 (8.98, 9.24)	0.016	0.082	0.785	0.324	0.309	0.519
TALDO1			0.801	0.831	6.80E–03	0.017	8.80E–03	5.90E–03
TKT			0.086	0.158	0.149	0.083	0.117	0.059
TPI1	9.68 (9.43, 9.92)	9.33 (9.08, 9.57)	0.03	0.099	0.094	0.114	0.118	0.064

ACACA: Acetyl-CoA carboxylase alpha, ACO1: Aconitase 1, ACO2: Aconitase 2, ACSL1: Acyl-CoA synthetase 1, ACSL3: Acyl-CoA synthetase 3, ACSL4: Acyl-CoA synthetase 4,ACSL5: Acyl-CoA synthetase 5, ACSL6: Acyl-CoA synthetase 6, AGXT1: Alanine-glyoxylate aminotransferase 1, ALDOA: Aldolase A, ALDOB: Aldolase B, ALDOC: Aldolase C,AMD1: Adenosylmethionine decarboxylase, CHKA: Choline kinase alpha, CS: Citrate synthase, DLD: Dihydrolipoamide dehydrogenase, DLST: Dihydrolipoamides-succinyltransferase, ENO1: Enolase 1, FASN: Fatty acid synthase, FH: Fumarate hydratase, G6PD: Glucose-6-phosphate dehydrogenase, GAPDH: Glyceraldehyde-3-phosphatedehydrogenase, GPI: Glucose-6-phosphate isomerase, HK1: Hexokinase 1, HK2: Hexokinase 2, IDH1: Isocitrate dehydrogenase 1, IDH2: Isocitrate dehydrogenase 2, IDH3A:Isocitrate dehydrogenase 3A, IDH3B: Isocitrate dehydrogenase 3B, IDH3G: Isocitrate dehydrogenase 3G, MDH1: Malate dehydrogenase 1, MDH2: Malate dehydrogenase 2, NPY:Neuropeptide Y, OAZ: Ornithine decarboxylase antizyme, ODC: Ornithine decarboxylase, OGDH: Oxoglutarate (alpha-ketoglutarate) dehydrogenase, OGDHL: Oxoglutarate dehydrogenase-like, PDHA: Pyruvate dehydrogenase alpha, PDHB: Pyruvate dehydrogenase beta, PGK1: Phosphoglycerate kinase, PGD: Phosphogluconate dehydrogenase, PGLS: 6-phosphogluconolactonase, PGM1: Phosphoglucomutase, PKM2: Pyruvate kinase, RBKS: Ribokinase, RPE: Ribulose-5-phosphate-3-epimerase, RPIA: Ribose 5-phosphate isomerase, SAT1: Spermidine/spermine-N1-acetyltransferase 1, SDHA: Succinate dehydrogenase complex, subunit A, SDHB: Succinate dehydrogenase complex, subunit B, SDHD: Succinate dehydrogenase complex, subunit D, SHMT1: Serine hydroxymethyltransferase 1, SMOX: Spermine oxidase, SMS: Spermine synthase, SRM: Spermidine synthase, SUCLG: Succinyl-CoA ligase, TALDO1: Transaldolase 1, TKT: Transketolase, TPI1: Triosephosphate isomerase 1. Since the gene expression values were log2 transformed, a difference in expression by one unit corresponds to a twofold mean change in probe intensities. *Benjamini-Hochberg corrected.

Supplementary Table S8: Differences in levels of quantified metabolites in the main cohort comparing ERG$_{high}$ samples to ERG$_{low}$ samples in low Gleason samples (Gleason ≤ 3 + 4), corrected for multiple testing with Benjamini-Hochberg correction and adjusted for stroma, cancer, benign epithelia and luminal space

Metabolites	ERG$_{low}$ (n = 12) Concentrations mmoles/kg wet weight, median (IQR)	ERG$_{high}$ (n = 17) Concentrations mmoles/kg wet weight, median (IQR)	p-value	p-value*	Adjusted for stromal content	Adjusted for cancer content	Adjusted for benign epithelial content	Adjusted for luminal space content
Alanine	1.99 (1.64 to 2.43)	2.60 (1.78 to 2.85)	0.651	0.83	0.766	0.702	0.622	0.667
Choline	0.78 (0.49 to 1.43)	1.23 (0.84 to 1.48)	0.087	0.16	0.364	0.143	0.075	0.082
Citrate	13.95 (8.62 to 17.06)	7.07 (3.43 to 7.82)	0.001	0.004	7.70E–05	1.10E–03	9.90E–04	8.00E–04
Creatine	2.23 (1.63 to 2.58)	2.23 (1.68 to 2.67)	0.679	0.83	0.708	0.634	0.663	0.742
Ethm	0.00 (0.00 to 0.00)	0.02 (0.00 to 0.41)	0.06	0.132	0.176	0.117	0.058	0.052
Glucose	0.23 (0.00 to 0.43)	0.00 (0.00 to 0.07)	0.12	0.203	0.222	0.203	0.126	0.095
Glutamate	4.58 (3.77 to 5.48)	4.55 (4.20 to 7.79)	0.068	0.136	0.161	0.259	0.087	0.065
Glutamine	2.35 (2.05 to 2.68)	3.18 (2.27 to 3.66)	0.005	0.018	0.012	8.40E–03	5.20E–03	5.80E–03
Glycine	1.71 (1.43 to 2.22)	2.68 (1.93 to 3.51)	< 0.001	0.002	8.00E–04	4.60E–04	1.70E–04	2.90E–04
GPC	0.67 (0.48 to 1.18)	0.98 (0.68 to 1.24)	0.908	0.951	0.762	0.855	0.755	0.783
GPE	0.33 (0 to 0.53)	0.00 (0.00 to 0.73)	< 0.001	< 0.001	1.30E–40	4.10E–10	1.30E–03	6.00E–32
Isoleucine	0.16 (0.00 to 0.18)	0.16 (0.11 to 0.28)	0.027	0.085	5.50E–04	0.015	0.029	0.016
Lactate	19.77 (16.68 to 22.66)	17.81 (15.69 to 22.01)	0.997	0.997	0.845	0.833	0.984	0.984
Leucine	0.33 (0.26 to 0.45)	0.44 (0.28 to 0.63)	0.364	0.501	0.447	0.329	0.353	0.37
Myo–inositol	9.94 (7.71 to 10.25)	8.79 (7.23 to 10.09)	0.795	0.875	0.702	0.744	0.796	0.78
PCh	0.43 (0.26 to 0.81)	0.90 (0.54 to 1.17)	0.039	0.107	0.014	0.033	0.032	0.043
PE	2.18 (1.49 to 2.54)	2.63 (2.26 to 3.71)	0.055	0.132	0.016	0.053	0.039	0.034
Putrescine	0.20 (0.00 to 0.83)	0.05 (0.00 to 0.16)	0.001	0.004	4.60E–03	0.075	0.029	1.00E–03
Scyllo–inositol	0.40 (0.27 to 0.87)	0.41 (0.32 to 0.61)	0.742	0.859	0.286	0.257	0.783	0.738
Spermine	3.56 (1.64 to 6.07)	1.18 (0.69 to 1.54)	< 0.001	< 0.001	9.10E–08	2.00E–05	4.50E–05	5.60E–06
Succinate	0.51 (0.39 to 0.73)	0.62 (0.52 to 0.94)	0.151	0.237	0.325	0.191	0.142	0.133
Taurine	4.83 (3.20 to 6.05)	4.23 (3.03 to 4.73)	0.344	0.501	0.488	0.46	0.333	0.519
Valine	1.99 (1.64 to 2.43)	2.60 (1.78 to 2.85)	0.651	0.83	0.026	0.021	0.013	8.40E–03

Ethm: Ethanolamine, GPC: Glycerophosphocholine, GPE: Glycerophosphoethanolamine, PCh: Phosphocholine, PE: Phosphoethanolamine *Benjamini-Hochberg corrected.

Supplementary Table S9: Differences in levels of quantified metabolites in the main cohort comparing ERG$_{high}$ samples to ERG$_{low}$ samples in high Gleason samples (Gleason ≥ 4 + 3), corrected for multiple testing with Benjamini-Hochberg correction and adjusted for stroma, cancer, benign epithelia and luminal space

Metabolites	ERG$_{low}$ (n = 18) Concentrations mmoles/kg wet weight, median (IQR)	ERG$_{high}$ (n = 17) Concentrations mmoles/kg wet weight, median (IQR)	p–value	p value*	Adjusted for stromal content	Adjusted for cancer content	Adjusted for benign epithelial content	Adjusted for luminal space content
Alanine	2.34 (1.83 to 3.34)	2.36 (1.97 to 2.78)	0.762	0.922	0.723	0.51	0.453	0.785
Choline	1.01 (0.66 to 1.43)	1.23 (0.90 to 1.66)	0.372	0.611	0.381	0.508	0.454	0.322
Citrate	7.58 (5.51 to 10.69)	2.95 (2.34 to 4.04)	< 0.001	0.003	1.60E–04	1.00E–04	1.30E–04	3.20E–04
Creatine	2.27 (1.71 to 2.75)	1.97 (1.60 to 2.22)	0.825	0.928	0.841	0.761	0.865	0.848
Ethm	0.00 (0.00 to 0.08)	0.00 (0.00 to 0.19)	0.6	0.812	0.619	0.79	0.969	0.676
Glucose	0.15 (0.00 to 1.00)	0.00 (0.00 to 0.04)	0.016	0.087	6.10E–03	0.068	0.313	0.017
Glutamate	4.67 (4.11 to 7.28)	6.31 (5.75 to 7.74)	0.241	0.504	0.294	0.54	0.517	0.215
Glutamine	2.80 (2.45 to 3.89)	2.93 (2.69 to 3.70)	0.907	0.948	0.812	0.461	0.361	0.946
Glycine	2.57 (1.85 to 3.08)	3.17 (2.11 to 3.65)	0.366	0.611	0.389	0.58	0.833	0.417
GPC	0.81 (0.54 to 1.43)	0.84 (0.44 to 1.08)	0.714	0.912	0.798	0.886	0.908	0.38
GPE	0.00 (0.00 to 0.55)	0.00 (0.00 to 0.36)	0.541	0.778	0.538	0.518	0.507	0.413
Isoleucine	0.18 (0.07 to 0.27)	0.24 (0.12 to 0.32)	0.188	0.48	0.183	0.133	5.50E–03	0.208
Lactate	22.02 (15.64 to 27.75)	16.45 (15.70 to 20.85)	0.065	0.249	0.064	0.046	0.056	0.04
Leucine	0.52 (0.34 to 0.65)	0.59 (0.45 to 0.83)	0.974	0.974	0.978	0.86	0.828	0.966
Myo–inositol	10.82 (6.72 to 12.76)	9.39 (7.43 to 11.35)	0.226	0.504	0.236	0.226	0.401	0.377
PCh	0.72 (0.45 to 1.03)	1.12 (0.74 to 1.36)	0.019	0.087	0.021	0.079	0.232	0.011
PE	2.62 (1.35 to 3.69)	2.95 (2.78 to 3.80)	0.087	0.286	0.088	0.087	0.081	0.03
Putrescine	0.11 (0.00 to 0.56)	0.00 (0.00 to 0.00)	0.008	0.058	7.70E–03	6.80E–03	0.096	0.01
Scyllo–inositol	0.38 (0.34 to 0.59)	0.50 (0.41 to 0.60)	0.366	0.611	0.487	0.295	0.191	0.604
Spermine	1.65 (1.15 to 2.27)	0.58 (0.39 to 1.02)	< 0.001	0.004	3.10E–04	4.30E–04	3.10E–03	4.90E–04
Succinate	0.60 (0.54 to 0.74)	0.63 (0.46 to 0.76)	0.847	0.928	0.824	0.516	0.637	0.963
Taurine	6.21 (4.05 to 7.15)	4.30 (3.70 to 6.49)	0.172	0.48	0.176	0.314	0.581	0.222
Valine	0.46 (0.28 to 0.58)	0.40 (0.28 to 0.52)	0.441	0.676	0.38	0.272	0.463	0.363

Ethm: Ethanolamine, GPC: Glycerophosphocholine, GPE: Glycerophosphoethanolamine, PCh: Phosphocholine, PE: Phosphoethanolamine *Benjamini-Hochberg corrected.

Supplementary Table S10: INMEX-analysis of KEGG-pathways in the main cohort

Pathway	Total	Expected	Hits	*P*-Value	Topology
Glutathione metabolism	75	6.5205	17	0.000155	0.79245
Purine metabolism	234	20.344	33	0.00267	0.66667
Glycerolipid metabolism	72	6.2597	14	0.002904	0.78378
Arginine and proline metabolism	102	8.8679	17	0.005846	0.63441
Cyanoamino acid metabolism	12	1.0433	4	0.015752	1.25
Pyruvate metabolism	64	5.5642	11	0.019979	0.78261
Glycolysis / Gluconeogenesis	91	7.9116	14	0.023299	0.57627
Glycosaminoglycan degradation	20	1.7388	5	0.024915	0.42105
Glycine, serine and threonine metabolism	68	5.9119	11	0.030255	0.66667
Pentose and glucuronate interconversions	52	4.5209	9	0.032529	0.89655
Lysine degradation	73	6.3466	11	0.04783	0.46512
beta–Alanine metabolism	50	4.347	8	0.063613	0.38095
Propanoate metabolism	52	4.5209	8	0.076871	0.63415
Glyoxylate and dicarboxylate metabolism	53	4.6078	8	0.084079	0.52
Valine, leucine and isoleucine degradation	82	7.1291	11	0.094732	0.43333
Pentose phosphate pathway	48	4.1731	7	0.11797	0.97561
Vitamin B6 metabolism	15	1.3041	3	0.13563	0.4
Ether lipid metabolism	51	4.4339	7	0.14941	0.34483
Fatty acid metabolism	83	7.216	10	0.17989	1.2621
Ascorbate and aldarate metabolism	35	3.0429	5	0.18309	0.83333
Alanine, aspartate and glutamate metabolism	56	4.8686	7	0.20941	0.40816
Aminoacyl–tRNA biosynthesis	87	7.5638	10	0.2201	0.27536
One carbon pool by folate	28	2.4343	4	0.22201	0.26667
Sphingolipid metabolism	67	5.825	8	0.22271	0.35849
Nicotinate and nicotinamide metabolism	39	3.3907	5	0.24769	0.43243
Pyrimidine metabolism	142	12.346	15	0.24795	0.49558
Arachidonic acid metabolism	100	8.694	11	0.24825	0.41026
Riboflavin metabolism	20	1.7388	3	0.24925	0.46154
Mucin type O–Glycan biosynthesis	32	2.7821	4	0.30123	0.15
Histidine metabolism	44	3.8254	5	0.33492	0.1875
Glycerophospholipid metabolism	119	10.346	12	0.33741	0.37333
Glycosphingolipid biosynthesis – ganglio series	14	1.2172	2	0.34713	0.26087
Glycosaminoglycan biosynthesis – chondroitin sulfate	14	1.2172	2	0.34713	0.42857
Pantothenate and CoA biosynthesis	35	3.0429	4	0.3626	0.58065
Glycosaminoglycan biosynthesis – heparan sulfate	5	0.4347	1	0.36563	0.33333
Butanoate metabolism	47	4.0862	5	0.38864	0.5625
Tryptophan metabolism	80	6.9552	8	0.39387	0.325
Glycosphingolipid biosynthesis – lacto and neolacto series	26	2.2604	3	0.39675	0.067797
Porphyrin and chlorophyll metabolism	70	6.0858	7	0.408	0.29787
Metabolism of xenobiotics by cytochrome P450	139	12.085	13	0.43406	0.4
Retinol metabolism	83	7.216	8	0.43514	0.51163
N–Glycan biosynthesis	50	4.347	5	0.44211	0.17143
Citrate cycle (TCA cycle)	50	4.347	5	0.44211	0.75
Butirosin and neomycin biosynthesis	7	0.60858	1	0.47135	0.66667
Taurine and hypotaurine metabolism	18	1.5649	2	0.47262	0.4

Sulfur metabolism	19	1.6519	2	0.50172	0.5
Drug metabolism – other enzymes	77	6.6944	7	0.51018	0.31579
Starch and sucrose metabolism	78	6.7813	7	0.52435	0.35185
Galactose metabolism	55	4.7817	5	0.52831	0.36735
Fructose and mannose metabolism	55	4.7817	5	0.52831	0.43243
Inositol phosphate metabolism	90	7.8246	8	0.52956	0.375
Drug metabolism – cytochrome P450	127	11.041	11	0.55364	0.14815
Fatty acid elongation	57	4.9556	5	0.56115	0.35616
Amino sugar and nucleotide sugar metabolism	84	7.303	7	0.60576	0.2716
Thiamine metabolism	11	0.95634	1	0.63303	0.36364
Valine, leucine and isoleucine biosynthesis	13	1.1302	1	0.69432	0.72727
Synthesis and degradation of ketone bodies	14	1.2172	1	0.72102	0.44444
Phenylalanine metabolism	29	2.5213	2	0.7326	0.36364
Caffeine metabolism	17	1.478	1	0.78799	0.11765
Selenocompound metabolism	33	2.869	2	0.79592	0.21429
Primary bile acid biosynthesis	63	5.4772	4	0.8118	0.34043
Glycosylphosphatidylinositol(GPI)–anchor biosynthesis	20	1.7388	1	0.83893	0.096774
Tyrosine metabolism	80	6.9552	5	0.83931	0.27586
Biosynthesis of unsaturated fatty acids	27	2.3474	1	0.91528	0.6875
Folate biosynthesis	32	2.7821	1	0.94652	0.11429
alpha–Linolenic acid metabolism	34	2.956	1	0.95552	0.1
Linoleic acid metabolism	34	2.956	1	0.95552	0.25
Terpenoid backbone biosynthesis	39	3.3907	1	0.97196	0.057143
Cysteine and methionine metabolism	63	5.4772	2	0.97836	0.072727
Fatty acid biosynthesis	49	4.2601	1	0.98888	0.11111
Steroid biosynthesis	54	4.6948	1	0.99301	0.029412
Steroid hormone biosynthesis	137	11.911	4	0.99854	0.17978

Supplementary Table S11: INMEX-analysis of KEGG-pathways in the main cohort in low Gleason (Gleason ≤ 3 + 4) samples

Pathway	Total	Expected	Hits	P.Value	Topology
Purine metabolism	234	15.941	34	8.98E–06	0.57471
Glutathione metabolism	75	5.1091	14	0.000393	0.75472
Pyrimidine metabolism	142	9.6733	18	0.006584	0.58407
Glycolysis / Gluconeogenesis	91	6.1991	13	0.007606	0.57627
Glycosaminoglycan degradation	20	1.3624	5	0.009281	0.42105
Valine, leucine and isoleucine degradation	82	5.586	11	0.021342	0.45556
Lysine degradation	73	4.9729	10	0.024199	0.46512
Arginine and proline metabolism	102	6.9484	12	0.041654	0.31183
Tryptophan metabolism	80	5.4498	10	0.042465	0.375
Cyanoamino acid metabolism	12	0.81746	3	0.043309	1.25
Nicotinate and nicotinamide metabolism	39	2.6568	6	0.045681	0.48649
beta–Alanine metabolism	50	3.4061	7	0.050052	0.33333
Glycerolipid metabolism	72	4.9048	9	0.052883	0.2973
Pentose and glucuronate interconversions	52	3.5423	7	0.05985	0.48276
Vitamin B6 metabolism	15	1.0218	3	0.077143	0.4
Sphingolipid metabolism	67	4.5642	8	0.081965	0.32075
Ascorbate and aldarate metabolism	35	2.3843	5	0.08539	0.83333
Pentose phosphate pathway	48	3.2699	6	0.10428	0.87805
Fatty acid metabolism	83	5.6541	9	0.10817	1.1456
Citrate cycle (TCA cycle)	50	3.4061	6	0.12099	0.75
Ether lipid metabolism	51	3.4742	6	0.12981	0.2069
Propanoate metabolism	52	3.5423	6	0.13894	0.58537
Glyoxylate and dicarboxylate metabolism	53	3.6105	6	0.14836	0.48
Drug metabolism – cytochrome P450	127	8.6515	12	0.15174	0.16667
Glycine, serine and threonine metabolism	68	4.6323	7	0.17714	0.53968
Amino sugar and nucleotide sugar metabolism	84	5.7222	8	0.20982	0.39506
Pantothenate and CoA biosynthesis	35	2.3843	4	0.21232	0.58065
Aminoacyl–tRNA biosynthesis	87	5.9266	8	0.23802	0.21739
Pyruvate metabolism	64	4.3598	6	0.26789	0.34783
Retinol metabolism	83	5.6541	7	0.33519	0.37209
Metabolism of xenobiotics by cytochrome P450	139	9.4689	11	0.34633	0.36364
Fatty acid elongation	57	3.883	5	0.34699	0.35616
Histidine metabolism	44	2.9974	4	0.35192	0.125
Riboflavin metabolism	20	1.3624	2	0.39995	0.30769
Starch and sucrose metabolism	78	5.3135	6	0.44142	0.33333
Arachidonic acid metabolism	100	6.8122	7	0.5286	0.4359
Glycosphingolipid biosynthesis – lacto and neolacto series	26	1.7712	2	0.53782	0.050847
Thiamine metabolism	11	0.74934	1	0.54049	0.36364
Biosynthesis of unsaturated fatty acids	27	1.8393	2	0.55861	0.6875
One carbon pool by folate	28	1.9074	2	0.57873	0.2
Inositol phosphate metabolism	90	6.131	6	0.5844	0.25
Glycosphingolipid biosynthesis – globo series	13	0.88559	1	0.60122	0.22222
Valine, leucine and isoleucine biosynthesis	13	0.88559	1	0.60122	0.72727
Drug metabolism – other enzymes	77	5.2454	5	0.61177	0.17544

Glycosaminoglycan biosynthesis – chondroitin sulfate	14	0.95371	1	0.62852	0.28571
Synthesis and degradation of ketone bodies	14	0.95371	1	0.62852	0.44444
Tyrosine metabolism	80	5.4498	5	0.64615	0.22989
N–Glycan biosynthesis	50	3.4061	3	0.6735	0.11429
Caffeine metabolism	17	1.1581	1	0.69976	0.11765
Glycerophospholipid metabolism	119	8.1065	7	0.71401	0.17333
Taurine and hypotaurine metabolism	18	1.2262	1	0.72034	0.13333
Galactose metabolism	55	3.7467	3	0.73553	0.22449
Fructose and mannose metabolism	55	3.7467	3	0.73553	0.27027
Glycosylphosphatidylinositol(GPI)–anchor biosynthesis	20	1.3624	1	0.75739	0.096774
Primary bile acid biosynthesis	63	4.2917	3	0.81507	0.25532
Butanoate metabolism	47	3.2017	2	0.84148	0.375
Porphyrin and chlorophyll metabolism	70	4.7685	3	0.86718	0.10638
Phenylalanine metabolism	29	1.9755	1	0.8722	0.18182
Mucin type O–Glycan biosynthesis	32	2.1799	1	0.89684	0.15
Alanine, aspartate and glutamate metabolism	56	3.8148	2	0.90443	0.12245
Linoleic acid metabolism	34	2.3161	1	0.91058	0.5
Cysteine and methionine metabolism	63	4.2917	2	0.93638	0.10909
Steroid hormone biosynthesis	137	9.3327	2	0.99944	0.10112

Supplementary Table S12: INMEX-analysis of KEGG-pathways in the main cohort in high Gleason (Gleason ≥ 4 + 3) samples

Pathway	Total	Expected	Hits	*P*-Value	Topology
Purine metabolism	234	5.4603	13	0.002268	0.41954
Pyrimidine metabolism	142	3.3135	9	0.004933	0.49558
Glutathione metabolism	75	1.7501	6	0.007251	0.49057
Glycosphingolipid biosynthesis – lacto and neolacto series	26	0.6067	3	0.02147	0.067797
Nicotinate and nicotinamide metabolism	39	0.91005	3	0.061006	0.37838
Glycine, serine and threonine metabolism	68	1.5868	4	0.072199	0.2381
Arginine and proline metabolism	102	2.3801	5	0.087048	0.16129
Retinol metabolism	83	1.9368	4	0.1264	0.37209
Arachidonic acid metabolism	100	2.3335	4	0.20307	0.12821
Cyanoamino acid metabolism	12	0.28002	1	0.24718	0.5
Drug metabolism – other enzymes	77	1.7968	3	0.26692	0.10526
Glycosaminoglycan biosynthesis – chondroitin sulfate	14	0.32668	1	0.28207	0.14286
Glycosphingolipid biosynthesis – ganglio series	14	0.32668	1	0.28207	0.17391
Vitamin B6 metabolism	15	0.35002	1	0.29891	0.1
Fatty acid metabolism	83	1.9368	3	0.3056	0.19417
N–Glycan biosynthesis	50	1.1667	2	0.32651	0.057143
Citrate cycle (TCA cycle)	50	1.1667	2	0.32651	0.29545
Caffeine metabolism	17	0.39669	1	0.33143	0.11765
Drug metabolism – cytochrome P450	127	2.9635	4	0.3441	0.50926
Taurine and hypotaurine metabolism	18	0.42002	1	0.34713	0.13333
Inositol phosphate metabolism	90	2.1001	3	0.35105	0.125
Glyoxylate and dicarboxylate metabolism	53	1.2367	2	0.35238	0.12
Fructose and mannose metabolism	55	1.2834	2	0.36944	0.16216
Riboflavin metabolism	20	0.46669	1	0.37746	0.15385
Metabolism of xenobiotics by cytochrome P450	139	3.2435	4	0.40908	0.36364
One carbon pool by folate	28	0.65337	1	0.48549	0.2
Glycerolipid metabolism	72	1.6801	2	0.50592	0.18919
Lysine degradation	73	1.7034	2	0.51337	0.13953
Folate biosynthesis	32	0.74671	1	0.53236	0.11429
Mucin type O–Glycan biosynthesis	32	0.74671	1	0.53236	0.15
Linoleic acid metabolism	34	0.79338	1	0.55419	0.5
Tryptophan metabolism	80	1.8668	2	0.56346	0.05
Tyrosine metabolism	80	1.8668	2	0.56346	0.068966
Pantothenate and CoA biosynthesis	35	0.81671	1	0.56473	0.12903
Valine, leucine and isoleucine degradation	82	1.9134	2	0.57709	0.044444
beta–Alanine metabolism	50	1.1667	1	0.6963	0.047619
Ether lipid metabolism	51	1.1901	1	0.70352	0.068966
Pentose and glucuronate interconversions	52	1.2134	1	0.71058	0.13793
Alanine, aspartate and glutamate metabolism	56	1.3067	1	0.73718	0.081633
Fatty acid elongation	57	1.3301	1	0.74344	0.041096
Primary bile acid biosynthesis	63	1.4701	1	0.77806	0.19149
Sphingolipid metabolism	67	1.5634	1	0.79854	0.037736
Porphyrin and chlorophyll metabolism	70	1.6334	1	0.81267	0.042553
Steroid hormone biosynthesis	137	3.1968	2	0.83921	0.05618

Starch and sucrose metabolism	78	1.8201	1	0.84576	0.037037
Amino sugar and nucleotide sugar metabolism	84	1.9601	1	0.86673	0.049383
Glycolysis / Gluconeogenesis	91	2.1234	1	0.88767	0.067797
Glycerophospholipid metabolism	119	2.7768	1	0.94357	0.053333

Supplementary Table S13: GSEA of KEGG-pathways in the main cohort, enrichment in ERG$_{low}$

PATHWAY	SIZE	ES	NES	NOM p-val	FDR q-val
DRUG METABOLISM CYTOCHROME P450	55	−0.6651254	−1.7795105	0.002132196	0.38855085
GLYCOLYSIS GLUCONEOGENESIS	56	−0.4682116	−1.6126508	0.013100437	0.8536236
HEMATOPOIETIC CELL LINEAGE	68	−0.6377796	−1.5941937	0.01632653	0.66214764
BETA ALANINE METABOLISM	19	−0.6408701	−1.563926	0.020242915	0.62843364
GLUTATHIONE METABOLISM	45	−0.5840763	−1.5582185	0.052738335	0.528509
PRION DISEASES	30	−0.5652002	−1.5239826	0.03285421	0.57103133
HYPERTROPHIC CARDIOMYOPATHY HCM	66	−0.5525483	−1.521536	0.04592902	0.49907997
CARDIAC MUSCLE CONTRACTION	58	−0.5234164	−1.5056521	0.033126295	0.4885571
METABOLISM OF XENOBIOTICS BY CYTOCHROME P450	50	−0.5358754	−1.4941844	0.017094018	0.47057903
VALINE LEUCINE AND ISOLEUCINE DEGRADATION	44	−0.556476	−1.491827	0.0741483	0.43163717
PPAR SIGNALING PATHWAY	51	−0.52109236	−1.4773304	0.027139874	0.43209052
FATTY ACID METABOLISM	35	−0.4665317	−1.4668512	0.07385229	0.42450482
DILATED CARDIOMYOPATHY	73	−0.51889294	−1.4528376	0.07660455	0.4315167
HISTIDINE METABOLISM	24	−0.5824359	−1.4319005	0.043912176	0.4642199
APOPTOSIS	75	−0.35446638	−1.4161711	0.0503876	0.48000923
TYROSINE METABOLISM	31	−0.51137507	−1.3706481	0.07236842	0.5867568
RETINOL METABOLISM	40	−0.4578147	−1.3549248	0.06313646	0.6055808
ARGININE AND PROLINE METABOLISM	46	−0.47907144	−1.3485328	0.09325397	0.5914392
PROXIMAL TUBULE BICARBONATE RECLAMATION	18	−0.5100017	−1.3183355	0.11677282	0.6578999
ALDOSTERONE REGULATED SODIUM REABSORPTION	33	−0.5453702	−1.3126537	0.1122449	0.64245903
GLYCOSPHINGOLIPID BIOSYNTHESIS LACTO AND NEOLACTO SERIES	17	−0.47345862	−1.2855277	0.15303983	0.6999029
PROPANOATE METABOLISM	31	−0.42541942	−1.2627256	0.2413793	0.74142903
MELANOGENESIS	82	−0.38168576	−1.2602468	0.12931034	0.71696585
COLORECTAL CANCER	54	−0.3426833	−1.2553287	0.16260162	0.7035733
NICOTINATE AND NICOTINAMIDE METABOLISM	19	−0.5683716	−1.249494	0.15118791	0.69425493
REGULATION OF ACTIN CYTOSKELETON	176	−0.32957068	−1.2489338	0.18371607	0.6691262
MTOR SIGNALING PATHWAY	43	−0.32117867	−1.2314632	0.15748031	0.7004713
FOCAL ADHESION	168	−0.38118944	−1.2233069	0.20910972	0.70139503
STEROID HORMONE BIOSYNTHESIS	36	−0.44694844	−1.2228109	0.16115703	0.6786272
B CELL RECEPTOR SIGNALING PATHWAY	70	−0.39102882	−1.2056568	0.24158415	0.7087562
GLIOMA	57	−0.33602458	−1.2041162	0.16297787	0.6909752
PANCREATIC CANCER	63	−0.33796537	−1.2031119	0.21850394	0.67281705
PHENYLALANINE METABOLISM	17	−0.50094825	−1.1898077	0.21638656	0.6888747
CELL ADHESION MOLECULES CAMS	112	−0.4456652	−1.1702002	0.26814517	0.7239068

VASCULAR SMOOTH MUSCLE CONTRACTION	89	−0.3525924	−1.1684161	0.2516269	0.70892125
PHOSPHATIDYLINOSITOL SIGNALING SYSTEM	65	−0.32815862	−1.1536041	0.26326963	0.73398
NATURAL KILLER CELL MEDIATED CYTOTOXICITY	101	−0.3900806	−1.1503791	0.3156823	0.72411835
MAPK SIGNALING PATHWAY	218	−0.2851731	−1.1377566	0.22937626	0.74028134
ADHERENS JUNCTION	62	−0.31532508	−1.1332755	0.2838983	0.7337195
WNT SIGNALING PATHWAY	129	−0.31501043	−1.1323106	0.24347825	0.7179757
ANTIGEN PROCESSING AND PRESENTATION	65	−0.43509716	−1.1301155	0.33467743	0.7065309
TRYPTOPHAN METABOLISM	33	−0.43367508	−1.1286054	0.27272728	0.6938948
PYRUVATE METABOLISM	38	−0.31179887	−1.1273845	0.3224401	0.68096334
CHEMOKINE SIGNALING PATHWAY	152	−0.3680516	−1.1199098	0.31451613	0.683802
ARRHYTHMOGENIC RIGHT VENTRICULAR CARDIOMYOPATHY ARVC	59	−0.38263643	−1.1110188	0.32415253	0.6903247
PROSTATE CANCER	83	−0.2898831	−1.1048385	0.30181086	0.69068277
P53 SIGNALING PATHWAY	61	−0.36322063	−1.1022831	0.31013918	0.6820439
VIRAL MYOCARDITIS	62	−0.4290487	−1.102143	0.35918367	0.66820705
CALCIUM SIGNALING PATHWAY	124	−0.32479998	−1.0972714	0.29817444	0.666126
LEISHMANIA INFECTION	58	−0.41226605	−1.094746	0.3478261	0.6588297
PYRIMIDINE METABOLISM	90	−0.27208874	−1.0883919	0.33539096	0.66017336
ENDOCYTOSIS	156	−0.25955555	−1.080931	0.3253012	0.6651845
FC EPSILON RI SIGNALING PATHWAY	62	−0.33219257	−1.0740509	0.35458168	0.66915274
INOSITOL PHOSPHATE METABOLISM	49	−0.29495132	−1.0660963	0.36134455	0.6755806
SNARE INTERACTIONS IN VESICULAR TRANSPORT	34	−0.33910057	−1.059901	0.40361446	0.6771761
LEUKOCYTE TRANSENDOTHELIAL MIGRATION	94	−0.3551334	−1.0428008	0.3970894	0.70556414
PARKINSONS DISEASE	103	−0.23767659	−1.0376086	0.39478958	0.70493156
ASCORBATE AND ALDARATE METABOLISM	18	−0.4042498	−1.0335944	0.41910332	0.7018473
FC GAMMA R MEDIATED PHAGOCYTOSIS	83	−0.29817593	−1.0322497	0.39959016	0.69348305
TIGHT JUNCTION	106	−0.27402833	−1.0200877	0.41908714	0.7086325
PATHWAYS IN CANCER	279	−0.26002625	−1.0173993	0.40248963	0.7030485
HEDGEHOG SIGNALING PATHWAY	39	−0.38406587	−1.0157797	0.40900195	0.69598156
OXIDATIVE PHOSPHORYLATION	107	−0.22497028	−1.0146433	0.42720306	0.68765503
PORPHYRIN AND CHLOROPHYLL METABOLISM	33	−0.33246934	−1.0108454	0.4486166	0.6846646
NEUROTROPHIN SIGNALING PATHWAY	114	−0.224331	−1.007843	0.4305835	0.6805984
EPITHELIAL CELL SIGNALING IN HELICOBACTER PYLORI INFECTION	65	−0.25337642	−0.9983357	0.4385246	0.69139266
COMPLEMENT AND COAGULATION CASCADES	52	−0.4048311	−0.9914725	0.46666667	0.69601685
RENAL CELL CARCINOMA	61	−0.27857324	−0.98580015	0.47572815	0.69756716
LINOLEIC ACID METABOLISM	16	−0.4168835	−0.97966015	0.47358122	0.7009005
GRAFT VERSUS HOST DISEASE	31	−0.47203773	−0.97773933	0.47912526	0.6949405
LONG TERM POTENTIATION	60	−0.29623806	−0.97161144	0.49275362	0.69866616
MELANOMA	58	−0.26884857	−0.9689658	0.52165353	0.6946373
PRIMARY IMMUNODEFICIENCY	29	−0.46399686	−0.9642447	0.5423387	0.6954509
ARACHIDONIC ACID METABOLISM	40	−0.3602123	−0.94764656	0.5665962	0.72030175
AXON GUIDANCE	119	−0.26357418	−0.9413953	0.55737704	0.723444
HUNTINGTONS DISEASE	157	−0.20142505	−0.9413445	0.53816044	0.7140554
CYTOKINE CYTOKINE RECEPTOR INTERACTION	183	−0.31934986	−0.93090165	0.5535714	0.7261811

ALLOGRAFT REJECTION	26	−0.48169452	−0.92606074	0.56363636	0.7267475
BASAL CELL CARCINOMA	40	−0.32223517	−0.9183921	0.59163344	0.733288
NOD LIKE RECEPTOR SIGNALING PATHWAY	48	−0.3053208	−0.9062703	0.5923077	0.7488158
DRUG METABOLISM OTHER ENZYMES	39	−0.30631727	−0.898475	0.6105675	0.755737
T CELL RECEPTOR SIGNALING PATHWAY	92	−0.26117206	−0.88521665	0.58171207	0.77357084
BUTANOATE METABOLISM	30	−0.35817483	−0.882406	0.62896824	0.76976794
TERPENOID BACKBONE BIOSYNTHESIS	15	−0.365085	−0.8818788	0.6188605	0.76160514
SYSTEMIC LUPUS ERYTHEMATOSUS	88	−0.31776744	−0.88100976	0.6051081	0.7542704
LYSOSOME	117	−0.23417728	−0.87776816	0.6113281	0.7519492
ECM RECEPTOR INTERACTION	70	−0.33946648	−0.86563253	0.64656967	0.7670777
RNA POLYMERASE	27	−0.26543185	−0.86558276	0.63723606	0.75846314
TYPE I DIABETES MELLITUS	32	−0.40558252	−0.8645699	0.6372549	0.75184464
TOLL LIKE RECEPTOR SIGNALING PATHWAY	76	−0.2723002	−0.86427546	0.63527054	0.74425393
AUTOIMMUNE THYROID DISEASE	34	−0.40742585	−0.8585488	0.6356275	0.7463378
JAK STAT SIGNALING PATHWAY	114	−0.24149683	−0.84474975	0.6859504	0.7643172
GALACTOSE METABOLISM	20	−0.24985434	−0.8222984	0.67701864	0.79740804
DORSO VENTRAL AXIS FORMATION	23	−0.27150398	−0.81813145	0.68136275	0.7966562
ASTHMA	21	−0.38351947	−0.7867748	0.7321063	0.8421756
PEROXISOME	67	−0.2196404	−0.7829821	0.7924528	0.8402849
INTESTINAL IMMUNE NETWORK FOR IGA PRODUCTION	34	−0.34740373	−0.7700342	0.7526427	0.85196656
PATHOGENIC ESCHERICHIA COLI INFECTION	50	−0.2641702	−0.7688507	0.7628866	0.8450977
O GLYCAN BIOSYNTHESIS	19	−0.28100225	−0.75185394	0.8367347	0.8638152
VIBRIO CHOLERAE INFECTION	51	−0.19520424	−0.7430977	0.8526971	0.868085
TASTE TRANSDUCTION	25	−0.2806909	−0.72318774	0.8623482	0.88615614
STEROID BIOSYNTHESIS	16	−0.2878557	−0.7064701	0.83657587	0.8987915
CYTOSOLIC DNA SENSING PATHWAY	42	−0.2202738	−0.6797214	0.86875	0.92031956
CELL CYCLE	108	−0.17019337	−0.63154733	0.9644269	0.9548837
GLYCOSYLPHOSPHATIDYLINOSITOL GPI ANCHOR BIOSYNTHESIS	24	−0.20134689	−0.5982353	0.868	0.9676463
PROTEASOME	41	−0.16021891	−0.54751354	0.8828125	0.97920084

Supplementary Table S14: GSEA of KEGG-pathways in the main cohort, enrichment in ERG_{high}

PATHWAY	SIZE	ES	NES	NOM *p*-val	FDR *q*-val
NOTCH SIGNALING PATHWAY	44	0.570646	1.8939956	0	0.055041593
PENTOSE AND GLUCURONATE INTERCONVERSIONS	22	0.539804	1.477712	0.07157895	1
CITRATE CYCLE TCA CYCLE	31	0.379956	1.4168941	0.097087376	1
SPLICEOSOME	106	0.315289	1.4022022	0.15665236	1
ENDOMETRIAL CANCER	45	0.37221	1.3950136	0.084337346	0.9674002
GLYCOSAMINOGLYCAN DEGRADATION	19	0.442775	1.3926028	0.120527305	0.8168873
RNA DEGRADATION	52	0.31642	1.3648297	0.15400411	0.82250255
TYPE II DIABETES MELLITUS	32	0.417495	1.3610706	0.057692308	0.7334166
ETHER LIPID METABOLISM	23	0.463213	1.2851268	0.14164905	0.97442824
SPHINGOLIPID METABOLISM	33	0.407494	1.283116	0.092929296	0.88598776
N GLYCAN BIOSYNTHESIS	43	0.387991	1.2674928	0.21501014	0.86686134
ADIPOCYTOKINE SIGNALING PATHWAY	56	0.352025	1.2573285	0.11890838	0.83355254
ABC TRANSPORTERS	35	0.405498	1.2435976	0.17958412	0.81864834
SELENOAMINO ACID METABOLISM	21	0.420925	1.2259183	0.25940594	0.82157844
VEGF SIGNALING PATHWAY	61	0.344259	1.2232822	0.16765286	0.77542835
GLYCOSAMINOGLYCAN BIOSYNTHESIS HEPARAN SULFATE	20	0.382288	1.219865	0.19569471	0.73947895
NITROGEN METABOLISM	16	0.533677	1.219607	0.1902834	0.6969087
PURINE METABOLISM	137	0.295115	1.2109817	0.15090543	0.68312067
GLYCEROLIPID METABOLISM	38	0.354304	1.1863607	0.17751479	0.720392
BASE EXCISION REPAIR	32	0.315088	1.1661458	0.24395162	0.7446007
PROGESTERONE MEDIATED OOCYTE MATURATION	74	0.331903	1.1658318	0.248	0.7097456
AMINOACYL TRNA BIOSYNTHESIS	40	0.340565	1.1531861	0.3215859	0.7137402
LYSINE DEGRADATION	41	0.305711	1.136898	0.27131784	0.72547805
ERBB SIGNALING PATHWAY	74	0.268035	1.1329427	0.23326571	0.7061263
LONG TERM DEPRESSION	53	0.325049	1.1232173	0.26185566	0.70176786
CHRONIC MYELOID LEUKEMIA	65	0.289681	1.1228493	0.29545453	0.6758072
GAP JUNCTION	73	0.349332	1.1220355	0.29400387	0.6528848
GLYCOSAMINOGLYCAN BIOSYNTHESIS CHONDROITIN SULFATE	19	0.43838	1.1136811	0.30812854	0.6488753
UBIQUITIN MEDIATED PROTEOLYSIS	127	0.247331	1.1013335	0.3417191	0.6578307
INSULIN SIGNALING PATHWAY	114	0.258285	1.1009762	0.3029703	0.63664
RIBOSOME	82	0.305605	1.0945854	0.37787056	0.63081235
PROTEIN EXPORT	22	0.379509	1.0859424	0.39918533	0.6306105
VASOPRESSIN REGULATED WATER REABSORPTION	38	0.288508	1.0723445	0.3472222	0.64245826
GLYOXYLATE AND DICARBOXYLATE METABOLISM	15	0.368145	1.0647148	0.4090909	0.63977957
BLADDER CANCER	41	0.338338	1.0609249	0.3580786	0.62997067
GLYCEROPHOSPHOLIPID METABOLISM	59	0.260367	1.0561461	0.36507937	0.622571
PENTOSE PHOSPHATE PATHWAY	24	0.308604	1.024085	0.44421053	0.6753275
ACUTE MYELOID LEUKEMIA	53	0.268619	1.0056236	0.43217054	0.6977903
RIG I LIKE RECEPTOR SIGNALING PATHWAY	54	0.284634	1.003308	0.44399184	0.6847781

GLYCINE SERINE AND THREONINE METABOLISM	29	0.363609	0.99739045	0.4529058	0.67987096
NEUROACTIVE LIGAND RECEPTOR INTERACTION	144	0.290831	0.9860564	0.4940476	0.6868037
MISMATCH REPAIR	22	0.296132	0.97391146	0.47268906	0.69548035
STARCH AND SUCROSE METABOLISM	38	0.297813	0.96517473	0.53195876	0.6982206
GNRH SIGNALING PATHWAY	84	0.261413	0.95846295	0.53154874	0.69538563
THYROID CANCER	29	0.288496	0.9438243	0.5331992	0.7112361
OOCYTE MEIOSIS	98	0.252336	0.94085276	0.5365854	0.7022131
BASAL TRANSCRIPTION FACTORS	28	0.243772	0.91830355	0.5484536	0.7325656
ONE CARBON POOL BY FOLATE	16	0.356132	0.9163467	0.552521	0.7210434
NON SMALL CELL LUNG CANCER	50	0.249019	0.9008779	0.5807087	0.7362479
TGF BETA SIGNALING PATHWAY	71	0.297991	0.88788795	0.6529081	0.74723715
SMALL CELL LUNG CANCER	77	0.248611	0.87011284	0.6429942	0.7668747
REGULATION OF AUTOPHAGY	24	0.226453	0.8181001	0.70623744	0.85259765
AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	41	0.236854	0.80509555	0.7211155	0.8610756
ALZHEIMERS DISEASE	141	0.201338	0.8043591	0.69246435	0.8465044
AMYOTROPHIC LATERAL SCLEROSIS ALS	44	0.249455	0.75974053	0.82077396	0.9104104
ALANINE ASPARTATE AND GLUTAMATE METABOLISM	28	0.253045	0.73419213	0.8477801	0.93348867
CYSTEINE AND METHIONINE METABOLISM	29	0.212136	0.72075784	0.8646465	0.9367848
HOMOLOGOUS RECOMBINATION	24	0.218325	0.7071195	0.8479657	0.9396755
FRUCTOSE AND MANNOSE METABOLISM	30	0.222205	0.70003974	0.88957053	0.93191516
BIOSYNTHESIS OF UNSATURATED FATTY ACIDS	20	0.278155	0.6484301	0.8965517	0.97063565
NUCLEOTIDE EXCISION REPAIR	42	0.1579	0.6299698	0.9227557	0.97013724
DNA REPLICATION	36	0.18302	0.5928331	0.9354839	0.9776261
OLFACTORY TRANSDUCTION	131	0.168927	0.4935225	0.989899	0.99249303

Supplementary Table S15: GSEA of KEGG-pathways in the main cohort, low Gleason (Gleason ≤ 3 + 4) samples, enrichment in ERG$_{low}$

PATHWAY	SIZE	ES	NES	NOM *p*-val	FDR *q*-val
FATTY ACID METABOLISM	35	−0.5748	−1.75733	0.003976	0.428658
VALINE LEUCINE AND ISOLEUCINE DEGRADATION	44	−0.67001	−1.73347	0.006342	0.270427
GLYCOLYSIS GLUCONEOGENESIS	56	−0.53319	−1.7318	0.00396	0.183222
DRUG METABOLISM CYTOCHROME P450	55	−0.64306	−1.6711	0.002024	0.243507
PROPANOATE METABOLISM	31	−0.53258	−1.66536	0.037815	0.209191
HEMATOPOIETIC CELL LINEAGE	68	−0.65124	−1.63053	0	0.240902
PPAR SIGNALING PATHWAY	51	−0.56741	−1.58869	0.002151	0.294395
TYROSINE METABOLISM	31	−0.55823	−1.51035	0.022587	0.467732
HISTIDINE METABOLISM	24	−0.62194	−1.50656	0.026639	0.430377
GLUTATHIONE METABOLISM	45	−0.55473	−1.49415	0.056863	0.42461
BETA ALANINE METABOLISM	19	−0.62699	−1.49271	0.056604	0.389144
METABOLISM OF XENOBIOTICS BY CYTOCHROME P450	50	−0.53636	−1.48458	0.02988	0.377124
RETINOL METABOLISM	40	−0.52352	−1.47665	0.02268	0.369106
CARDIAC MUSCLE CONTRACTION	58	−0.54267	−1.4481	0.076152	0.417485
HYPERTROPHIC CARDIOMYOPATHY HCM	66	−0.54303	−1.43106	0.057613	0.436884
PYRUVATE METABOLISM	38	−0.3643	−1.41729	0.085774	0.44844
DILATED CARDIOMYOPATHY	73	−0.52784	−1.39136	0.08125	0.498288
NICOTINATE AND NICOTINAMIDE METABOLISM	19	−0.63499	−1.37941	0.079399	0.50922
ARACHIDONIC ACID METABOLISM	40	−0.49731	−1.36389	0.025105	0.526413
VIRAL MYOCARDITIS	62	−0.52378	−1.36242	0.130952	0.50457
CELL ADHESION MOLECULES CAMS	112	−0.51387	−1.34666	0.116564	0.523981
B CELL RECEPTOR SIGNALING PATHWAY	70	−0.41894	−1.3323	0.165975	0.54144
BUTANOATE METABOLISM	30	−0.50098	−1.32435	0.136821	0.540635
PRIMARY IMMUNODEFICIENCY	29	−0.59903	−1.31789	0.184874	0.535373
PROXIMAL TUBULE BICARBONATE RECLAMATION	18	−0.4973	−1.30566	0.121339	0.549594
FC EPSILON RI SIGNALING PATHWAY	62	−0.39485	−1.29086	0.122699	0.572681
VASCULAR SMOOTH MUSCLE CONTRACTION	89	−0.39253	−1.2908	0.164905	0.551471
REGULATION OF ACTIN CYTOSKELETON	176	−0.34923	−1.28702	0.145791	0.541866
GLYCOSPHINGOLIPID BIOSYNTHESIS LACTO AND NEOLACTO SERIES	17	−0.49558	−1.28193	0.158635	0.535908
LEISHMANIA INFECTION	58	−0.48024	−1.26	0.188017	0.579365
OXIDATIVE PHOSPHORYLATION	107	−0.25188	−1.25928	0.223158	0.563622
LINOLEIC ACID METABOLISM	16	−0.52279	−1.25469	0.152361	0.556829
PHENYLALANINE METABOLISM	17	−0.5513	−1.25461	0.163223	0.540131
LEUKOCYTE TRANSENDOTHELIAL MIGRATION	94	−0.42807	−1.25319	0.160825	0.527582
ADHERENS JUNCTION	62	−0.35074	−1.24188	0.177291	0.539059
NATURAL KILLER CELL MEDIATED CYTOTOXICITY	101	−0.40518	−1.24066	0.214592	0.527223
EPITHELIAL CELL SIGNALING IN HELICOBACTER PYLORI INFECTION	65	−0.28888	−1.24058	0.131466	0.5131
PARKINSONS DISEASE	103	−0.26831	−1.24027	0.235294	0.500656
FC GAMMA R MEDIATED PHAGOCYTOSIS	83	−0.34748	−1.22995	0.207039	0.512762
CHEMOKINE SIGNALING PATHWAY	152	−0.39003	−1.22606	0.202479	0.509208
FOCAL ADHESION	168	−0.37517	−1.22213	0.194861	0.505357
PHOSPHATIDYLINOSITOL SIGNALING SYSTEM	65	−0.34458	−1.2177	0.169661	0.50295

PANCREATIC CANCER	63	−0.34855	−1.21172	0.208511	0.505366
ENDOCYTOSIS	156	−0.28012	−1.20086	0.170782	0.517217
ANTIGEN PROCESSING AND PRESENTATION	65	−0.46206	−1.19376	0.287234	0.522465
PYRIMIDINE METABOLISM	90	−0.2891	−1.18693	0.221766	0.525949
ARGININE AND PROLINE METABOLISM	46	−0.40073	−1.18367	0.189655	0.522112
PRION DISEASES	30	−0.43579	−1.18351	0.240816	0.511485
ASCORBATE AND ALDARATE METABOLISM	18	−0.47857	−1.18189	0.273092	0.504342
SYSTEMIC LUPUS ERYTHEMATOSUS	88	−0.41173	−1.17882	0.237323	0.500253
STEROID HORMONE BIOSYNTHESIS	36	−0.43597	−1.17244	0.224	0.503355
WNT SIGNALING PATHWAY	129	−0.31684	−1.1669	0.225873	0.505087
TIGHT JUNCTION	106	−0.32125	−1.14441	0.243083	0.543737
MTOR SIGNALING PATHWAY	43	−0.29399	−1.14222	0.273101	0.537842
MELANOGENESIS	82	−0.34923	−1.13872	0.264151	0.535669
GLYCEROLIPID METABOLISM	38	−0.33329	−1.12194	0.278481	0.561586
LYSOSOME	117	−0.28709	−1.1122	0.317328	0.571848
MAPK SIGNALING PATHWAY	218	−0.28331	−1.10765	0.269737	0.571201
ALDOSTERONE REGULATED SODIUM REABSORPTION	33	−0.45758	−1.10048	0.335541	0.576414
INOSITOL PHOSPHATE METABOLISM	49	−0.31283	−1.09895	0.331959	0.569941
TOLL LIKE RECEPTOR SIGNALING PATHWAY	76	−0.32024	−1.09699	0.316222	0.565621
COLORECTAL CANCER	54	−0.29807	−1.08723	0.327766	0.575742
SNARE INTERACTIONS IN VESICULAR TRANSPORT	34	−0.3634	−1.07756	0.382892	0.586732
TYPE I DIABETES MELLITUS	32	−0.52346	−1.07534	0.40257	0.582921
APOPTOSIS	75	−0.26781	−1.06279	0.335484	0.600432
VIBRIO CHOLERAE INFECTION	51	−0.26823	−1.06249	0.331276	0.591874
CYTOKINE CYTOKINE RECEPTOR INTERACTION	183	−0.35607	−1.05478	0.377682	0.598656
TRYPTOPHAN METABOLISM	33	−0.40505	−1.05363	0.368973	0.592153
CALCIUM SIGNALING PATHWAY	124	−0.31068	−1.05021	0.368644	0.590896
HEDGEHOG SIGNALING PATHWAY	39	−0.38985	−1.04884	0.378099	0.584801
T CELL RECEPTOR SIGNALING PATHWAY	92	−0.28932	−1.04161	0.397872	0.591842
AUTOIMMUNE THYROID DISEASE	34	−0.49558	−1.04001	0.440426	0.586462
ASTHMA	21	−0.50737	−1.03382	0.428571	0.59034
ALLOGRAFT REJECTION	26	−0.54805	−1.0317	0.471215	0.586535
AXON GUIDANCE	119	−0.27966	−1.03134	0.36646	0.579415
JAK STAT SIGNALING PATHWAY	114	−0.28618	−1.02503	0.417671	0.584776
ARRHYTHMOGENIC RIGHT VENTRICULAR CARDIOMYOPATHY ARVC	59	−0.34725	−1.02209	0.430353	0.582792
GLIOMA	57	−0.28234	−1.02111	0.436059	0.57718
COMPLEMENT AND COAGULATION CASCADES	52	−0.42698	−1.02096	0.448637	0.570092
STARCH AND SUCROSE METABOLISM	38	−0.316	−1.02012	0.415612	0.564567
GRAFT VERSUS HOST DISEASE	31	−0.50635	−1.01979	0.487288	0.558214
GAP JUNCTION	73	−0.30982	−1.01693	0.409871	0.556332
ADIPOCYTOKINE SIGNALING PATHWAY	56	−0.27146	−1.01226	0.435374	0.558633
PATHWAYS IN CANCER	279	−0.25379	−1.00408	0.418605	0.567434
LONG TERM DEPRESSION	53	−0.29765	−0.99491	0.431535	0.577069
NON SMALL CELL LUNG CANCER	50	−0.27431	−0.99275	0.494759	0.574468
PATHOGENIC ESCHERICHIA COLI INFECTION	50	−0.32015	−0.9716	0.483607	0.606761
GNRH SIGNALING PATHWAY	84	−0.26309	−0.96915	0.493802	0.605055

MELANOMA	58	−0.27091	−0.95161	0.561181	0.630692
LONG TERM POTENTIATION	60	−0.27998	−0.9468	0.544699	0.632693
FRUCTOSE AND MANNOSE METABOLISM	30	−0.29582	−0.94543	0.543651	0.628576
GLYCEROPHOSPHOLIPID METABOLISM	59	−0.23849	−0.93851	0.57551	0.635078
HUNTINGTONS DISEASE	157	−0.19111	−0.93586	0.507592	0.633028
RENAL CELL CARCINOMA	61	−0.25307	−0.93248	0.541667	0.632276
BASAL CELL CARCINOMA	40	−0.31747	−0.92139	0.60334	0.64651
ALZHEIMERS DISEASE	141	−0.22165	−0.9201	0.53139	0.642262
VEGF SIGNALING PATHWAY	61	−0.25292	−0.91574	0.622129	0.643416
TERPENOID BACKBONE BIOSYNTHESIS	15	−0.36513	−0.90732	0.578189	0.650887
PROSTATE CANCER	83	−0.24162	−0.90552	0.597895	0.647442
PORPHYRIN AND CHLOROPHYLL METABOLISM	33	−0.30215	−0.89355	0.612774	0.662094
PEROXISOME	67	−0.24244	−0.88397	0.659919	0.67278
AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	41	−0.24287	−0.88212	0.675676	0.669515
GALACTOSE METABOLISM	20	−0.2704	−0.87744	0.612971	0.671198
DRUG METABOLISM OTHER ENZYMES	39	−0.30971	−0.87682	0.638211	0.665909
INTESTINAL IMMUNE NETWORK FOR IGA PRODUCTION	34	−0.3731	−0.84708	0.645435	0.711469
STEROID BIOSYNTHESIS	16	−0.34029	−0.81211	0.696121	0.764039
RNA POLYMERASE	27	−0.23329	−0.79392	0.72211	0.787158
P53 SIGNALING PATHWAY	61	−0.25305	−0.78393	0.795031	0.796277
ALANINE ASPARTATE AND GLUTAMATE METABOLISM	28	−0.26272	−0.78148	0.796334	0.792567
BIOSYNTHESIS OF UNSATURATED FATTY ACIDS	20	−0.30046	−0.72265	0.814196	0.869183
GLYCOSYLPHOSPHATIDYLINOSITOL GPI ANCHOR BIOSYNTHESIS	24	−0.20336	−0.65816	0.869121	0.929164
PROTEASOME	41	−0.12487	−0.47125	0.949686	0.995191

Supplementary Table S16: GSEA of KEGG-pathways in the main cohort, low Gleason (Gleason ≤ 3 + 4) samples, enrichment in ERG$_{high}$

PATHWAY	SIZE	ES	NES	NOM *p*-val	FDR *q*-val
NOTCH SIGNALING PATHWAY	44	0.516836	1.717935	0.001942	0.301009
AMINOACYL TRNA BIOSYNTHESIS	40	0.462563	1.56396	0.059917	0.617156
RNA DEGRADATION	52	0.354821	1.419818	0.118609	1
NITROGEN METABOLISM	16	0.588849	1.329333	0.118367	1
CITRATE CYCLE TCA CYCLE	31	0.329703	1.304396	0.142342	1
SPHINGOLIPID METABOLISM	33	0.417907	1.277418	0.118952	1
ABC TRANSPORTERS	35	0.413801	1.277238	0.148148	1
ONE CARBON POOL BY FOLATE	16	0.446341	1.241598	0.219959	1
TYPE II DIABETES MELLITUS	32	0.409184	1.20927	0.201195	1
PENTOSE PHOSPHATE PATHWAY	24	0.331508	1.207402	0.213018	1
GLYOXYLATE AND DICARBOXYLATE METABOLISM	15	0.400063	1.203525	0.245902	1
ETHER LIPID METABOLISM	23	0.443994	1.194557	0.198853	1
GLYCOSAMINOGLYCAN BIOSYNTHESIS HEPARAN SULFATE	20	0.343143	1.071101	0.349693	1
SPLICEOSOME	106	0.22487	1.047405	0.391837	1
GLYCOSAMINOGLYCAN DEGRADATION	19	0.358777	1.041911	0.423554	1
N GLYCAN BIOSYNTHESIS	43	0.283753	1.040282	0.413927	1
SMALL CELL LUNG CANCER	77	0.299533	1.031481	0.406977	1
BLADDER CANCER	41	0.317529	1.031392	0.412121	1
OLFACTORY TRANSDUCTION	131	0.321648	1.020145	0.440162	1
BASE EXCISION REPAIR	32	0.27957	1.018931	0.430672	1
PURINE METABOLISM	137	0.229031	1.012979	0.449807	1
PENTOSE AND GLUCURONATE INTERCONVERSIONS	22	0.375712	0.980591	0.516832	1
NEUROACTIVE LIGAND RECEPTOR INTERACTION	144	0.287321	0.970639	0.497154	1
DNA REPLICATION	36	0.306206	0.965192	0.480808	1
GLYCOSAMINOGLYCAN BIOSYNTHESIS CHONDROITIN SULFATE	19	0.374668	0.959322	0.517578	1
LYSINE DEGRADATION	41	0.276211	0.950357	0.538023	1
TGF BETA SIGNALING PATHWAY	71	0.323419	0.932605	0.573674	1
DORSO VENTRAL AXIS FORMATION	23	0.318803	0.913324	0.569573	1
UBIQUITIN MEDIATED PROTEOLYSIS	127	0.19352	0.909368	0.55814	1
ECM RECEPTOR INTERACTION	70	0.341929	0.900604	0.611632	1
GLYCINE SERINE AND THREONINE METABOLISM	29	0.31782	0.899822	0.619718	1
PROGESTERONE MEDIATED OOCYTE MATURATION	74	0.24859	0.88205	0.654297	1
MISMATCH REPAIR	22	0.267798	0.872425	0.608696	1
CELL CYCLE	108	0.229027	0.872099	0.646943	1
ERBB SIGNALING PATHWAY	74	0.192177	0.860207	0.731313	1
RIG I LIKE RECEPTOR SIGNALING PATHWAY	54	0.230037	0.852649	0.697495	1
OOCYTE MEIOSIS	98	0.220263	0.847816	0.732284	1
VASOPRESSIN REGULATED WATER REABSORPTION	38	0.240948	0.844498	0.646341	1
AMYOTROPHIC LATERAL SCLEROSIS ALS	44	0.281079	0.83691	0.672414	1

INSULIN SIGNALING PATHWAY	114	0.187648	0.79114	0.801181	1
O GLYCAN BIOSYNTHESIS	19	0.275072	0.784538	0.810176	1
HOMOLOGOUS RECOMBINATION	24	0.236923	0.777339	0.755144	1
NEUROTROPHIN SIGNALING PATHWAY	114	0.179373	0.77006	0.825203	1
PROTEIN EXPORT	22	0.236003	0.764155	0.708098	1
CYSTEINE AND METHIONINE METABOLISM	29	0.220384	0.753994	0.841176	1
BASAL TRANSCRIPTION FACTORS	28	0.19441	0.750105	0.825203	1
SELENOAMINO ACID METABOLISM	21	0.244671	0.747654	0.793587	0.996678
RIBOSOME	82	0.195386	0.737388	0.690909	0.991515
ENDOMETRIAL CANCER	45	0.199174	0.732926	0.799599	0.978085
CHRONIC MYELOID LEUKEMIA	65	0.193513	0.729681	0.813627	0.963081
ACUTE MYELOID LEUKEMIA	53	0.198753	0.728204	0.884086	0.946406
TASTE TRANSDUCTION	25	0.243484	0.68766	0.948077	0.979054
NOD LIKE RECEPTOR SIGNALING PATHWAY	48	0.215297	0.672106	0.900778	0.978148
NUCLEOTIDE EXCISION REPAIR	42	0.155527	0.611816	0.902153	1
THYROID CANCER	29	0.193289	0.605692	0.926878	0.993572
REGULATION OF AUTOPHAGY	24	0.177812	0.605173	0.937736	0.97612
CYTOSOLIC DNA SENSING PATHWAY	42	0.143015	0.491659	0.992481	0.99366

Supplementary Table S17: GSEA of KEGG-pathways in the main cohort, high Gleason (Gleason ≥ 4 + 3) samples, enrichment in ERG$_{low}$

PATHWAY	SIZE	ES	NES	NOM *p*-val	FDR *q*-val
GLUTATHIONE METABOLISM	45	−0.58109	−1.49226	0.062753	1
DRUG METABOLISM CYTOCHROME P450	55	−0.52313	−1.45852	0.018256	1
GLYCOLYSIS GLUCONEOGENESIS	56	−0.40615	−1.44262	0.053254	1
P53 SIGNALING PATHWAY	61	−0.45124	−1.38952	0.070342	1
APOPTOSIS	75	−0.33517	−1.38622	0.045455	1
BETA ALANINE METABOLISM	19	−0.53945	−1.33119	0.124	1
ARGININE AND PROLINE METABOLISM	46	−0.48317	−1.31523	0.110442	1
GLYCOSPHINGOLIPID BIOSYNTHESIS LACTO AND NEOLACTO SERIES	17	−0.46753	−1.30379	0.120623	1
HISTIDINE METABOLISM	24	−0.54079	−1.3033	0.09	1
MELANOGENESIS	82	−0.38026	−1.29089	0.093361	1
ALDOSTERONE REGULATED SODIUM REABSORPTION	33	−0.53263	−1.27956	0.136095	1
HYPERTROPHIC CARDIOMYOPATHY HCM	66	−0.43965	−1.27941	0.187373	0.9968
METABOLISM OF XENOBIOTICS BY CYTOCHROME P450	50	−0.43719	−1.25665	0.117647	1
PYRUVATE METABOLISM	38	−0.36316	−1.23956	0.208413	1
PROSTATE CANCER	83	−0.30743	−1.21731	0.155419	1
MTOR SIGNALING PATHWAY	43	−0.30711	−1.18744	0.1917	1
O GLYCAN BIOSYNTHESIS	19	−0.45238	−1.18317	0.239837	1
DILATED CARDIOMYOPATHY	73	−0.40563	−1.18086	0.264463	1
TYROSINE METABOLISM	31	−0.43806	−1.17859	0.211765	0.984351
PRION DISEASES	30	−0.44281	−1.17312	0.29505	0.957145
CARDIAC MUSCLE CONTRACTION	58	−0.34436	−1.14871	0.26938	1
PPAR SIGNALING PATHWAY	51	−0.39044	−1.12762	0.262425	1
GLIOMA	57	−0.31679	−1.12513	0.25102	1
TASTE TRANSDUCTION	25	−0.42644	−1.07992	0.373016	1
STEROID HORMONE BIOSYNTHESIS	36	−0.37043	−1.07989	0.325671	1
FATTY ACID METABOLISM	35	−0.3198	−1.07466	0.351515	1
PYRIMIDINE METABOLISM	90	−0.26562	−1.07115	0.339286	1
CYSTEINE AND METHIONINE METABOLISM	29	−0.31216	−1.0571	0.397895	1
HEMATOPOIETIC CELL LINEAGE	68	−0.42636	−1.046	0.397638	1
TRYPTOPHAN METABOLISM	33	−0.39935	−1.03939	0.393375	1
NICOTINATE AND NICOTINAMIDE METABOLISM	19	−0.4544	−1.03812	0.371717	1
VALINE LEUCINE AND ISOLEUCINE DEGRADATION	44	−0.38296	−1.03049	0.423387	1
PROXIMAL TUBULE BICARBONATE RECLAMATION	18	−0.39738	−1.0253	0.41966	0.996142
GLYCINE SERINE AND THREONINE METABOLISM	29	−0.38587	−1.01423	0.426195	1
ARACHIDONIC ACID METABOLISM	40	−0.38977	−0.99708	0.441955	1
ARRHYTHMOGENIC RIGHT VENTRICULAR CARDIOMYOPATHY ARVC	59	−0.34466	−0.98655	0.48394	1
AMYOTROPHIC LATERAL SCLEROSIS ALS	44	−0.30689	−0.96829	0.48855	1
PHENYLALANINE METABOLISM	17	−0.41873	−0.96135	0.516	1
LINOLEIC ACID METABOLISM	16	−0.39795	−0.94534	0.564777	1

NOD LIKE RECEPTOR SIGNALING PATHWAY	48	−0.31988	−0.93946	0.53229	1
OLFACTORY TRANSDUCTION	131	−0.33618	−0.93875	0.53816	1
ADHERENS JUNCTION	62	−0.25968	−0.93686	0.537473	1
FOCAL ADHESION	168	−0.29155	−0.93682	0.521739	1
TERPENOID BACKBONE BIOSYNTHESIS	15	−0.39638	−0.93531	0.560078	0.984399
PROPANOATE METABOLISM	31	−0.31655	−0.92299	0.546667	0.99571
WNT SIGNALING PATHWAY	129	−0.25453	−0.92235	0.572614	0.975564
PHOSPHATIDYLINOSITOL SIGNALING SYSTEM	65	−0.26402	−0.91989	0.566596	0.961487
RETINOL METABOLISM	40	−0.29044	−0.91814	0.634	0.945594
ALANINE ASPARTATE AND GLUTAMATE METABOLISM	28	−0.32715	−0.9138	0.59802	0.936682
CALCIUM SIGNALING PATHWAY	124	−0.2778	−0.91263	0.577963	0.9206
PATHWAYS IN CANCER	279	−0.23088	−0.9015	0.640657	0.931577
PORPHYRIN AND CHLOROPHYLL METABOLISM	33	−0.29245	−0.89562	0.598441	0.928584
AXON GUIDANCE	119	−0.25724	−0.89225	0.664	0.918983
LONG TERM POTENTIATION	60	−0.27098	−0.88623	0.655319	0.916726
CYTOSOLIC DNA SENSING PATHWAY	42	−0.28958	−0.88511	0.591356	0.902483
DRUG METABOLISM OTHER ENZYMES	39	−0.29008	−0.88207	0.672549	0.893022
PANCREATIC CANCER	63	−0.24008	−0.87586	0.667355	0.891418
DORSO VENTRAL AXIS FORMATION	23	−0.26575	−0.87206	0.633401	0.884406
VASCULAR SMOOTH MUSCLE CONTRACTION	89	−0.25807	−0.84826	0.658174	0.921096
HEDGEHOG SIGNALING PATHWAY	39	−0.30239	−0.81983	0.735537	0.968674
COLORECTAL CANCER	54	−0.22429	−0.81853	0.772257	0.955406
COMPLEMENT AND COAGULATION CASCADES	52	−0.33958	−0.8171	0.68	0.942883
BLADDER CANCER	41	−0.26128	−0.81446	0.761044	0.933214
HUNTINGTONS DISEASE	157	−0.17929	−0.81213	0.683168	0.923699
NUCLEOTIDE EXCISION REPAIR	42	−0.20196	−0.80325	0.714004	0.927309
DNA REPLICATION	36	−0.24967	−0.79526	0.678208	0.929258
CYTOKINE CYTOKINE RECEPTOR INTERACTION	183	−0.27434	−0.7898	0.749515	0.925759
CHEMOKINE SIGNALING PATHWAY	152	−0.26392	−0.77536	0.762295	0.938934
EPITHELIAL CELL SIGNALING IN HELICOBACTER PYLORI INFECTION	65	−0.1986	−0.76677	0.833652	0.941243
RENAL CELL CARCINOMA	61	−0.20876	−0.75937	0.79065	0.940721
CELL CYCLE	108	−0.19763	−0.7411	0.85567	0.957485
GALACTOSE METABOLISM	20	−0.22163	−0.73411	0.790476	0.955116
SMALL CELL LUNG CANCER	77	−0.20928	−0.72354	0.89834	0.957566
B CELL RECEPTOR SIGNALING PATHWAY	70	−0.23307	−0.72088	0.819106	0.948205
RNA POLYMERASE	27	−0.2329	−0.69867	0.792887	0.965847
CELL ADHESION MOLECULES CAMS	112	−0.25326	−0.66414	0.89	0.991479
HOMOLOGOUS RECOMBINATION	24	−0.19277	−0.6397	0.932939	1
PROTEASOME	41	−0.18083	−0.62023	0.806911	1
ANTIGEN PROCESSING AND PRESENTATION	65	−0.22812	−0.59258	0.919075	1
PARKINSONS DISEASE	103	−0.13483	−0.57441	0.925996	1
OXIDATIVE PHOSPHORYLATION	107	−0.13049	−0.55556	0.900952	0.99953
GLYCOSYLPHOSPHATIDYLINOSITOL GPI ANCHOR BIOSYNTHESIS	24	−0.17695	−0.51881	0.924419	0.998504
AMINOACYL TRNA BIOSYNTHESIS	40	−0.12514	−0.40487	0.993776	0.998799

Supplementary Table S18: GSEA of KEGG-pathways in the main cohort, high Gleason (Gleason ≥ 4 + 3) samples, enrichment in ERG$_{high}$

PATHWAY	SIZE	ES	NES	NOM p-val	FDR q-val
NOTCH SIGNALING PATHWAY	44	0.5045062	1.6738925	0	0.62938935
PENTOSE AND GLUCURONATE INTERCONVERSIONS	22	0.5123342	1.4166561	0.08817204	1
GLYCOSAMINOGLYCAN BIOSYNTHESIS CHONDROITIN SULFATE	19	0.542504	1.3943979	0.07114624	1
SELENOAMINO ACID METABOLISM	21	0.47533673	1.3648516	0.10714286	1
ENDOMETRIAL CANCER	45	0.36318752	1.3550841	0.086519115	1
PROTEIN EXPORT	22	0.48945096	1.3471158	0.19421488	1
SPLICEOSOME	106	0.31644973	1.3434168	0.17	1
ETHER LIPID METABOLISM	23	0.44876847	1.3325243	0.06832298	1
GAP JUNCTION	73	0.40344366	1.2925855	0.13872832	1
LONG TERM DEPRESSION	53	0.35514474	1.2659172	0.1002004	1
TYPE II DIABETES MELLITUS	32	0.36255944	1.2274318	0.15039062	1
UBIQUITIN MEDIATED PROTEOLYSIS	127	0.27910286	1.227381	0.24418604	1
TIGHT JUNCTION	106	0.2995418	1.1919937	0.18426104	1
GLYOXYLATE AND DICARBOXYLATE METABOLISM	15	0.41961658	1.1883075	0.28985506	1
PROGESTERONE MEDIATED OOCYTE MATURATION	74	0.329471	1.171812	0.25494072	1
LYSINE DEGRADATION	41	0.2979997	1.171253	0.19444445	1
ABC TRANSPORTERS	35	0.3798634	1.1644273	0.2576336	1
RNA DEGRADATION	52	0.2598646	1.1500181	0.30754352	1
RIBOSOME	82	0.3375702	1.129038	0.3858586	1
GLYCEROLIPID METABOLISM	38	0.3245482	1.1109859	0.3046092	1
STARCH AND SUCROSE METABOLISM	38	0.34416395	1.1031003	0.32040817	1
GLYCEROPHOSPHOLIPID METABOLISM	59	0.2613262	1.091921	0.28870293	1
RIG I LIKE RECEPTOR SIGNALING PATHWAY	54	0.3146344	1.0732448	0.35976788	1
CITRATE CYCLE TCA CYCLE	31	0.29832336	1.0654979	0.39300412	1
VEGF SIGNALING PATHWAY	61	0.3033966	1.0607156	0.37058824	1
SPHINGOLIPID METABOLISM	33	0.3293276	1.0508296	0.35849056	1
PURINE METABOLISM	137	0.26441482	1.0403163	0.39648438	1
N GLYCAN BIOSYNTHESIS	43	0.32496336	1.0089209	0.45759368	1
VIBRIO CHOLERAE INFECTION	51	0.2632098	0.99514633	0.43838385	1
GLYCOSAMINOGLYCAN BIOSYNTHESIS HEPARAN SULFATE	20	0.3191126	0.9949526	0.49115914	1
GLYCOSAMINOGLYCAN DEGRADATION	19	0.3005399	0.965958	0.5010309	1
THYROID CANCER	29	0.29003474	0.9657133	0.5121951	1
BASE EXCISION REPAIR	32	0.26781192	0.963137	0.49707603	1
MISMATCH REPAIR	22	0.303365	0.95994514	0.5051546	1
ALZHEIMERS DISEASE	141	0.23473425	0.94902	0.5030303	1
MAPK SIGNALING PATHWAY	218	0.2395718	0.94869226	0.560241	1
BASAL TRANSCRIPTION FACTORS	28	0.25926867	0.9475964	0.5235294	1
NON SMALL CELL LUNG CANCER	50	0.27112022	0.9459981	0.52566737	1
VASOPRESSIN REGULATED WATER REABSORPTION	38	0.23592153	0.93527263	0.5808967	1
GNRH SIGNALING PATHWAY	84	0.25247583	0.92241526	0.5753425	1
OOCYTE MEIOSIS	98	0.24481033	0.9178716	0.5795678	1

ERBB SIGNALING PATHWAY	74	0.22490917	0.9112399	0.6229508	1
NEUROACTIVE LIGAND RECEPTOR INTERACTION	144	0.2674146	0.90514165	0.68801653	1
AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	41	0.28475586	0.89959615	0.57758623	1
ASCORBATE AND ALDARATE METABOLISM	18	0.3372187	0.887614	0.6293996	1
TGF BETA SIGNALING PATHWAY	71	0.28805873	0.8791876	0.64176244	1
T CELL RECEPTOR SIGNALING PATHWAY	92	0.27073228	0.8785631	0.61044174	1
PENTOSE PHOSPHATE PATHWAY	24	0.27586928	0.87095606	0.6260504	1
ENDOCYTOSIS	156	0.20839994	0.86836064	0.6947162	1
FC EPSILON RI SIGNALING PATHWAY	62	0.272123	0.8679006	0.612326	1
ADIPOCYTOKINE SIGNALING PATHWAY	56	0.254351	0.8639951	0.70726913	1
ECM RECEPTOR INTERACTION	70	0.33320397	0.8552361	0.6287425	1
TYPE I DIABETES MELLITUS	32	0.4039086	0.84916973	0.64123714	1
JAK STAT SIGNALING PATHWAY	114	0.24348678	0.8405615	0.6825397	1
SYSTEMIC LUPUS ERYTHEMATOSUS	88	0.31594077	0.8378146	0.69896907	1
MELANOMA	58	0.23131153	0.8347135	0.80846775	1
INSULIN SIGNALING PATHWAY	114	0.19043854	0.83018816	0.74810606	1
SNARE INTERACTIONS IN VESICULAR TRANSPORT	34	0.24903408	0.82990885	0.7033399	1
NITROGEN METABOLISM	16	0.36240193	0.82630676	0.7225807	1
PEROXISOME	67	0.23194745	0.82561415	0.7344961	0.99797565
REGULATION OF AUTOPHAGY	24	0.22092693	0.82355046	0.7057654	0.9860101
BASAL CELL CARCINOMA	40	0.28743526	0.82145005	0.78149605	0.97474784
CHRONIC MYELOID LEUKEMIA	65	0.20110641	0.8113076	0.7887324	0.981718
NEUROTROPHIN SIGNALING PATHWAY	114	0.17418417	0.81061876	0.8224852	0.9678766
ASTHMA	21	0.3933154	0.79921764	0.7057654	0.9763035
ONE CARBON POOL BY FOLATE	16	0.32356194	0.798125	0.68604654	0.9636205
INOSITOL PHOSPHATE METABOLISM	49	0.2127568	0.7812778	0.8507752	0.9808415
LYSOSOME	117	0.21278796	0.7681107	0.8032787	0.99078923
BIOSYNTHESIS OF UNSATURATED FATTY ACIDS	20	0.31573716	0.72537094	0.8079332	1
LEISHMANIA INFECTION	58	0.28457937	0.7233578	0.8353909	1
BUTANOATE METABOLISM	30	0.28861216	0.7105619	0.85626286	1
PATHOGENIC ESCHERICHIA COLI INFECTION	50	0.24425834	0.70852464	0.8627859	1
LEUKOCYTE TRANSENDOTHELIAL MIGRATION	94	0.24133033	0.7067604	0.8313953	1
REGULATION OF ACTIN CYTOSKELETON	176	0.17993955	0.6648534	0.9596154	1
TOLL LIKE RECEPTOR SIGNALING PATHWAY	76	0.21081783	0.65017754	0.9047619	1
FC GAMMA R MEDIATED PHAGOCYTOSIS	83	0.19004686	0.6495817	0.90335304	1
ACUTE MYELOID LEUKEMIA	53	0.16568136	0.6410138	0.9823875	1
FRUCTOSE AND MANNOSE METABOLISM	30	0.19758251	0.6064993	0.95669293	1
NATURAL KILLER CELL MEDIATED CYTOTOXICITY	101	0.2088917	0.6036799	0.92622954	1
PRIMARY IMMUNODEFICIENCY	29	0.29064822	0.58000416	0.88188976	1
ALLOGRAFT REJECTION	26	0.3025116	0.57801336	0.93890023	1
INTESTINAL IMMUNE NETWORK FOR IGA PRODUCTION	34	0.261487	0.5624501	0.94949496	1
GRAFT VERSUS HOST DISEASE	31	0.26586834	0.5504549	0.9738956	1
STEROID BIOSYNTHESIS	16	0.22409485	0.55034274	0.96825397	1
AUTOIMMUNE THYROID DISEASE	34	0.26633847	0.5500026	0.9534413	0.98951966
VIRAL MYOCARDITIS	62	0.21385355	0.5458711	0.9785575	0.9793871

Supplementary Table S19: Pearson correlations between individual metabolites and selected tissue composition parameters among cancer samples in the main cohort

Metabolites	Cancer content		Stroma content		Benign epithelium content		Relative luminal space	
	Correlations	*p*-value	Correlations	*p*-value	Correlations	*p*-value	Correlations	*p*-value
Alanine	0.2295	0.0253	–0.2269	0.027	–0.0803	0.439	–0.0087	0.9337
Choline	0.5395	< 0.001	–0.5233	< 0.001	–0.3218	0.0002	–0.2185	0.0132
Citrate	–0.0618	0.5518	–0.1113	0.2828	0.2662	0.0091	0.436	< 0.001
Creatine	–0.0642	0.4699	0.0024	0.9783	0.1075	0.2254	0.0397	0.6564
Ethm	–0.0341	0.701	0.0277	0.7557	0.0267	0.7643	–0.0687	0.4412
Glucose	–0.6205	< 0.001	0.514	< 0.001	0.4716	< 0.001	0.1263	0.1554
Glutamate	0.5898	< 0.001	–0.4871	< 0.001	–0.45	< 0.001	–0.1672	0.0592
Glutamine	0.4315	< 0.001	–0.3368	0.0001	–0.3519	< 0.001	–0.2222	0.0117
Glycine	–0.4881	< 0.001	0.3687	< 0.001	0.4122	< 0.001	0.2114	0.0166
GPC	0.491	< 0.001	–0.4792	< 0.001	–0.2894	0.0009	–0.0006	0.9946
GPE	–0.0362	0.6839	0.0357	0.6876	0.0208	0.8146	–0.0531	0.5518
Isoleucine	0.2005	0.0227	–0.1439	0.1036	–0.178	0.0436	–0.1519	0.087
Lactate	0.4693	< 0.001	–0.4064	< 0.001	–0.338	0.0001	–0.1843	0.038
Leucine	0.368	< 0.001	–0.2978	0.0006	–0.2878	0.0009	–0.2375	0.007
Myo–inositol	0.0784	0.3771	–0.1006	0.2568	–0.0184	0.8359	–0.0644	0.4704
PCh	0.4436	< 0.001	–0.2875	0.001	–0.4296	< 0.001	–0.1757	0.0472
PE	0.3601	< 0.001	–0.2105	0.0167	–0.3752	< 0.001	–0.2391	0.0066
Putrescine	–0.3059	0.0004	0.0277	0.7555	0.4935	< 0.001	0.1555	0.0796
Scyllo–inositol	0.1097	0.216	–0.0281	0.752	–0.1559	0.0776	0.0436	0.6254
Spermine	–0.0485	0.6405	–0.1098	0.2897	0.2407	0.0188	0.4536	< 0.001
Succinate	0.374	< 0.001	–0.3517	< 0.001	–0.2358	0.0071	–0.0454	0.6106
Taurine	–0.0895	0.3131	0.0881	0.3211	0.052	0.5586	–0.1108	0.2129
Valine	0.2057	0.0193	–0.1288	0.1457	–0.2045	0.0201	–0.182	0.0398

Ethm: Ethanolamine, GPC: Glycerophosphocholine, GPE: Glycerophosphoethanolamine, PCh: Phosphocholine, PE: Phosphoethanolamine.

Paper II

Research Paper

A novel non-canonical Wnt signature for prostate cancer aggressiveness

Elise Sandsmark[1], Ailin Falkmo Hansen[1], Kirsten M. Selnæs[1], Helena Bertilsson[2,3], Anna M. Bofin[4], Alan J. Wright[5], Trond Viset[6], Elin Richardsen[7,8], Finn Drabløs[3], Tone F. Bathen[1], May-Britt Tessem[1,*], Morten B. Rye[2,3,9,*]

[1]Department of Circulation and Medical Imaging, Faculty of Medicine, NTNU - Norwegian University of Science and Technology, Trondheim, Norway

[2]Department of Urology, St. Olavs Hospital, Trondheim University Hospital, Norway

[3]Department of Cancer Research and Molecular Medicine, Faculty of Medicine, NTNU - Norwegian University of Science and Technology, Trondheim, Norway

[4]Department of Laboratory Medicine, Children's and Women's Health, Faculty of Medicine, NTNU - Norwegian University of Science and Technology, Trondheim, Norway

[5]Cancer Research UK Cambridge Institute, University of Cambridge, United Kingdom

[6]Department of Pathology and Medical Genetics, St. Olavs Hospital, Trondheim University Hospital, Norway

[7]Department of Medical Biology, UiT - The Arctic University of Norway, Tromsø, Norway

[8]Department of Clinical Pathology, University Hospital of North Norway, Tromsø, Norway

[9]St. Olavs Hospital, Trondheim University Hospital, Norway

[*]These authors have contributed equally to this work

Correspondence to: Elise Sandsmark, *email:* elise.sandsmark@gmail.com
          Morten B. Rye, *email:* morten.rye@ntnu.no

ABSTRACT

Activation of the Canonical Wnt pathway (CWP) has been linked to advanced and metastatic prostate cancer, whereas the Wnt5a-induced non-canonical Wnt pathway (NCWP) has been associated with both good and poor prognosis. A newly discovered NCWP, Wnt5/Fzd2, has been shown to induce epithelial-to-mesenchymal transition (EMT) in cancers, but has not been investigated in prostate cancer. The aim of this study was to investigate if the CWP and NCWP, in combination with EMT, are associated with metabolic alterations, aggressive disease and biochemical recurrence in prostate cancer. An initial analysis was performed using integrated transcriptomics, *ex vivo* and *in vivo* metabolomics, and histopathology of prostatectomy samples (n=129), combined with at least five-year follow-up. This analysis detected increased activation of NCWP through Wnt5a/ Fzd2 as the most common mode of Wnt activation in prostate cancer. This activation was associated with increased expression of EMT markers and higher Gleason score. The transcriptional association between NCWP and EMT was confirmed in five other publicly available patient cohorts (1519 samples in total). A novel gene expression signature of concordant activation of NCWP and EMT (NCWP-EMT) was developed, and this signature was significantly associated with metastasis and shown to be a significant predictor of biochemical recurrence. The NCWP-EMT signature was also associated with decreased concentrations of the metabolites citrate and spermine, which have previously been linked to aggressive prostate cancer. Our results demonstrate the importance of NCWP and EMT in prostate cancer aggressiveness, suggest a novel gene signature for improved risk stratification, and give new molecular insight.

INTRODUCTION

Increased activation of the Wnt signaling pathway (WP) is associated with development, progression, and metastasis of many cancers [1]. In prostate cancer, the WP has been associated with aggressive, late stage disease, and metastasis [2–5]; however, its potential for early prediction of aggressiveness is still unclear. Previous studies are mainly performed in prostate cancer cell lines [6–9], and proper validation in human tissue is lacking. The WP is proposed as a therapeutic target in prostate cancer treatment [10], and reduced proliferation has been detected as a result of targeted Wnt-inhibitor drugs in cell lines [11, 12]. However, to develop Wnt-targeted drugs for human prostate cancer, an increased understanding of the molecular mechanisms *in vivo* is needed.

Wnt ligands bind to Frizzled (Fzd) receptors to activate the WP, which then induces signal transduction cascades. The WP is generally divided into a β-catenin-dependent canonical WP (CWP), and a β-catenin-independent non-canonical WP (NCWP). The importance of the CWP in carcinogenesis was first discovered in colorectal cancer, where mutations of the *APC* gene, a part of the β-catenin destruction complex (Figure 1A), resulted in stabilization and nuclear translocation of β-catenin [13]. This β-catenin translocation is a hallmark of CWP activation, and can drive tumor invasion and metastasis through a process of epithelial-to-mesenchymal transition (EMT) [14]. During EMT, epithelial cancer cells develop into less adhesive and more motile mesenchymal-like cells, which increases the cancer's potential for invasion and metastasis [15]. There is mounting evidence associating EMT in prostate cancer with increased aggressiveness [16]. Several studies support the activation of CWP in advanced and metastatic prostate cancer [7, 17], but little evidence exists for localized and locally advanced prostate cancer.

The NCWP is commonly divided into two pathways, the planar cell polarity (PCP), and the Wnt/Calcium pathway (Figure 1B-1C). Few studies have addressed the significance of NCWP in prostate cancer. Most attention has been focused on the role of the non-canonical ligand Wnt5a, a key activator of the NCWP. Wnt5a is generally found to be upregulated in prostate cancer, but results are inconsistent regarding its association with good [18–20] or poor prognosis [21]. Recently, a new NCWP involving Wnt5a and the receptor Frizzled2 (Fzd2) was discovered (Figure 1D) and shown to promote tumor progression and EMT in several cancer cell lines and a mouse xenograft model [22]. In the same study, a Wnt5/Fzd2 based gene set was also shown to accurately predict metastasis and survival in a small cohort (n=46) of patients with hepatocellular carcinoma. However, this study did not address the *in vivo* relevance of the NCWP in larger patient cohorts or in prostate cancer tissue.

Metabolic reprogramming is a hallmark of cancer [23], and the WP has been suggested as an emerging mediator of cancer cell metabolism [24, 25]. Wnt5a-mediated NCWP has been directly related to alterations of the energy metabolism in melanoma and breast cancer cells [26]. Selected metabolic alterations detected in tissue samples by high resolution magic angle spinning magnetic resonance spectroscopy (HR-MAS MRS) can be translated for use in a clinical setting by magnetic resonance spectroscopy imaging (MRSI). Differences in (choline + creatine + spermine)/citrate ratio between
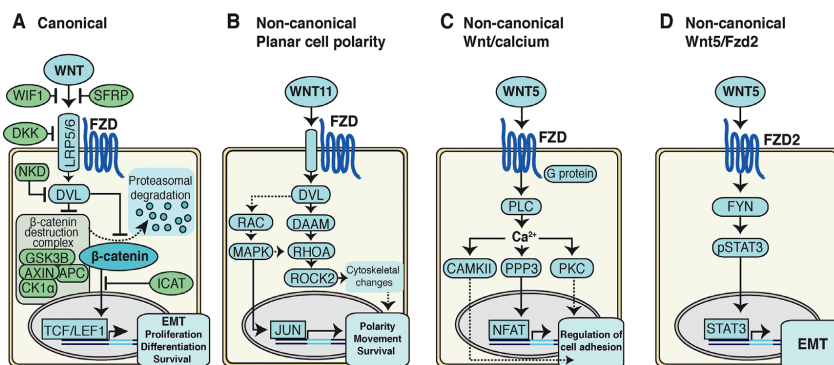
Figure 1: Schematics of Wnt signaling pathways in cancer cells. A. Canonical Wnt pathway. In the absence of Wnt signaling, the β-catenin destruction complex labels β-catenin for proteasomal degradation. In the presence of Wnt signaling, the destruction complex is inhibited, resulting in stabilization and nuclear translocation of β-catenin, activating transcription of target genes. **B.** Non-canonical planar cell polarity (PCP) pathway activates signaling cascades resulting in cytoskeletal changes, as well as alterations in cell polarity, movement and survival. **C.** Non-canonical Wnt/Calcium pathway signaling activates intracellular calcium, which in turn reduce cell adhesion through further signaling. **D.** Non-canonical Wnt5/Fzd2 pathway. Wnt5 signals via the FZD2 receptor and FYN activates STAT3 transcription leading to epithelial-mesenchymal transition (EMT) in cancer cells.

low and high histopathological Gleason score have previously been detected using *in vivo* MRSI of patients [27], and citrate and spermine are suggested as the main contributors to discriminating on the basis of tumor aggressiveness from tissue HR-MAS MRS analysis [28]. To date, metabolic alterations associated with the WP have not been investigated in prostate cancer.

The aim of this study was to investigate if the CWP and NCWP, in combination with EMT markers, are activated and associated with aggressive disease and metabolic alterations in human prostate cancer. To approach these questions, we first used a patient cohort where integrated omics analyses were performed on the same samples from fresh-frozen prostatectomy-tissue slices, including transcriptomics, tissue *ex vivo* and *in vivo* patient metabolomics, and detailed histopathological evaluation [29]. Histopathology allowed us to control for tissue heterogeneity, particularly the fraction of stroma, which is a major complicating factor when analyzing tissue samples [30]. The findings were confirmed in publicly available prostate cancer cohorts (n=1519 samples in total), and in a separate immunohistochemistry cohort. The analysis suggests that the NCWP, and not the CWP, is the most active WP for *in vivo* prostate cancer, and that this activity correlates with markers for EMT. Our approach allowed for the development of a novel NCWP-EMT gene signature significantly associated with recurrent and metastatic cancer and metabolic biomarkers. This signature may help differentiate aggressive from indolent prostate cancer.

RESULTS AND DISCUSSION

Patient and sample characteristics of the *main* and the *immunohistochemistry cohorts* are presented in Table 1. The five *validation cohorts* (in total 1519 samples) are presented in the methods section with more information listed in Supplementary Table 1.

The canonical Wnt pathway is not activated in prostate cancer

To investigate if the CWP is activated in prostate cancer, we compared gene expression of the central CWP genes between cancer and normal samples of the *main cohort* using sample subsets balanced and unbalanced for stroma content according to histopathology (Figure 2A, Supplementary Table 2, Methods). The level of β-catenin (*CTNNB1*), the key component of the CWP pathway, showed no significant altered expression in cancer compared to normal, and two of the main components of the β-catenin destruction complex, *GSK3B* and *AXIN1*, were significantly upregulated in cancer. This may suggest increased activity of β-catenin destruction in prostate cancer, contrary to what is expected when the CWP is turned on. Additionally, the Wnt ligand genes associated with the CWP were not significantly changed in cancer compared to normal samples. Other important findings

are reduced expression of the receptor *FZD1*, increased expressions of the antagonist *SFRP4* and casein kinase *CSNK1E*, which support the absence of CWP activation. Although some variations were observed (Figure 2A), the lack of upregulation of the main CWP genes suggests no increased expression activity of the CWP in prostate cancer in our *main cohort*.

Translocation of β-catenin from the membrane to the nucleus is the hallmark of CWP activation, and to validate the findings above, we performed β-catenin immunohistochemistry (IHC) on the *immunohistochemistry cohort* (Figure 3A-3B). All the samples (n=40) had weak or non-detectable nuclear staining (SI≤2). Most of the samples (n=30) had strong membranous β-catenin staining (SI=9), indicating no activation of the CWP. Ten samples had weak or moderate membranous staining (SI≤6), indicating reduced membranous expression without increased nuclear expression of β-catenin. These findings demonstrate that the CWP is not activated in prostate cancer in our *immunohistochemistry cohort*, which is in concordance with the gene expression results from the *main cohort*. We therefore conclude that there is little evidence of CWP activation in prostate cancer compared to normal prostate tissue investigated in two independent cohorts.

We further investigated alterations in the CWP between *low Gleason (≤3+4)* and *high Gleason (≥4+3)* samples (Figure 2A). There were no significant gene expression alterations detected for β-catenin (*CTNNB1*), the Wnt ligands, the receptor-complex and the destruction complex (Supplementary Table 2). Of the CWP inhibitors, both *SFRP2* and *SFRP4* were upregulated in *high Gleason* compared to *low Gleason* cancer samples, which is contradictory to CWP activation. However, the inhibitor of β-catenin translocation, *ICAT* (*CTNNBIP1*), was downregulated, and the CWP transcription factors *LEF1* and *TCF* were upregulated in *high Gleason* cancer, which could indicate activation of downstream components of the pathway independently of the β-catenin destruction complex. To conclude, the overall analysis suggests no significant increase in CWP activation through the canonical destruction complex, neither in cancer compared to normal nor in *high Gleason* cancer.

There is currently no consensus in the literature regarding CWP activation in prostate cancer, and our findings are contradictory to several previous studies suggesting increased CWP in prostate cancer [7, 9, 17]. The CWP has previously been associated with advanced disease such as androgen resistant prostate cancer in cell lines [7], and prostate cancer bone metastasis in human tissue and cell lines [8, 17]. The fact that our cohorts consist of radical prostatectomy tissue, from localized or locally advanced disease, may explain the absence of CWP activation. The CWP may therefore still be of importance in advanced, metastatic prostate cancer, but might not prove useful for early risk stratification. Furthermore, several previous studies reporting increased CWP signaling

Table 1: Patients and sample characteristics of the two cohorts

		Main cohort	Immunohistochemistry cohort
Patients		**n=41**	**n=40**
Age (median, range)	Years	64 (48-69)	62 (48-73)
sPSA (median, range)	Before Surgery (ng/mL)	9.1 (4.0-45.8)	8.9 (5.2-18.0)
Clinical pT stage (patients)	pT1c	-	7
	pT2	28	20
	pT3	13	10
	Unknown	-	3
Tissue samples		**n=129**	**n=40**
Sample weight (mean, range)	(mg)	12.7 (3.0-21.9)	12.6 (7.6-21.0)
Gleason score of tissue samples	Benign	34	- *
	6	24	5
	7	41	25
	8	15	5
	9	15	4
	10	-	1
Gleason grade groups	*Low Gleason (≤3+4)*	48	21
	High Gleason (≥4+3)	47	19

* 50 benign samples were excluded from further analysis in the *immunohistochemistry cohort*
sPSA – serum PSA, pT stage – pathological tumor stage.

were using prostate cancer cell lines [6–9]. The disparity could therefore also highlight a difference between *in vitro* cell lines and human prostate tissue, emphasizing the importance of validation studies in human tissue, especially for identification of potential targets for personalized drug therapy.

In our *main cohort*, the central CWP genes showed an expression pattern that was indicative of substantial stromal influence when comparing normal against cancer tissue (Figure 2A). This trend was particularly strong for genes that, directly or partly, regulate the activity of the β-catenin destruction complex, and indicates a difference of CWP activity when cancer is compared to stroma, but not when compared to benign epithelium. Thus, at least some of the discrepancies from previous studies of CWP in prostate cancer may be explained by uneven sampling of stroma content between cancer and normal samples which has previously been observed in tissue samples from prostate cancer patient cohorts [30, 31].

Wnt5a-induced non-canonical Wnt signaling is increased in *high Gleason* prostate cancer

The NCWP, including the Wnt/Calcium, PCP and the new Wnt5/Fzd2 pathways, were investigated (Figure 2B, Supplementary Table 2). When comparing cancer with normal samples, we found no alterations in any of

the pathway components apart from downregulation of the ligand *WNT5B*, and upregulation of the calcium pathway component *PLCB2*, suggesting no increased activation of the NCWP in prostate cancer in general. However, when *high Gleason* samples were compared with *low Gleason* samples, significantly increased expressions were detected for three of the four key genes of the Wnt5/Fzd2 pathway; the ligand *WNT5A* (p<0.001), the receptor *FZD2* (p=0.003) and the midstream kinase component *FYN* (p<0.001) (Figure 2B). No significant expression change was detected for the last key component, the transcription factor *STAT3*. For the Wnt/Calcium pathway, only *PLCB2* was upregulated in *high Gleason* cancer (Figure 2B), and none of the central components of the PCP pathway were altered (Supplementary Table 2). In summary, these data suggest upregulation of the Wnt5/Fzd2 pathway in *high Gleason* prostate cancer.

For validation, IHC of WNT5A was performed on the *immunohistochemistry cohort* (Figure 3C-3D). Of the 40 cancer samples, 32 had strong (SI=9) and 8 had moderate or weak staining (SI≤6). There was no association between the staining intensity and Gleason grade for this cohort.

Wnt5a has been suggested as a biomarker in prostate cancer, but its prognostic outcome has been inconsistent [18–21]. The increased *WNT5A* gene expression in *high Gleason* cancer samples compared to *low Gleason* samples
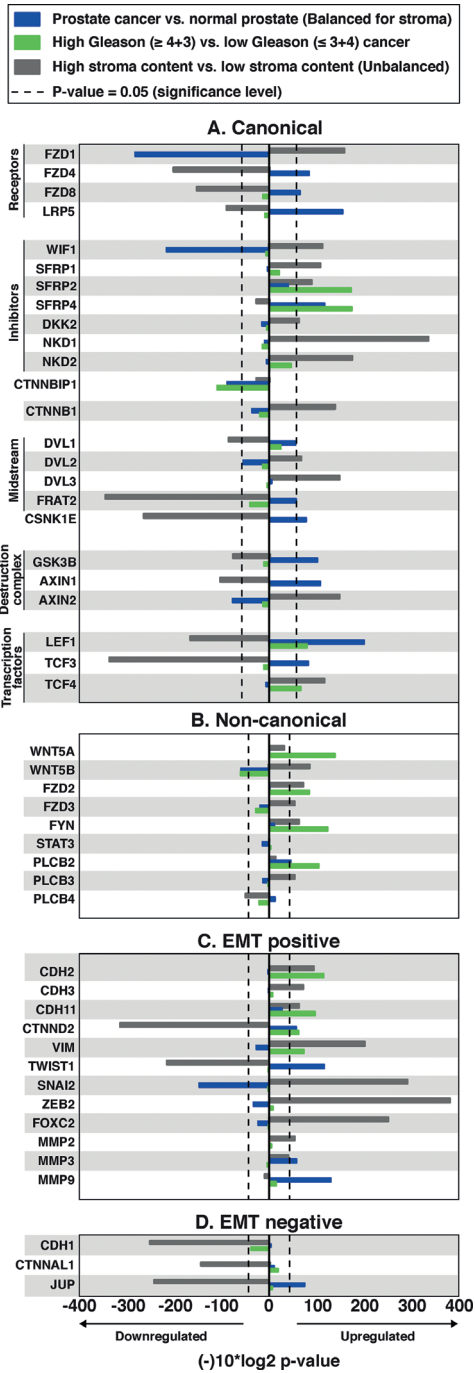
Figure 2: Alterations in central Wnt and EMT genes in prostate cancer compared with normal samples (balanced for stroma), *high Gleason* **compared with** *low Gleason* **prostate cancer, and high stroma content compared with low stroma content (unbalanced) tissue samples.** The x-axis displays log10(p-value) fold change, multiplied by −1 for upregulated genes, and 1 for downregulated genes. P-values for prostate cancer vs. normal prostate tissue are balanced for stroma content; unbalanced p-values are available in Supplementary Table 2. **A.** The central canonical genes show a pattern of no further activation in cancer or *high Gleason* cancer, but show a confounding stroma effect, especially of the genes of the destruction complex. **B.** The central non-canonical genes generally show an upregulation of Wnt5/Fzd2 genes in *high Gleason* cancer. **C.** The central epithelial-mesenchymal transition (EMT) positive genes indicate ongoing EMT, especially in *high Gleason* cancer. **D.** The central EMT negative genes.

is in agreement with results from Yamamoto et al. who reported increased Wnt5a IHC staining of prostatectomy tissue samples with high Gleason grade [21]. This oncogenic effect of Wnt5a in prostate cancer progression is also supported by studies of cell lines, where Wnt5a has been shown to improve migration capacity [32], induce androgen resistance in prostate cancer metastases [33], and induce bone metastasis [8]. Contrary to this, other IHC studies of prostatectomy tissue samples have detected a tumor-suppressing role of Wnt5a in prostate cancer; increased Wnt5a IHC expression has been associated with increased

10 years survival [18], and a lower risk of biochemical recurrence [19, 20]. This was, however, only true for low Gleason grade samples in one of the studies [20]. This apparent opposing role of Wnt5a in prostate cancer may be explained by the paradoxical effect of Wnt5a in other cancers. In melanoma, pancreatic and gastric cancer, Wnt5a expression is associated with worse prognosis, but in colon and thyroid cancer Wnt5a expression is associated with better prognosis as reviewed by McDonald and Silver, and Zhu et al. [34, 35]. The tumor-promoting role of Wnt5a can be caused by activation of NCWP [35], whereas the tumor-suppressing role may be caused by inhibition of the CWP [36]. Because of this conflicting role in different cancer types, we suspect that Wnt5a alone may not be a useful biomarker for prostate cancer.

EMT markers are upregulated in *high Gleason* prostate cancer

The Wnt5/Fzd2 NCWP has previously been linked with EMT studies on various cancer cell-lines, but not in prostate cancer [22]. We therefore evaluated the gene expression of the most central EMT positive and negative markers in prostate cancer in the *main cohort* (Figure 2C and 2D). When comparing *high Gleason* with *low Gleason* samples, significant upregulations were detected for the expression of EMT positive markers in *high Gleason*; N-cadherin (*CDH2*), OB-cadherin (*CDH11*), vimentin (*VIM*) and Delta-2-catenin (*CTNND2*) (Figure 2C). In addition, a non-significant downregulation of E-cadherin (*CDH1*), an EMT negative marker, was observed in *high Gleason* samples (fold-change=-0.25, p=0.07; Figure 2D), suggesting ongoing EMT in *high Gleason* samples. In the *immunohistochemistry cohort*, IHC of N-cadherin showed membranous staining (SI≥2) in only two, both *high Gleason*, of the forty cancer samples (Figure 3E-3F). Reduced, moderate membranous staining of E-cadherin (SI=6), was detected in five samples while the remaining samples had strong membranous staining (SI=9) (Figure 3G-3H). However, the reduced E-cadherin staining did not correspond to N-cadherin staining, as hypothesized for the N- to E-cadherin switch proposed to be important for EMT in prostate cancer [37]. Inspection of the principal component analysis (PCA) score plots for the *main* and *validation* cohorts also confirmed consistent N-cadherin upregulation correlating with *high Gleason* and EMT genes, while the anticorrelation to E-cadherin was inconsistent between the cohorts, in accordance with observations in the *immunohistochemistry cohort* (Figure 4A, 4C-4G). In conclusion, the increased levels of several EMT positive genes, suggests ongoing EMT in a subset of mainly *high Gleason* prostate cancer samples. This was partly supported by the IHC, although the number of samples in the *immunohistochemistry cohort* was too few to make a conclusion.

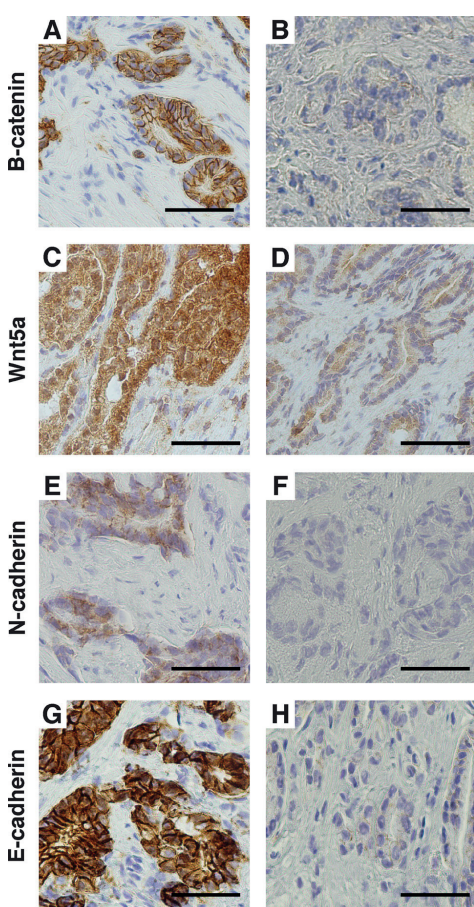**Figure 3: Immunohistochemical staining of the immunohistochemistry cohort. A.** Strong membranous β-catenin staining and **B.** weak β-catenin staining. **C.** Strong Wnt5a staining and **D.** weak Wnt5a staining. **E.** Positive membranous N-cadherin staining and **F.** negative N-cadherin staining. **G.** Strong membranous E-cadherin staining and **H.** weak E-cadherin staining. Magnification x400. Bar 50µm.

A novel 15 gene non-canonical Wnt pathway - EMT (NCWP-EMT) signature

To further investigate the relationship between the expression of Wnt and EMT genes, PCA analysis was performed on the expression profiles of 48 central Wnt and EMT genes (Methods). The first two principal components clearly highlighted a separate cluster of 15 genes related to the Wnt5a/Fzd2 pathway and EMT (Figure 4A). This gene set included 11 genes, which were also upregulated in *high Gleason* samples. In addition, two inhibitors of the CWP (*NKD2* and *SFRP1*), and two EMT positive markers (*CDH3* and *MMP9*) were part of the PCA cluster and included in the gene set. Because of the clear relationship to Wnt5/Fzd2 NCWP and EMT, we will refer to this set of genes collectively as the NCWP-EMT genes.

Using all cancer samples in the *main cohort*, we calculated an average Pearson's correlation r of 0.34 between all 15 gene using pairwise correlations. This is comparable or higher than the average correlation between genes in previously validated prostate cancer signatures [38, 39] (Figure 4B), including signatures for the established TMPRSS2-ERG gene fusion (average Pearson's r=0.30). The pattern of the NCWP-EMT gene set from the *main cohort* was validated in PCA analysis of the Wnt-genes in the five publicly available cohorts (n=1519 samples in total, Supplementary Table 1). The same 48 central Wnt-genes, in addition to *WNT1*, *WNT3* and *WNT3A* which were lacking data in the *main cohort,* were used. All cohorts confirm the NCWP-EMT component as the most important source of variation in the gene expression, although there were some variations in the highlighted genes (Figure 4C-4G and Supplementary Figure 1). The CWP was either insignificant or spanning a separate axis of variation with little correlation to EMT. Interestingly, *WNT5A* expression pattern varied considerably with respect to the NCWP-EMT axis. Overall, these data show the NCWP-EMT gene cluster to be robust over large prostate cancer patient cohorts, and the 15 NCWP-EMT genes to be accessible for a concordant NCWP-EMT gene expression signature.

The continuous single sample gene set enrichment analysis (GSEA) score of the novel NCWP-EMT signature was significantly correlated with the Gleason score of the samples (Pearson's r of 0.49, p<0.001). When the samples were categorized according to the NCWP-EMT score as *low*, *intermediate*, and *high*, the distribution of *low/high Gleason* samples in the groups were as following: NCWP-EMT *low* (n=25/n=7), NCWP-EMT *intermediate* (n=17/n=14), and NCWP-EMT *high* (n=6/n=26). As expected most samples with *high* NCWP-EMT score also were *high Gleason* samples; however, some samples were *low Gleason*, and vice versa for samples with *low* NCWP-EMT score. This indicates that the NCWP-EMT signature might add an additional dimension for stratification, compared to Gleason grade alone. The NCWP-EMT signature may therefore, with further refinements and validation, be a useful addition to the selection criteria for active surveillance in prostate cancer patients.

The novel NCWP-EMT signature also showed significant association with previously published mesenchyme and cytokine gene signatures (Supplementary Figure 2), and highly significant gene ontology (GO) terms related to cell adhesion, extracellular matrix, inflammation and immune response which are features commonly associated with EMT (Supplementary Table 3). The same analysis based on the expression level of *WNT5A* alone, did not produce any significant GO terms, further supporting the hypothesis that Wnt5a alone is an ambiguous biomarker in prostate cancer.

The NCWP-EMT gene signature is associated with metabolic alterations

We further investigated the metabolic alterations of 23 metabolites between samples with *low*, *intermediate*, and *high* activation of the developed NCWP-EMT gene expression signature (Supplementary Table 4) in the *main cohort*. The most prominent alterations were observed for the metabolites citrate and the polyamine spermine (Table 2), which showed significantly decreased concentration in the *high* NCWP-EMT compared to *low* NCWP-EMT samples. This alteration was also observed for *high* NCWP-EMT samples when compared with *intermediate* NCWP-EMT samples, but not when comparing *intermediate* with *low* NCWP-EMT samples. This suggests citrate and spermine alterations to be more profound in the samples with *high* NCWP-EMT score compared to *low* and *intermediate* score NCWP-EMT. In addition, there were alterations in the concentration of phosphoethanolamine and taurine between the *low* and the *intermediate score* group (p=0.002, p=0.028 respectively).

Decreased concentrations of citrate and spermine have been associated with aggressive prostate cancer [28, 40], and our results therefore suggest the NCWP-EMT signature to be associated with an aggressive metabolic profile. Reduced citrate can be a result of increased energy production through the Krebs cycle in prostate cancer [41]. Previously, Wnt5a signaling has been identified as a regulator of the energy metabolism in melanoma cancer cells [26], and alterations of this metabolism have also been associated with EMT in cancer [42]. Another study detected that reduced polyamine content promoted EMT in non-tumor MDCK cells [43]. We therefore hypothesize that NCWP-EMT activation is associated with alterations in citrate and spermine metabolism in prostate cancer, although the direct mechanisms require further investigation.

To investigate the potential clinical translation of the metabolic findings, we inspected the gene signature score with matched pre-surgical *in vivo* MRSI from the same patients. Reduced citrate/creatine and spermine/creatine ratios were detected for *high* NCWP-EMT score samples when compared with *low* NCWP-EMT score (Table 2). Although we had a limited number of matched samples
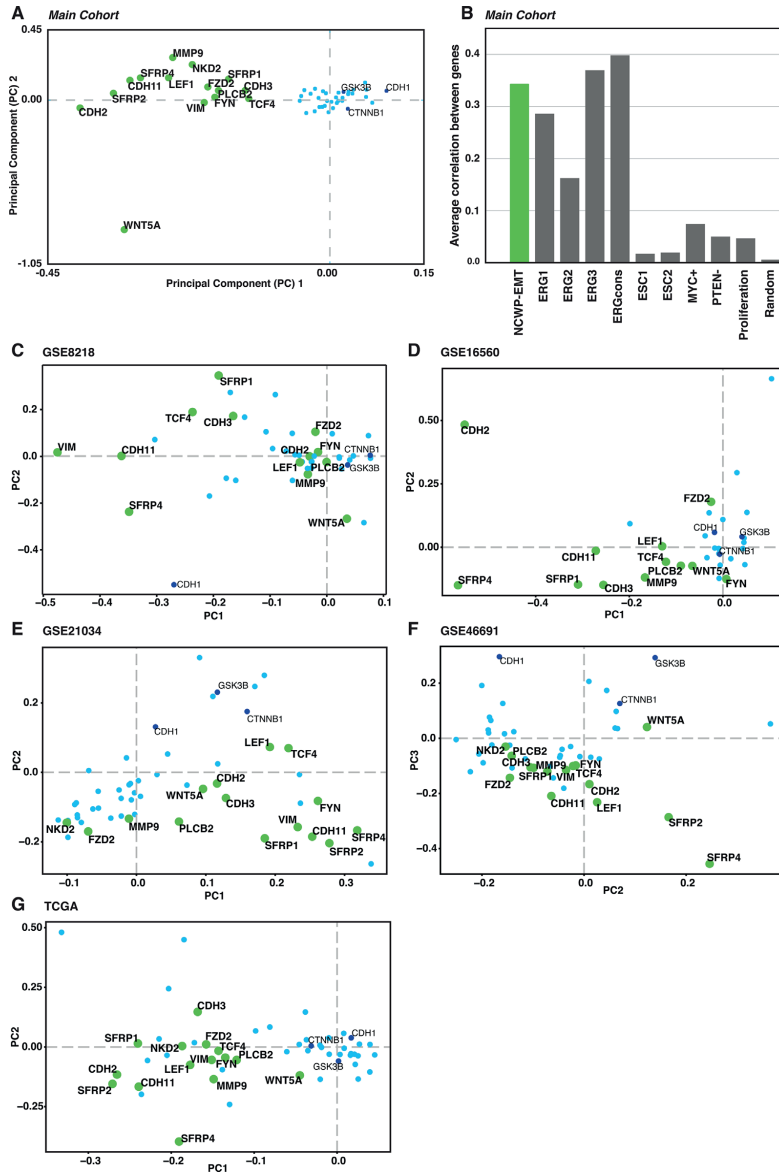
Figure 4: The NCWP-EMT gene expression signature. A. Two component PCA plot reveals a group of 15 of 48 genes, mainly connected to Wnt5a/Fzd2 non-canonical Wnt pathway, epithelial-mesenchymal transition (EMT), and inhibitors of the canonical Wnt pathway, collectively termed NCWP-EMT (*CDH2, CDH3, CDH11, FYN, FZD2, LEF1, MMP9, NKD2, PLCB2, SFRP1, SFRP2, SFRP4, VIM, TCF4 WNT5A*). **B.** The Pearson correlation of co-expression of the genes in the NCWP-EMT signature is as good or better compared with other recognized genes expression signatures in prostate cancer. Random marks 200 randomly selected genes for validation. **C-G.** The NCWP-EMT signature confirmed in the *validation cohorts*, although there were some variations in the highlighted genes. High-resolution versions of the PCA plots including all gene names, and Pearson correlation of the *validation cohorts* are available in Supplementary Figure 1.

Table 2: Alterations in citrate and spermine metabolism

Signature score	Metabolite concentration (mmol/kg wet weight) *ex vivo* and metabolites amount/ratios *in vivo*			p-values[a]		
	Low	*Intermediate(Int)*	*High*	*Low* vs. *High*	*Int.* vs *High*	*Low* vs. *Int.*
	Median (IQR)	Median (IQR)	Median (IQR)			
Ex vivo (n=95)	(n=32)	(n=31)	(n=32)			
Citrate	7.31 (5.57-11.56)	6.38 (4.56-11.58)	3.55 (2.08-7.25)	$3.38 \cdot 10^{-4}$*	0.018*	0.282
Spermine	1.55 (1.02-2.36)	1.23 (0.67-2.27)	0.75 (0.39-1.43)	$3.38 \cdot 10^{-4}$*	0.028*	0.113
In vivo (n=22)	(n=10)	(n=7)	(n=5)			
Citrate/Creatine	7.36 (5.81-8.79)	4.45 (3.34-7.79)	2.77 (1.48-3.00)	0.0056*	0.027*	0.030*
Spermine/ Creatine	0.83 (0.44-1.04)	0.50 (0.04-1.11)	0.00 (0.00-0.02)	0.0057*	0.027*	0.101

IQR – Interquartile range
* Indicates significance at p<0.05
[a] P-values from LMM adjusting for multiple samples per patient, and corrected for multiple testing by Benjamini and Hochberg procedure.

in the *main cohort* (n=22), the results support our findings from the tissue analysis, and demonstrates that the MR biomarkers can reflect the NCWP-EMT signature also in non-invasive MRSI examinations.

Citrate and spermine are stored within the luminal space of the glands in prostate tissue, and the observed metabolic alterations can be due to cell metabolism or morphological changes. In the *main cohort*, the citrate and spermine concentrations were correlated with luminal space (Spearman's rho=0.30/p=0.003, rho=0.31/p=0.003, respectively). This was a weaker correlation than between citrate and spermine concentrations and the NCWP-EMT signature score (Spearman's rho=0.42/p<0.001, rho=0.38/p<0.001, respectively). LMM, adjusting for luminal space as well as other tissue heterogeneity and Gleason score, still showed the same metabolic alterations to be significant (Supplementary Table 5). These results suggest the alterations observed in citrate and spermine concentrations are a combination of changes in both luminal space and reprogramming of metabolism in samples with *high* NCWP-EMT score. There was no relationship between Wnt5a expression and metabolite concentrations in either the *main* or *immunohistochemistry cohort* (Supplementary Table 6). This supports that Wnt5a should be used as a biomarker in combination with other pathway components, such as our NCWP-EMT signature.

NCWP-EMT signature may help predict biochemical recurrence

In the *main cohort* the five-year biochemical recurrence free rates were 100%, 75% and 46% for the patients in the *low*, *intermediate* and *high* NCWP-EMT score groups, respectively, and the Kaplan-Meier plot showed a significant separation between the groups (log-rank p=0.035) (Figure 5A). Validation of recurrence was possible in the GSE21034 cohort (131 samples, 27 with recurrence), and showed a similar pattern with 10-year biochemical recurrence free rates of 81%, 73% and 57% in patients with *low, intermediate* and *high* NCWP-EMT score, respectively. However, there was no significant separation in the Kaplan-Meier curves for this cohort (log-rank p=0.522) (Figure 5B). For this validation dataset there was only one sample per patient, not necessarily extracted from the most aggressive cancer foci, which may reduce the precision of the NCWP-EMT grouping for biochemical recurrence analysis. In addition, many of the patients in the validation dataset were lost to follow-up early, and therefore censored in the analysis (Figure 5B), causing reduces reliability of the curves. In the GSE46691 cohort, samples with *high* NCWP-EMT scores were significantly associated with metastases (545 samples, 212 with metastasis, p-value<0.001, chi-square test, Supplementary Figure 3). With the significant separation in our data, and the similar trend in the validation datasets, we therefore suggest that increased NCWP-EMT signature score is associated with an increased risk of biochemical recurrence and metastases. This strengthens the NCWP-EMT signature, and the activation of the Wnt5/Fzd2 pathway, as markers of aggressive prostate cancer.

Patients in the *main cohort* with a post-operative Gleason score of 7 showed a five-year biochemical recurrence free survival of 100%, 89% and 67% with *low*, *intermediate* and *high* NCWP-EMT score, respectively (Figure 5C). Although not statistically significant, possibly due to the low number of patients (n=23), this separation with no crossing indicates that the NCWP-EMT gene

signature might be useful for improved risk stratification in the challenging group of patients with Gleason score 7.

Univariate cox proportional hazards analyses identified NCWP-EMT, Gleason score and pathological T-stage as significant predictors of biochemical recurrence (Table 3). Multivariate analysis showed both NCWP-EMT and post-operative Gleason score to be significant predictors of biochemical recurrence (Table 3). The multivariate model included a significant interaction term between NCWP-EMT and post-operative Gleason score, implying that the hazards ratio of these variables were dependent on the value of the other variable. For patient with low post-operative

Gleason score ($\leq$ 7), the hazard ratio for NCWP-EMT was 1.61, indicating that increased NCWP-EMT signature score gives a significant higher risk of biochemical recurrence for this group. To compare the NCWP-EMT and post-operative Gleason score as predictors of biochemical recurrence, two additional Cox proportional hazards models, each excluding either NCWP-EMT or post-operative Gleason score, were tested (Supplementary Table 8). The Akaike information criterion (AIC) represent the goodness of fit as well as the complexity of the model, and can be compared between models, where the lower AIC provides a better model fit. The model including post-operative Gleason
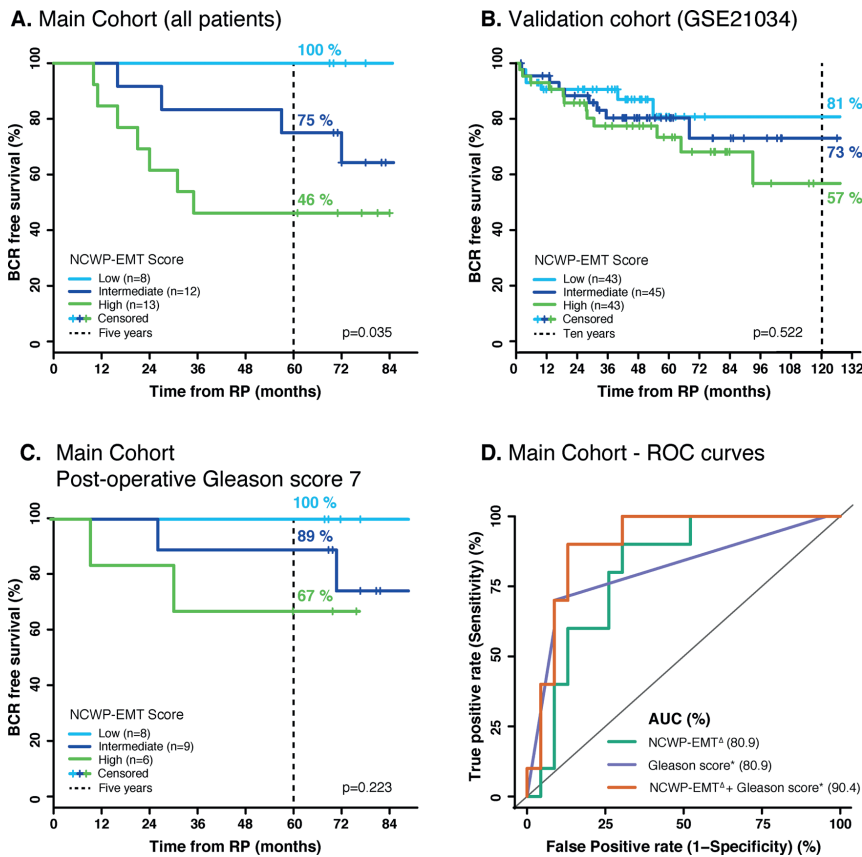
Figure 5: Kaplan-Meier and ROC curves of biochemical recurrence. A. The *main cohort* shows clear separation in biochemical recurrence free survival between the *low*, *intermediate* and *high* NCWP-EMT signature groups. **B.** A validation cohort (GSE21034) shows the same pattern, although not a significant separation. **C.** A similar pattern was also shown for the patient of the *main cohort* with a post-operative Gleason score of 7. **D.** The ROC curves of biochemical recurrence after 5 years show the same AUC of post-operative Gleason score and NCWP-EMT, but an increased AUC when combined. $^\Delta$ Continuous NCWP-EMT signature score, * continuous post-operative Gleason score. Abbreviations: BCR - biochemical recurrence, RP – radical prostatectomy, ROC – Receiver operating characteristic, and AUC – area under the curve.

Table 3: Univariate and multivariate Cox proportional hazards analyses of biochemical recurrence

Variables	Univariate		Multivariate – All variables (AIC = 60.15)	
	Hazard ratio (95% CI)	P-values	Hazard ratio (95% CI)	P-values
Post-operative Gleason score (≤7^Δ and ≥8)	7.66 (2.20-26.62)	0.001*	19.46 (2.67-142.9)	0.003*
Pathological T-stage (≤T2c^Δ and ≥T3a)	6.88 (2.06-23.01)	0.002*	8.27 (0.89-77.15)	0.064
Pre-operative PSA (<10^Δ and ≥10)	2.17 (0.69-7.13)	0.204	2.89 (0.72-11.67)	0.14
NCWP-EMT Continuous score/100 (-4.4–5.4)	1.37 (1.08-1.73)	0.009*	Low GS 1.61 (1.06-2.44)	Low GS 0.028*
			High GS 0.59 (0.35-0.99)	High GS 0.044*
NCWP-EMT and Post-operative Gleason score (≤7^Δ and ≥8) (interaction term)	-	-	0.37 (0.18-0.74)	0.005*

^{Δ} Indicates the category used as a reference in each analysis.

* Indicates significant p-value.

Low GS – Hazard ratio/p-value for patients with post-operative Gleason score ≤7

High GS – Hazard ratio/p-value for patients with post-operative Gleason score ≥8

had a slightly lower AIC (AIC=64.24) compared to the model including NCWP-EMT (AIC=65.61), suggesting post-operative Gleason to be a slightly better predictor of biochemical recurrence than NCWP-EMT. However, the model containing all variables, had the lowest AIC (AIC=60.15) demonstrating improved prediction of biochemical recurrence when NCWP-EMT and post-operative Gleason score were modelled together.

Similar findings were also visualized by using logistic regression and receiver operating characteristic (ROC) curves with the depended variable being biochemical recurrence after 5-year follow-up. The area under the curve (AUC) of the ROC-curve were the same for NCWP-EMT and post-operative Gleason score (AUC=80.9), and in combination they provided increased sensitivity and specificity (AUC=90.4) (Figure 5D). In conclusion, our results suggest that the NCWP-EMT signature could be a useful addition in prediction of biochemical recurrence in prostate cancer.

CONCLUSIONS

The present study showed no alterations in the CWP in prostate cancer, but revealed an increased expression of NCWP and EMT markers in a subgroup of mainly *high Gleason* grade prostate cancer samples. A novel gene expression signature (NCWP-EMT) for this expression profile was presented and confirmed in several publicly available patient cohorts. *High* NCWP-EMT score was

associated with reduced concentrations of the metabolites citrate and spermine both *ex vivo*, and in a clinical non-invasive setting using *in vivo* patient MRSI. The novel NCWP-EMT signature was also shown to be a predictor of biochemical recurrence and was associated with metastasis, indicating that upregulation of the NCWP and EMT is linked to more aggressive prostate cancer. The novel NCWP-EMT signature may therefore be useful for risk stratification and molecular subtyping of prostate cancer patients. The NCWP and its relation to EMT, cancer aggressiveness and tumor metabolism warrants further attention in prostate cancer studies.

MATERIALS AND METHODS

Patients and tissue samples

In the *main cohort*, human prostate tissue was collected from 41 localized and locally advanced prostate cancer patients. The tissue harvesting was performed on fresh-frozen prostatectomy specimens using a standardized method thoroughly described by Bertilsson et al. [29]. A total of 95 cancer tissue samples, and 34 adjacent normal tissue samples were collected (median 3, range 1-6 samples per patient). At least five years of follow-up data were successfully retrieved for 33 patients in the *main cohort*, including the date of biochemical recurrence (PSA of at least 0.2 ng/mL) and/or last negative PSA measurement. To validate the results of the *main cohort*, an additional

cohort of 90 needle biopsies from 90 localized and locally advanced cancer patients were harvested and snap frozen within seconds after prostatectomy. Of these, only the samples with histopathological confirmed cancer were used as the *immunohistochemistry cohort* for this study (n=40). The patients in both cohorts received no prostate cancer treatment prior to surgery and had no detected metastasis at diagnosis. The Regional Committee of Medical and Health Research Ethics (REC), Central Norway approve both cohorts, and all patients gave written, informed consent. Validation was performed in four prostate cancer microarray datasets available through the Gene Expression Omnibus with GEO accessions GSE8218 (65 samples) [44], GSE16560 (281 samples) [45], GSE21034 (131 samples) [46], GSE46691 (545 samples) [47], and one data set from The Cancer Genome Atlas (TCGA, 497 samples) [48], in total 1519 samples (Supplementary Table 1). These datasets are collectively termed the *validation cohorts*. Biochemical recurrence was validated in the GSE21034 cohort, and metastasis in the GSE46691 cohort.

Histopathology

In the *main cohort*, tissue slices for histopathological evaluation were cryosectioned from each tissue sample prior to HR-MAS MRS [29]. All cryosections were stained with Haematoxylin and Eosin, and the histopathological evaluations were performed according to the clinical criteria for prostate cancer, by an experienced pathologist specialized in uropathology (TV). The percentage of Gleason grades, cancer, normal glandular epithelia, and stromal tissue were reported for each sample. Reproducibility of the histopathological scoring was assessed independently by a second pathologist specialized in uropathology (ER), and the overall kappa (κ) coefficient for interobserver agreement of Gleason score was 0.66 indicating substantial agreement. The first reading was used in this study due to slight degradation of the cryosections between the readings (5 years, slides kept dry and dark). Luminal space was quantified in each sample by a color-based segmentation method (Positive Pixel Count algorithm in ImageScope v.8, Aperio Technologies) [49]. The samples in the *immunohistochemistry cohort* were formalin fixed and paraffin embedded for sectioning after HR-MAS MRS analysis, and histopathological evaluation was done according to the same protocol as the *main cohort*. In both cohorts, we investigated differences between low and high Gleason grade by sorting the tissue samples into two groups, where samples in the *low Gleason* group had a Gleason score ≤ 3+4 and samples in the *high Gleason* group had a Gleason score ≥ 4+3 (Table 1).

HR-MAS MRS and MRSI experiments and quantification

For both the *main* and the *immunohistochemistry* cohort, proton HR-MAS MRS was acquired using a

Bruker Avance DRX600 Spectrometer (Bruker Biospin, Germany) equipped with a dual $^1H/^{13}C$ MAS probe. Absolute quantification of the spectra was performed using LCModel [50] with a basis set of 23 metabolites, and reported in mmol/kg wet weight. Full procedure and parameters of the HR-MAS MRS acquisition and LCModel quantification have earlier been described by Giskeødegård et al. [28]. *In vivo* patient MRSI examination of the prostate, performed using a 3T system (Magnetom Trio, Siemens, Germany) prior to prostatectomy, was available on a subset of the patients in the *main cohort* (n=9). Choline, citrate, creatine and spermine were quantified using LCModel, and creatine was used as an internal standard for normalization (metabolites to creatine ratios). HR-MAS cancer samples from the same patients were spatially matched to an *in vivo* voxel (n=22). Further details on the MRSI acquisition, quantification, and spatial matching are previously described by Selnæs et al. [27].

Gene expression, selection of genes, and controlling for confounding stroma

In the *main cohort*, gene expression analysis was performed after HR-MAS MRS on the exact same tissue sample, using an Illumina TotalPrep RNA Amplification Kit (Ambion Inc.) and an Illumina Human HT-12v4 Expression Bead Chip (Illumina), as described by Bertilsson et al. [51]. The microarray data has previously been published in Array Expression with access number: E-MTAB-1041. Genes relevant to both the WP and EMT were carefully chosen by investigating literature and publicly available pathway maps (KEGG as per March 2015) [2, 3, 5, 22]], resulting in 196 genes (Supplementary Table 2). To control for the effect of confounding stroma tissue when identifying differentially expressed genes, we used a recently published strategy of balancing the stroma content between sample groups [30]. This strategy makes it possible to separate molecular signals relevant to cancer from signals originating due to different stroma fractions between the sample groups. Briefly described, the strategy selects samples to ensure an equal average fraction of stroma tissue (according to histopathology) in each sample group termed a *balanced* differential expression analysis. In contrast, an *unbalanced* analysis is also performed to highlight differentially expressed gene due to different average fractions of stroma tissue.

Immunohistochemistry (IHC)

In the *immunohistochemistry cohort*, IHC was performed with mouse monoclonal antibodies against Wnt5a (Sigma-Aldrich, clone 3A4, dilution 1:50), N-cadherin (Dako, clone 6G11, dilution 1:30), and E-cadherin/NCH-38 (Dako, clone NCH-38, dilution 1:100) and polyclonal rabbit antibodies against β-catenin/*CTNNB1* (PRESTIGE antibodies Sigma, dilution 1:300). The sections were counter-stained with

Haematoxylin. Assessment was performed manually, and all the IHC sections were evaluated based on the average staining intensity (0-3) multiplied by the percentage of positive cancer cells (0-3), obtaining a total staining index (SI) (0-9). A SI of 0 was regarded as negative, 1-2 as weak positive staining; 3-6 as moderate, and 9 as strong positive staining (Supplementary Table 7). An experienced pathologist (AMB) validated the scoring.

Statistical analysis

The WP and EMT genes were compared for differential expression between normal and cancer samples, and between *low* and *high Gleason* samples by t-test. All the 196 genes were considered, but to ease data analysis and presentation a subgroup 48 key and/or significantly altered genes are presented as the central genes, however, a full table of the p-values is given in Supplementary Table 2. PCA was used to further investigate and visualize the unsupervised relationship between the expressions of these central WP and EMT genes. Based on the PCA score plot, a distinct set of genes was selected to make a gene expression signature termed NCWP-EMT. The co-expression between the signature genes was investigated by Pearson's correlation, and compared to other recognized gene expression signatures in prostate cancer. The distinct gene-signature pattern from PCA and Pearson's correlation between signature genes were confirmed in the *validation cohorts*. Single sample GSEA was performed to give each of the cancer samples in the *main* and *validation cohorts* a score representing the expression of the genes in the NCWP-EMT signature [52]. The samples in each cohort were sorted into three equal sized groups of *low, intermediate*, and *high* NCWP-EMT signature scores, where the *high score* group had the highest pathway activity. Features associated with NCWP-EMT were investigated by Gene Ontology (GO) using the Database for Annotation and Visualization and Integrated Discovery (DAVID). Biochemical recurrence free survival for the NCWP-EMT score groups were plotted by Kaplan-Meier curves and tested by log-rank test in the *main* and GSE21034 cohort, where for the individual patient's highest NCWP-EMT score was used in the *main cohort*. The association between NCWP-EMT and metastasis in the GSE46691 cohort was tested using a contingency table and chi-squared test. Univariate and multivariate cox proportional hazards statistics were used to investigate the role of the NCWP-EMT signature in prediction of biochemical recurrence. Prior to analysis, post-operative Gleason score, pathological T-stage and pre-operative PSA were dichotomized (Table 3), and together with the continuous NCWP-EMT signature score selected for multivariate analysis. Biochemical recurrence at five-year follow-up was selected to plot ROC curves of NCWP-EMT score, post-operative Gleason score and both combined. Linear mixed model (LMM) was used to account for multiple samples per patient, when investigating the relationship between NCWP-EMT score groups and metabolite concentrations. The analyses were repeated with additional adjustment for Gleason grade, and tissue heterogeneity including the proportion of cancer, benign epithelium, stroma and luminal space in the individual tissue sample. The *immunohistochemistry cohort* consisted of one sample per patient, and t-test was used to investigate the association between IHC and metabolite concentrations. Prior to analysis, all metabolite values were log transformed to obtain normalized residuals, and p-values were corrected for multiple testing using Benjamini-Hochberg false discovery rate. P-values <0.05 were considered significant. The statistical analyses were performed in R (version 3.2.0, R Foundation for Statistical Computing).

CONFLICTS OF INTEREST

The Authors do not have any conflicts of interest.

REFERENCES

1. Polakis P. Wnt signaling in cancer. Cold Spring Harb Perspect Biol. 2012; 4.

2. Kypta RM, Waxman J. Wnt/beta-catenin signalling in prostate cancer. Nature reviews Urology. 2012; 9:418-428.

3. Verras M, Sun Z. Roles and regulation of Wnt signaling and beta-catenin in prostate cancer. Cancer letters. 2006; 237:22-32.

4. Yardy GW, Brewster SF. Wnt signalling and prostate cancer. Prostate Cancer Prostatic Dis. 2005; 8:119-126.

5. Majid S, Saini S, Dahiya R. Wnt signaling pathways in urological cancers: past decades and still growing. Molecular cancer. 2012; 11:7.

6. Bisson I, Prowse DM. WNT signaling regulates self-renewal and differentiation of prostate cancer cells with stem cell characteristics. Cell Res. 2009; 19:683-697.

7. Wan X, Liu J, Lu JF, Tzelepi V, Yang J, Starbuck MW, Diao L, Wang J, Efstathiou E, Vazquez ES, Troncoso P, Maity SN, Navone NM. Activation of beta-catenin signaling in androgen receptor-negative prostate cancer cells. Clin Cancer Res. 2012; 18:726-736.

8. Dai J, Hall CL, Escara-Wilke J, Mizokami A, Keller JM, Keller ET. Prostate cancer induces bone metastasis through Wnt-induced bone morphogenetic protein-dependent and independent mechanisms. Cancer Res. 2008; 68:5785-5794.

9. Verras M, Brown J, Li X, Nusse R, Sun Z. Wnt3a growth factor induces androgen receptor-mediated transcription and enhances cell growth in human prostate cancer cells. Cancer Res. 2004; 64:8860-8866.

10. Anastas JN, Moon RT. WNT signalling pathways as therapeutic targets in cancer. Nature reviews Cancer. 2013; 13:11-26.

11. Lu W, Tinsley HN, Keeton A, Qu Z, Piazza GA, Li Y. Suppression of Wnt/beta-catenin signaling inhibits prostate cancer cell proliferation. Eur J Pharmacol. 2009; 602:8-14.

12. Grandy D, Shan J, Zhang X, Rao S, Akunuru S, Li H, Zhang Y, Alpatov I, Zhang XA, Lang RA, Shi DL, Zheng JJ. Discovery and characterization of a small molecule inhibitor of the PDZ domain of dishevelled. J Biol Chem. 2009; 284:16256-16263.

13. Korinek V, Barker N, Morin PJ, van Wichen D, de Weger R, Kinzler KW, Vogelstein B, Clevers H. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science. 1997; 275:1784-1787.

14. Heuberger J, Birchmeier W. Interplay of cadherin-mediated cell adhesion and canonical Wnt signaling. Cold Spring Harb Perspect Biol. 2010; 2:a002915.

15. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. J Clin Invest. 2009; 119:1420-1428.

16. Khan MI, Hamid A, Adhami VM, Lall RK, Mukhtar H. Role of epithelial mesenchymal transition in prostate tumorigenesis. Current pharmaceutical design. 2015; 21:1240-1248.

17. Chen G, Shukeir N, Potti A, Sircar K, Aprikian A, Goltzman D, Rabbani SA. Up-regulation of Wnt-1 and beta-catenin production in patients with advanced metastatic prostate carcinoma: potential pathogenetic and prognostic implications. Cancer. 2004; 101:1345-1356.

18. Thiele S, Gobel A, Rachner TD, Fuessel S, Froehner M, Muders MH, Baretton GB, Bernhardt R, Jakob F, Gluer CC, Bornhauser M, Rauner M, Hofbauer LC. WNT5A has anti-prostate cancer effects *in vitro* and reduces tumor growth in the skeleton *in vivo*. Journal of bone and mineral research. 2015; 30:471-480.

19. Syed Khaja AS, Helczynski L, Edsjo A, Ehrnstrom R, Lindgren A, Ulmert D, Andersson T, Bjartell A. Elevated level of Wnt5a protein in localized prostate cancer tissue is associated with better outcome. PLoS One. 2011; 6:e26539.

20. Khaja AS, Egevad L, Helczynski L, Wiklund P, Andersson T, Bjartell A. Emphasizing the role of Wnt5a protein expression to predict favorable outcome after radical prostatectomy in patients with low-grade prostate cancer. Cancer medicine. 2012; 1:96-104.

21. Yamamoto H, Oue N, Sato A, Hasegawa Y, Yamamoto H, Matsubara A, Yasui W, Kikuchi A. Wnt5a signaling is involved in the aggressiveness of prostate cancer and expression of metalloproteinase. Oncogene. 2010; 29:2036-2046.

22. Gujral TS, Chan M, Peshkin L, Sorger PK, Kirschner MW, MacBeath G. A noncanonical Frizzled2 pathway regulates epithelial-mesenchymal transition and metastasis. Cell. 2014; 159:844-856.

23. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144:646-674.

24. Sherwood V. WNT Signaling: an Emerging Mediator of Cancer Cell Metabolism? Molecular and cellular biology. 2015; 35:2-10.

25. Sethi JK, Vidal-Puig A. Wnt signalling and the control of cellular metabolism. Biochem J. 2010; 427:1-17.

26. Sherwood V, Chaurasiya SK, Ekstrom EJ, Guilmain W, Liu Q, Koeck T, Brown K, Hansson K, Agnarsdottir M, Bergqvist M, Jirstrom K, Ponten F, James P, Andersson T. WNT5A-mediated beta-catenin-independent signalling is a novel regulator of cancer cell metabolism. Carcinogenesis. 2014; 35:784-794.

27. Selnaes KM, Gribbestad IS, Bertilsson H, Wright A, Angelsen A, Heerschap A, Tessem MB. Spatially matched *in vivo* and *ex vivo* MR metabolic profiles of prostate cancer -- investigation of a correlation with Gleason score. NMR in biomedicine. 2013; 26:600-606.

28. Giskeodegard GF, Bertilsson H, Selnaes KM, Wright AJ, Bathen TF, Viset T, Halgunset J, Angelsen A, Gribbestad IS, Tessem MB. Spermine and citrate as metabolic biomarkers for assessing prostate cancer aggressiveness. PloS one. 2013; 8:e62375.

29. Bertilsson H, Angelsen A, Viset T, Skogseth H, Tessem MB, Halgunset J. A new method to provide a fresh frozen prostate slice suitable for gene expression study and MR spectroscopy. The Prostate. 2011; 71:461-469.

30. Tessem MB, Bertilsson H, Angelsen A, Bathen TF, Drablos F, Rye MB. A Balanced Tissue Composition Reveals New Metabolic and Gene Expression Markers in Prostate Cancer. PloS one. 2016; 11:e0153727.

31. Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, Kalyana-Sundaram S, Wei JT, Rubin MA, Pienta KJ, Shah RB, Chinnaiyan AM. Integrative molecular concept modeling of prostate cancer progression. Nat Genet. 2007; 39:41-51.

32. Jin F, Qu X, Fan Q, Wang L, Tang T, Hao Y, Dai K. Regulation of prostate cancer cell migration toward bone marrow stromal cell-conditioned medium by Wnt5a signaling. Mol Med Rep. 2013; 8:1486-1492.

33. Lee GT, Kang DI, Ha YS, Jung YS, Chung J, Min K, Kim TH, Moon KH, Chung JM, Lee DH, Kim WJ, Kim IY. Prostate cancer bone metastases acquire resistance to androgen deprivation via WNT5A-mediated BMP-6 induction. Br J Cancer. 2014; 110:1634-1644.

34. McDonald SL, Silver A. The opposing roles of Wnt-5a in cancer. Br J Cancer. 2009; 101:209-214.

35. Zhu N, Qin L, Luo Z, Guo Q, Yang L, Liao D. Challenging role of Wnt5a and its signaling pathway in cancer metastasis (Review). Experimental and therapeutic medicine. 2014; 8:3-8.

36. Topol L, Jiang X, Choi H, Garrett-Beal L, Carolan PJ, Yang Y. Wnt-5a inhibits the canonical Wnt pathway by promoting GSK-3-independent beta-catenin degradation. J Cell Biol. 2003; 162:899-908.

37. Gravdal K, Halvorsen OJ, Haukaas SA, Akslen LA. A switch from E-cadherin to N-cadherin expression indicates epithelial to mesenchymal transition and is of strong and independent importance for the progress of prostate cancer. Clin Cancer Res. 2007; 13:7003-7011.

38. Markert EK, Mizuno H, Vazquez A, Levine AJ. Molecular classification of prostate cancer using curated expression signatures. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:21276-21281.

39. Rye MB, Bertilsson H, Drablos F, Angelsen A, Bathen TF, Tessem MB. Gene signatures ESC, MYC and ERG-fusion are early markers of a potentially dangerous subtype of prostate cancer. BMC medical genomics. 2014; 7:50.

40. van der Graaf M, Schipper RG, Oosterhof GO, Schalken JA, Verhofstad AA, Heerschap A. Proton MR spectroscopy of prostatic tissue focused on the detection of spermine, a possible biomarker of malignant behavior in prostate cancer. Magma (New York, NY). 2000; 10:153-159.

41. Costello LC, Franklin RB. The clinical relevance of the metabolism of prostate cancer; zinc and tumor suppression: connecting the dots. Molecular cancer. 2006; 5:17.

42. Li L, Li W. Epithelial-mesenchymal transition in human cancer: Comprehensive reprogramming of metabolism, epigenetics, and differentiation. Pharmacol Ther. 2015.

43. Compagnone A, Bandino A, Meli F, Bravoco V, Cravanzola C, Parola M, Colombatto S. Polyamines modulate epithelial-to-mesenchymal transition. Amino Acids. 2012; 42:783-789.

44. Wang Y, Xia XQ, Jia Z, Sawyers A, Yao H, Wang-Rodriquez J, Mercola D, McClelland M. In silico estimates of tissue components in surgical samples based on expression profiling data. Cancer Res. 2010; 70:6448-6455.

45. Sboner A, Demichelis F, Calza S, Pawitan Y, Setlur SR, Hoshida Y, Perner S, Adami HO, Fall K, Mucci LA, Kantoff PW, Stampfer M, Andersson SO, et al. Molecular sampling of prostate cancer: a dilemma for predicting disease progression. BMC Med Genomics. 2010; 3:8.

46. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, Antipin Y, Mitsiades N, Landers T, et al. Integrative genomic profiling of human prostate cancer. Cancer Cell. 2010; 18:11-22.

47. Erho N, Crisan A, Vergara IA, Mitra AP, Ghadessi M, Buerki C, Bergstralh EJ, Kollmeyer T, Fink S, Haddad Z, Zimmermann B, Sierocinski T, Ballman KV, et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. PLoS One. 2013; 8:e66855.

48. The Cancer Genome Atlas. (TCGA). [Available from: http://cancergenome.nih.gov/].

49. Langer DL, van der Kwast TH, Evans AJ, Plotkin A, Trachtenberg J, Wilson BC, Haider MA. Prostate tissue composition and MR measurements: investigating the relationships between ADC, T2, K(trans), v(e), and corresponding histologic features. Radiology. 2010; 255:485-494.

50. Provencher SW. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. Magnetic resonance in medicine. 1993; 30:672-679.

51. Bertilsson H, Tessem MB, Flatberg A, Viset T, Gribbestad I, Angelsen A, Halgunset J. Changes in gene transcription underlying the aberrant citrate and choline metabolism in human prostate cancer samples. Clin Cancer Res. 2012; 18:3261-3269.

52. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545-15550.

A novel non-canonical Wnt signature for prostate cancer aggressiveness

SUPPLEMENTARY FIGURES AND TABLES

A The *main cohort*

(*Continued*)

Figure S1
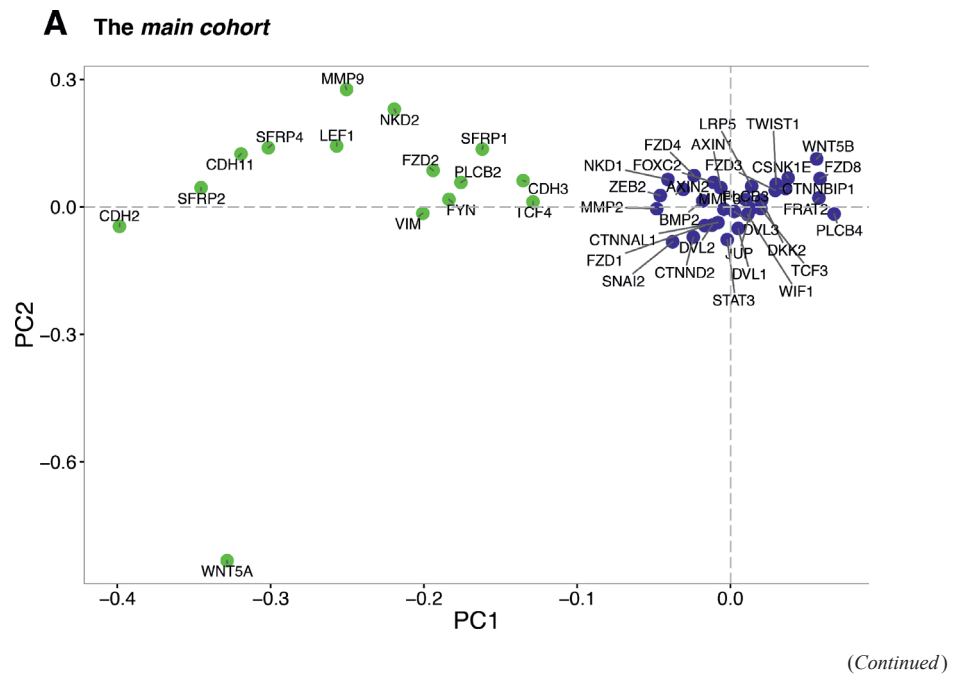
B GSE8218
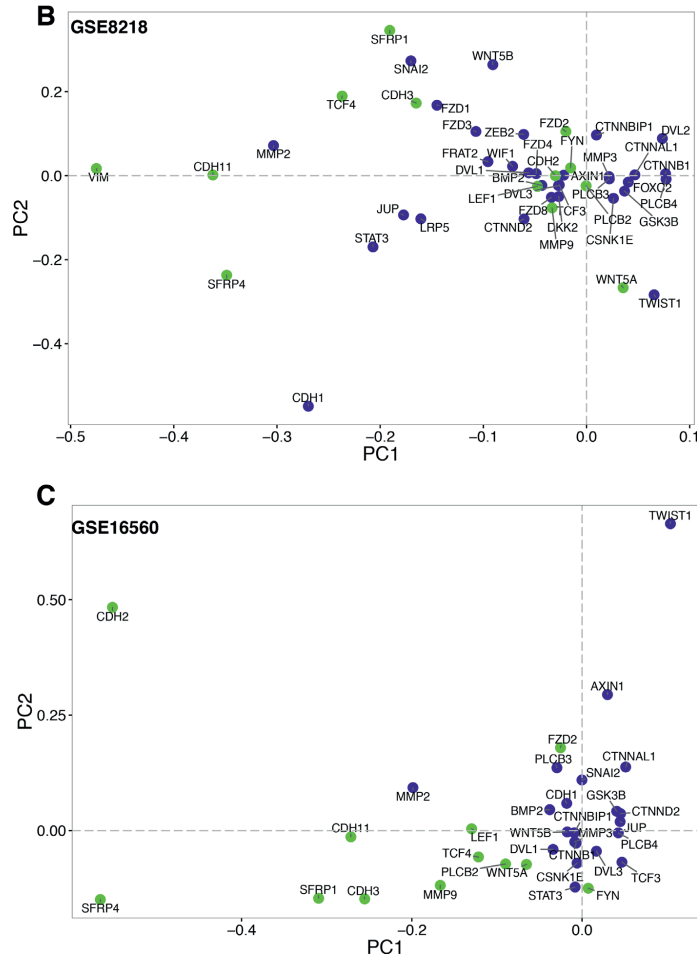
C GSE16560
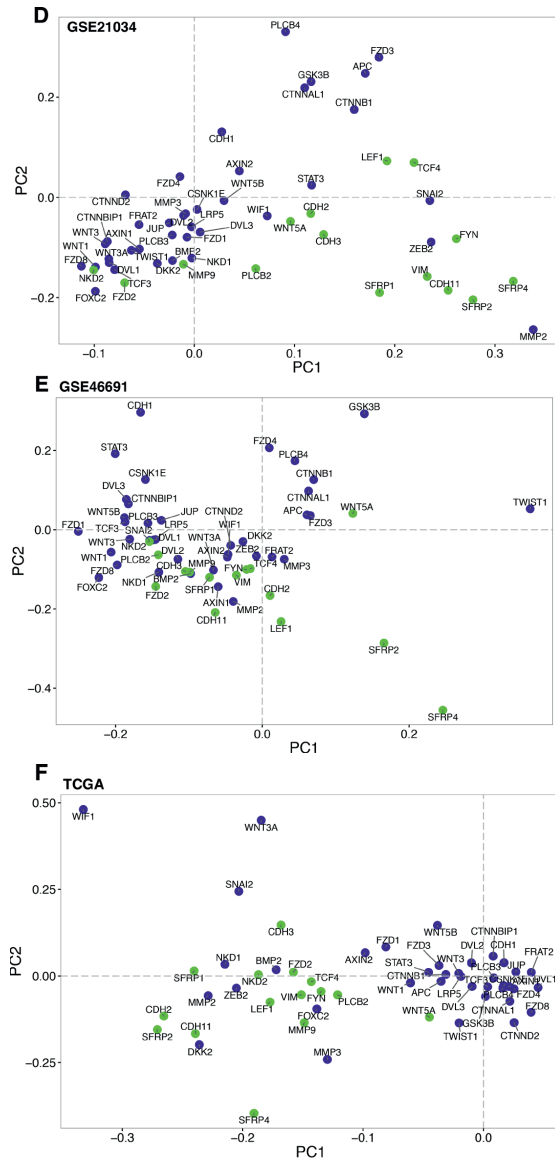
(*Continued*)

Supplementary Figure 1: A-F. PCA plots of the central WNT-EMT genes in the main cohort (A) and validation cohorts **D-E.** Genes in the NCWP-EMT signature is marked in green.

(*Continued*)

G   GSE8218
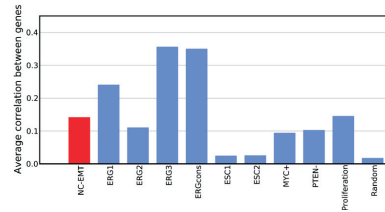
H   GSE16560
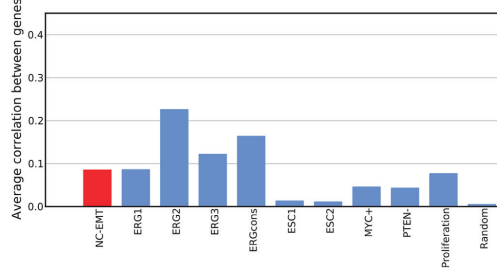
I   GSE21034
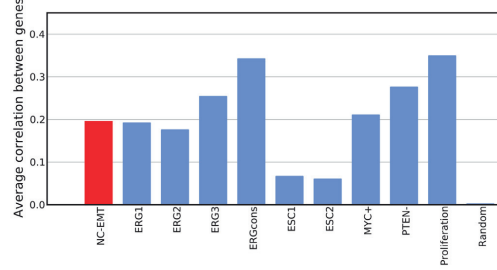
J   GSE46691
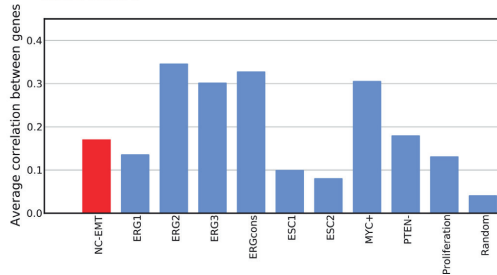
Supplementary Figure 1 *(Continued)*: **G-J.** NCWP-EMT gene signature co-expression correlations (red), compared to other gene signatures (blue). (see Supplementary Table S1 for references on the validation cohorts.)
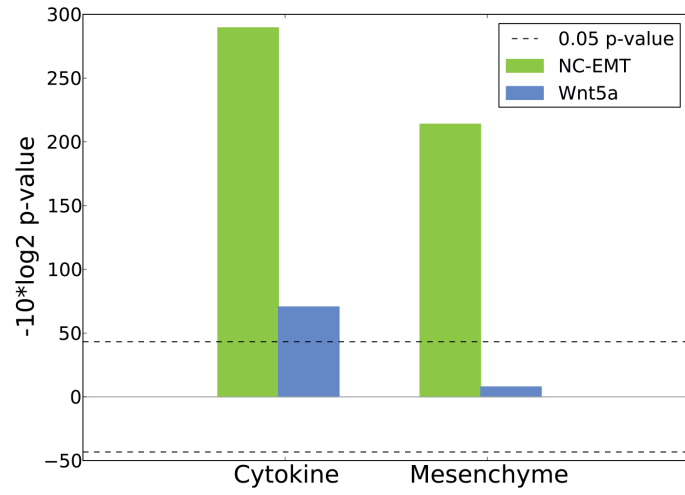
Supplementary Figure 2: P-values plot of NCWP-EMT (green) and Wnt5a (blue) coexpression with cytokine and mesenchyme gene expression signatures.
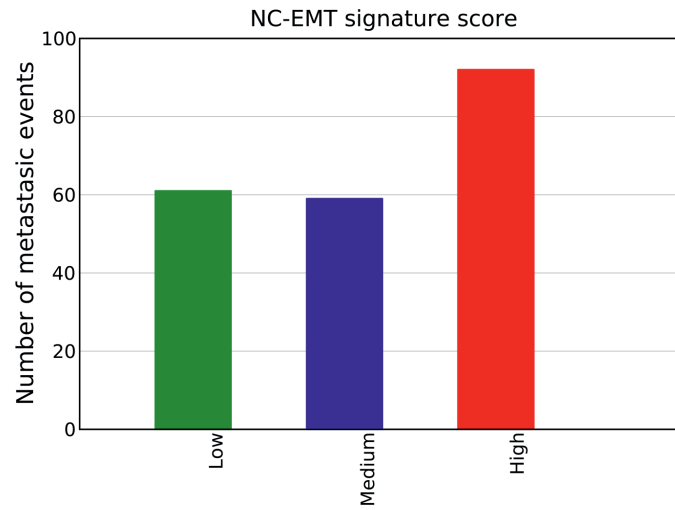
Supplementary Figure 3: The number of metastatic events was significantly higher in high NCWP-EMT score patients (red) compared to low (green) and intermediate (blue) score patients, Chi-square test p-value 0.00029.

Supplementary Table 1: Overview of the validation cohorts used in the study

GEO accession	Total number of PCa samples	Low Gleason <= 7	High Gleason >= 8	Wnt-genes present in data (total=53)	NC-EMT genes present in data (total=15)	Type of follow up data (events)	Reference
GSE8218	65	54	11	45	13	survival	(1)
GSE16560	281	200	81	33	12	survival	(2)
GSE21034	131	120	11	53	15	recurrence (27)	(3)
GSE46691	545	334	211	53	15	metastasis (212)	(4)
TCGA	497	297	200	53	15	No data	(5)

1. Wang Y, Xia XQ, Jia Z, Sawyers A, Yao H, Wang-Rodriquez J, et al. In silico estimates of tissue components in surgical samples based on expression profiling data. Cancer Res. 2010;70:6448-55.

2. Sboner A, Demichelis F, Calza S, Pawitan Y, Setlur SR, Hoshida Y, et al. Molecular sampling of prostate cancer: a dilemma for predicting disease progression. BMC medical genomics. 2010;3:8.

3. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic profiling of human prostate cancer. Cancer cell. 2010;18:11-22.

4. Erho N, Crisan A, Vergara IA, Mitra AP, Ghadessi M, Buerki C, et al. Discovery and Validation of a Prostate Cancer Genomic Classifier that Predicts Early Metastasis Following Radical Prostatectomy. PLoS One. 2013;8.

5. The Cancer Genome Atlas (TCGA) [Available from: http://cancergenome.nih.gov/].

Table S2: Overview of all the selected genes used for analysis of the Wnt Pathway, and the p-values for alterations in expression between cancer and normal, and high and low Gleason samples. In addition pathway classification and relevance in prostate cancer based on previous literature is noted, as well as our own classification of the genes in relation to prostate cancer.

Gene	P-values from t-test				Pathway	Relevance, prostate cancer[c]	Classification
	Cancer/normal			High/Low GL grade			
	All data	Balanced[a]	Unbalanced[b]				
AES	0.0007 dw	0.002dw	ns	ns	C	No	CA conf.
APC	Not found	Not found	Not found	Not found	C	Maybe	Unavailable
APC2	Not found	Not found	Not found	Not found	C	No	Unavailable
AR	ns	ns	ns	ns	C?		Not a marker
AXIN1	9.3e-6 up	0.004 up	0.005 up	ns	C	Maybe	CA pos
AXIN2	1.7e-7 dw	0.02 dw	0.0005 dw	ns	C	Maybe	Str. pos
BAMBI	4.1e-6 up	0.02 up	0.0002 up	ns	C	No	Str. neg
BIRC5	0.01 up	ns (0.09 up)	ns	0.0001 up	C	No	GL high
BMP2	0.03 dw	ns	0.04 dw	ns	Epithelial		Str. pos
BTRC	0.0001 up	ns	3.8e-5 up	0.006 dw	C	Maybe	Str. neg, GL Low
CACYBP	ns	ns	ns	ns	C	No	Not a marker
CAMK2A	0.001 dw	ns	0.0002 dw	ns	NC: Ca2+	Maybe	Str. pos
CAMK2B	4.7e-8 up	ns	6.6e-11 up	0.03 dw	NC: Ca2+	Maybe	Str. neg
CAMK2D	ns	ns	0.009 dw	ns	NC: Ca2+	Maybe	Str. pos
CAMK2G	2.5e-20 dw	3.4e-5dw	4.4e-18/dw	ns	NC: Ca2+	Maybe	Str. pos
CARM1	ns	ns	0.005 dw	ns	C? AR-B-cat	Maybe	Str. pos
CCND1	2.4e-7 dw	ns	2.4e-8 dw	ns	C, NC: PCP	Maybe	Str. pos
CCND2	1.1e-13 dw	0.002 dw	2.7e-12 dw	ns	C, NC: PCP	Highly	Str. pos
CCND3	ns	ns	0.008 dw	0.02 up	C, NC: PCP	Maybe	Str. pos
CD44	0.004 dw	ns	0.0007 dw	ns	C? AR-B-cat	Highly	Str. pos
CDH1	2.2e-5 up	ns	3.2e-8 up	ns (0.07 dw)	Cell adhesion	Highly	Str. neg
CDH11	ns	ns	0.02 dw	0.001 up	Mesenchymal	Highly	GL high
CDH2	0.01 dw	ns	0.002 dw	0.0003 up	Cell adhesion	Highly	Str. pos, GL high
CDH3	0.05 dw	ns	0.008 dw	ns	Cell adhesion	Maybe	Str. pos
CDH4	Not found	Not found	Not found	Not found	Cell adhesion	No	Unavailable
CDH5	ns	ns	ns	0.0003 up	Cell adhesion	No	GL high
CDH6	ns	ns	ns	ns	Cell adhesion	No	Not a marker
CER1	ns	ns	ns	ns	C	No	Not a marker
CHD8	ns	ns	ns	ns	C	Maybe	Not a marker
COX2	Not found	Not found	Not found	Not found	C?	No	Unavailable
CREBBP	0.005 dw	ns	0.03 dw	ns	C	No	Str. pos
CSNK1A1	ns	ns	ns	ns	C	No	Not a marker
CSNK1A1L	ns	ns	ns	ns	C	No	Not a marker
CSNK1E	2.6e-8 up	0.02 up	1.2e-6up	0.02 dw	C	No	Str. neg, GL low
CSNK2A1	0.006up	ns	0.01 up	0.05 up	C	No	Str. neg, GL high
CSNK2A2	ns	ns	ns	ns	C	No	Not a marker
CSNK2B	0.05 up	ns	ns	ns	C	No	Not a marker
CTBP1	0.0009 up	0.01 up	0.002up	ns	C	No	CA pos
CTBP2	0.0006 up	ns	0.006 up	0.05 dw	C	No	Str. neg
CTNNA1	ns	ns	0.04 dw	ns	Cell adhesion	Maybe	Not a marker
CTNNAL1	0.0002 up	ns	4.2e-5 up	ns	Cell adhesion	No	Str. neg
CTNNB1	0.0004 dw	ns	0.0008 dw	ns	C	Highly	Str. pos
CTNNBIP1	ns	0.01 dw	ns	0.003 dw	C	Maybe	CA conf., GL low
CTNNBL1	1.4e-6 dw	ns (0.09 dw)	5.7e-5 dw	ns	C	No	Str. pos
CTNND1	0.007 dw	ns	0.002 dw	0.0003 up	Cell adhesion	Highly	Str. pos
CTNND2	1.2e-9 up	0.02 up	3.7e-10 up	0.01 up	Cell adhesion	Maybe	Str. neg, GL high
CUL1	6.0e-7 dw	ns	3.6e-7 dw	ns	C	No	Str. pos
CXXC4	0.0002 up	0.02 up	0.01 up	ns	C	No	CA pos
DAAM1	0.006dw	ns	0.03 dw	ns	NC: PCP	No	Str. pos
DAAM2	8.3e-9 dw	0.04 dw	3.0e-8 dw	ns	NC: PCP	No	Str. pos
DKK1	ns	ns	ns	ns	C	Highly	Not a marker
DKK2	0.03 dw	ns	0.04 dw	ns	C	No	Str. pos
DKK3	1.9e-11 dw	0.0009 dw	2.4e-9 dw	ns	C ?	Maybe	Str. pos
DKK4	Not found	Not found	Not found	Not found	C	No	Unavailable
DVL1	0.0006 up	ns (0.06 up)	0.01 up	ns	C, NC: PCP	Highly	Str. neg
DVL2	0.002 dw	ns (0.06 dw)	0.03 dw	ns	C, NC: PCP	No	Str. pos
DVL3	0.03 dw	ns	0.0006 dw	ns	C, NC: PCP	No	Str. pos
EDN1	ns	ns	ns	ns	C?	No	Not a marker
ELK1	1.6e-10 up	6.8e-5 up	2.0e-5 up	ns	W5	No	CA pos
EP300	ns	ns	ns	ns	C	No	Not a marker
ERG	3.4e-9up	5.4e-5up	0.0002 up	ns	C?	Maybe	CA pos

[a] Balanced for stroma content. [b] Unbalanced for stroma content. [c] Based on previous findings: Highly = Alterations found in prostate cancer, Maybe = alterations found in other cancers etc.. [d]Our classification of the gene based on the p-values. GL - Gleason grade, ns - Not significant, Up - Up regulated in cancer/high GL, Dw - Down regulated in cancer/high GL, C - Canonical, NC - non-canonical, PCP - Planar cell polarity pathway, Ca2+ - Calcium pathway, W5 - Wnt5/Fzd2 pathway, Str - Stroma, CA - cancer, B - benign epithelium, Conf - Confounded, Red-NCWP-EMT gene signature, Red and Blue - central genes used in paper.

Gene	P-values from t-test				Pathway	Relevance, prostate cancer[c]	Classification
	Cancer/normal			High/Low GL grade			
	All data	Balanced[a]	Unbalanced[b]				
FBXW11	0.03 dw	ns	0.004 dw	ns	C	No	Str. pos
FHL2	5.6e-12 dw	0.04 dw	7.2e-17 dw	ns	C? AR-B-cat	Maybe	Str. pos
FN1	ns	ns	ns	ns	Mesenchymal	Maybe	Not a marker
FOSL1	ns	ns	ns	0.02 dw	C	No	GL low
FOXC2	1.9e-7 dw	ns	2.8e-8 dw	ns	Mesenchymal	Maybe	Str. pos
FRAT1	ns	ns	ns	ns	C	No	Not a marker
FRAT2	1.5e-8 up	0.05 up	1.6e-8 up	ns	C	No	Str. neg
FYN	ns	ns	0.009 dw	0.0002 up	W5	Maybe	Str. pos, GL high
FZD1	1.7e-11 dw	4.2e-7 dw	0.0003		C	Maybe	CA neg
FZD10	ns	ns	ns	ns	C, NC: PCP	No	Not a marker
FZD2	ns (0.1 dw)	ns	0.009 dw	0.003 up	C, NC: Ca2+, W5	Maybe	Str. pos, GL high
FZD3	0.009 dw	ns	0.03 dw	ns	C, NC: (PKA?, Ca2+?)	No	Str. pos
FZD4	4.5e-7 up	0.01 up	2.7e-5 up	ns	C	Highly	Str. neg
FZD5	ns	ns	ns	ns (0.07 dw)	C	No	Not a marker
FZD6	ns	ns	ns	ns	NC: Ca2+	Highly	Not a marker
FZD7	1.5e-12dw	0.02 dw	1.4e-13 dw	ns	C NC:l PCP	Maybe	Str. pos
FZD8	1.6e-5 up	0.03 up	0.0004 up	ns	C	No	Str. neg
FZD9	2.2e-11 dw	8.5e-5 dw	5.8e-7 dw	ns	NC: ERK	No	CA neg
GPC4	0.02 dw	ns	0.002 dw	ns	NC: PCP	No	Str. pos
GRIP1	Not found	Not found	Not found	Not found	C? AR-B-cat	Maybe	Unavailable
GSK3B	0.0001 up	0.005 up	0.02 up	ns	C	Maybe	CA pos
IGF1	5.2e-5 dw	0.005 dw	0.003 dw	ns	C? AR-B-cat	Maybe	CA neg
JUN	0.03 dw	ns	ns	0.0008 dw	C, NC: PCP	Highly	GL low
JUP	5.2e-10 up	0.006 up	4.2e-9 up	ns	Cell adhesion	Maybe	Str. neg
LEF1	2.9e-9 up	2.9e-5 up	0.0002 up	0.02 up	C	Highly	CA pos, GL high
LRP5	5.3e-6 up	0.0003 up	0.01 up	0.05 dw	C	Maybe	CA pos
LRP6	Not found	Not found	Not found	Not found	C	No	Unavailable
MAP2K1	0.0002 up	ns	0.0002 up	ns	W5	No	Str. neg
MAP2K2	ns	ns	0.02 up	0.04 dw	W5	No	Str. neg
MAP3K7	ns	ns	ns	ns	C	No	Not a marker
MAPK10	0.0022 dw	0.04 dw	0.01 dw	ns	NC: PCP	No	CA neg
MAPK8	ns	ns	ns	ns	NC: PCP	No	Not a marker
MAPK9	ns	ns	ns	ns	NC: PCP	No	Not a marker
MMP2	ns	ns	0.03 dw	ns	Mesenchymal	Highly	Str. pos
MMP3	ns	0.02 up	ns (0.06 dw)	ns	Mesenchymal		CA conf.
MMP7	ns	0.008 up	0.01 dw	ns	C	Maybe	CA conf.
MMP9	0.0005 up	0.0001 up	ns	ns	Mesenchymal		CA conf.
MYC	1.9e-5 up	0.01 up	0.0005 up	ns	C	Highly	CA pos
NFAT5	ns	ns	ns	ns	NC: Ca2+	No	Not a marker
NFATC1	ns	ns	ns	ns	NC: Ca2+	No	Not a marker
NFATC2	Not found	Not found	Not found	Not found	NC: Ca2+	No	Unavailable
NFATC3	2.7e-5 dw	0.01 dw	4.5e-5 dw	ns	NC: Ca2+	No	Str. pos
NFATC4	2.3e-7 dw	ns (0.06 dw)	1.0e-6 dw	ns	NC: Ca2+	No	Str. pos
NKD1	1.5e-5 dw	ns	3.0e-8 dw	ns	C, NC: PCP	No	Str. pos
NKD2	0.001 dw	ns	0.0001 dw	ns (0.09 up)	C, NC: PCP	No	Str. pos
NLK	ns	ns	0.05 up	ns	C	No	Str. neg
PCDH11Y	Not found			Not found	Cell adhesion	No	Unavailable
PLCB1	0.04 dw	ns	ns	ns	NC: Ca2+	No	Not a marker
PLCB2	ns	0.04 up	ns	0.0007 up	NC: Ca2+	No	GL high
PLCB3	0.02 dw	ns	0.03 dw	ns	NC: Ca2+	No	Str. pos
PLCB4	0.02 up	ns	0.03 up	ns	NC: Ca2+	No	Str. neg
PORCN	ns	ns	ns	ns	C	No	Not a marker
PPARD	ns	ns	ns	ns	C	No	Not a marker
PPP3CA	2.0e-10 up	0.002 up	4.0e-8 up	ns	NC: Ca2+	No	Str. neg
PPP3CB	1.1e-16 dw	0.0004 dw	1.0e-14 dw	ns	NC: Ca2+	No	Str. neg
PPP3CC	ns	ns	ns	ns	NC: Ca2+	No	Not a marker
PPP3R1	ns	ns	0.005 dw	ns	NC: Ca2+	No	Str. pos
PPP3R2	Not found	Not found	Not found	Not found	NC: Ca2+	No	Unavailable
PRICKLE1	0.0003 dw	ns	1.2e-6 dw	ns	NC: PCP	No	Str. pos
PRICKLE2	6.4e-18 dw	0.001 dw	3.8e-18 dw	ns	NC: PCP	No	Str. pos
PRKACA	Not found	Not found	Not found	Not found	C	No	Unavailable
PRKACB	ns	ns	ns	0.0001 dw	C	No	GL low
PRKACG	Not found	Not found	Not found	Not found	C	No	Unavailable
PRKCA	8.2e-14 dw	0.0008 dw	3.2e-11 dw	ns	NC: Ca2+	No	Str. pos
PRKCB	1.6e-9 dw	0.002 dw	2.8e-10 dw	ns	NC: Ca2+	No	Str. pos
PRKCG	Not found	Not found	Not found	Not found	NC: Ca2+	No	Unavailable
PRKX	ns	ns	ns	ns	C	No	Not a marker
PSEN1	ns	ns	ns	ns	C	No	Not a marker

[a] Balanced for stroma content.[b] Unbalanced for stroma content. [c] Based on previous findings: Highly = Alterations found in prostate cancer, Maybe = alterations found in other cancers etc.. [d]Our classification of the gene based on the p-values. GL - Gleason grade, ns - Not significant, Up - Up regulated in cancer/high GL, Dw - Down regulated in cancer/high GL, C - Canonical, NC - non-canonical, PCP - Planar cell polarity pathway, Ca2+ - Calcium pathway, W5 - Wnt5/Fzd2 pathway, Str - Stroma, CA - cancer, B - benign epithelium, Conf - Confounded, Red-NCWP-EMT gene signature, Red and Blue - central genes used in paper.

Gene	P-values from t-test				Pathway	Relevance, prostate cancer[c]	Classification
	Cancer/normal			High/Low GL grade			
	All data	Balanced[a]	Unbalanced[b]				
RAC1	ns	ns	0.04 dw	ns	NC: PCP	No	Str. pos
RAC2	ns	ns	ns	0.01 up	NC: PCP	No	Not a marker
RAC3	1.5e-14 up	5.4e-5 up	2.4e-10 up	ns	NC: PCP	No	CA pos
RBX1	ns	ns	ns	ns	C	No	Not a marker
RHOA	0.005 dw	ns	0.004 dw	ns	NC: PCP	Maybe	Str. pos
ROCK1	0.01 dw	0.03 dw	ns	ns	NC: PCP	No	CA conf.
ROCK2	2.7e-9 dw	0.01 dw	1.1e-7 dw	ns	NC: PCP	No	Str. pos
RUVBL1	8.4e-6 up	0.02 up	0.0006 up	ns	C	No	Str. neg
SENP2	ns	ns	ns	ns	C	No	Not a marker
SFRP1	0.04 dw	ns	0.004 dw	ns	C, NC: PCP	Highly	Str. pos
SFRP2	ns	ns	0.01 dw	0.0001 up	C	Highly	GL high
SFRP4	0.0009 up	0.002 up	ns	0.0001 up	C	Highly	B neg, GL high
SFRP5	Not found	Not found	Not found	Not found	C	No	Unavailable
SIAH1	ns	ns	0.008 dw	ns	C	No	Str. pos
SKP1	0.0002 dw	ns	3.4e-5 dw	ns	C	No	Str. pos
SMAD2	ns	ns	ns	ns	C	No	Not a marker
SMAD3	6.7e-8 dw	ns	2.9e-8 dw	ns	C	No	Str. pos
SMAD4	0.007 dw	0.05 dw	ns	ns	C	No	Not a marker
SNAI1	Not found	Not found	Not found	Not found	Mesenchymal	Highly	Unavailable
SNAI2	6.1e-14 dw	3.7e-5 dw	1.7e-9dw	ns	Mesenchymal	Highly	CA neg
SOX10	Not found	Not found	Not found	Not found	Mesenchymal	No	Unavailable
SOX17	0.0003 dw	ns	0.004 dw	ns	C	No	Str. pos
SRC	ns	ns	ns	0.0003 dw	W5	Maybe	GL low
STAT3	ns	ns	ns	ns	W5	Maybe	Not a marker
TBL1X	2.9e-9 dw	ns	6.8e-12 dw	0.05 dw	C	No	Str. pos
TBL1XR1	ns	ns	ns	ns	C	No	Not a marker
TBL1Y	ns	ns	ns	ns	C	No	Not a marker
TCF3	1.4e-9 up	0.01 up	3.0e-8 up	ns	C	Maybe	Str. neg
TCF4	0.01 dw	ns	0.003 dw	0.03 up	C	Highly	Str. pos, GL high
TCF7	ns	ns	ns	ns	C	No	Not a marker
TCF7L1	2.6e-9 dw	0.02 dw	3.1e-8 dw	ns	C	No	Str. pos
TCF7L2	ns	ns	ns	ns	C	No	Not a marker
TGFB1	Not found	Not found	Not found	Not found	Mesenchymal	Maybe	Unavailable
TGFB2	1.0e-7 dw	ns	4.7e-8 dw	ns	Mesenchymal	Maybe	Str. pos
TLE1	8.2e-7 up	0.0002 up	0.006 up	0.009 up	C	No	CA pos, GL high
TLE2	7.7e-12 dw	9.8e-5 dw	1.9e-7 dw	ns	C	No	CA neg
TLE3	4.3e-6 up	ns	1.8e-6 up	ns	C	No	Str. neg
TLE4	0.0002 dw	ns	0.0002 dw	ns	C	No	Str. pos
TLE6	Not found	Not found	Not found	Not found	C	No	Unavailable
TP53	ns	ns	ns	0.02 dw	C	Maybe	GL low
TWIST1	6.8e-11 up	0.0003 up	3,5e-7 up	ns	Mesenchymal	Highly	CA pos
VANGL1	Not found	Not found	Not found	Not found	NC: PCP	No	Unavailable
VANGL2	1.3e-5 dw	0.006 dw	0.003 dw	0.0005 dw	NC: PCP	No	CA neg, GL low
VIM	3.1e-6 dw	ns	9.2e-7 dw	0.002 up	Mesenchymal	Highly	Str. pos, GL high
WIF1	1.1e-8 dw	1.3e-5 dw	0.003 dw	ns	C	Highly	CA neg
WNT1	Not found	Not found	Not found	Not found	C	Highly	Unavailable
WNT10A	Not found	Not found	Not found	Not found		No	Unavailable
WNT10B	Not found	Not found	Not found	Not found		No	Unavailable
WNT11	0.01 dw	ns	ns	ns	NC: PCP	Highly	Not a marker
WNT13	Not found	Not found	Not found	Not found		No	Unavailable
WNT16	ns	ns	ns	ns	C	Maybe	Not a marker
WNT2	ns	ns	ns	ns	C	Highly	Not a marker
WNT2B	ns	ns (0.06 up)	0.02/ns dw	ns	C	No	Not a marker
WNT3	Not found	Not found	Not found	Not found			Unavailable
WNT3A	Not found	Not found	Not found	Not found	C	Maybe	Unavailable
WNT4	ns	ns	ns	ns			Not a marker
WNT5A	ns	ns	ns	6.7e.5 up	NC: Ca2+, W5, (PCP?)	Highly	GL high
WNT5B	3.7e-5 dw	0.02 dw	0.003 dw	0.02 dw	NC: Ca2+, W5	Maybe	CA neg, GL low
WNT6	Not found	Not found	Not found	Not found	C	Highly	Unavailable
WNT7A	ns	ns	ns	ns		No	Not a marker
WNT7B	ns	ns (0.1 up)	ns	ns	C	Maybe	Not a marker
WNT8A	Not found	Not found	Not found	Not found		No	Unavailable
WNT8B	Not found	Not found	Not found	Not found		No	Unavailable
WNT9A	Not found	Not found	Not found	Not found		No	Unavailable
WNT9B	Not found	Not found	Not found	Not found		No	Unavailable
ZEB1	Not found	Not found	Not found	Not found	Mesenchymal		Unavailable
ZEB2	3.3e-10 dw	ns	3.9e-12 dw	ns	Mesenchymal		Str. pos

[a] Balanced for stroma content. [b] Unbalanced for stroma content. [c] Based on previous findings: Highly = Alterations found in prostate cancer, Maybe = alterations found in other cancers etc.. [d]Our classification of the gene based on the p-values. GL - Gleason grade, ns - Not significant, Up - Up regulated in cancer/high GL, Dw - Down regulated in cancer/high GL, C - Canonical, NC - non-canonical, PCP - Planar cell polarity pathway, Ca2+ - Calcium pathway, W5 - Wnt5/Fzd2 pathway, Str - Stroma, CA - cancer, B - benign epithelium, Conf - Confounded, Red-NCWP-EMT gene signature, Red and Blue - central genes used in paper.

Supplementary Table 3: Most significant GO-terms for top 1000 differentially expressed genes between samples with high and low NCWP-EMT GSEA score, and between samples with high and low expression levels of Wnt5a

Source	Term	Benjamini corrected p-value
NC-EMT		
Terms related to cell surface and extracellular functions (Mesenchymal):		
SP_PIR_KEYWORDS	signal	5.3e-27
UP_SEQ_FEATURE	signal peptide	5.7e-26
SP_PIR_KEYWORDS	glycoprotein	3.3e-20
UP_SEQ_FEATURE	glycosylation site: N-linked(GlcNAc?)	9.0e-16
GOTERM_CC_FAT	extracellular region part	6.4e-16
UP_SEQ_FEATURE	topological domain: Extracellular	9.0e-12
GOTERM_CC_FAT	proteinaceous extracellular matrix	7.2e-13
GOTERM_BP_FAT	biological adhesion	1.2e-11
GOTERM_BP_FAT	cell adhesion	1.5e-11
SP_PIR_KEYWORDS	Secreted	8.9e-11
GOTERM_CC_FAT	extracellular matrix	1.2e-11
GOTERM_CC_FAT	extracellular region	3.4e-10
GOTERM_CC_FAT	plasma membrane part	1.4e-10
GOTERM_CC_FAT	integral to plasma membrane	3.4e-10
GOTERM_CC_FAT	plasma membrane	2.9e-10
GOTERM_CC_FAT	intrinsic to plasma membrane	4.2e-10
GOTERM_CC_FAT	extracellular space	2.2e-6
Terms related immune response and inflammation (Cytokine):		
GOTERM_BP_FAT	positive regulation of immune system process	7.0e-15
GOTERM_BP_FAT	positive regulation of immune response	4.9e-11
GOTERM_BP_FAT	response to wounding	7.5e-9
GOTERM_BP_FAT	inflammatory response	8.1e-9
GOTERM_BP_FAT	defense response	9.3e-9
GOTERM_BP_FAT	positive regulation of cell activation	1.4e-9
GOTERM_BP_FAT	regulation of lymphocyte activation	7.8e-9
GOTERM_BP_FAT	regulation of T cell activation	1.5e-8
GOTERM_BP_FAT	regulation of leukocyte activation	2.3e-8
GOTERM_BP_FAT	positive regulation of response to stimulus	2.0e-7
GOTERM_BP_FAT	activation of immune response	1.3e-5
GOTERM_BP_FAT	immune effector process	5.2e-5
WNT5A		
SP_PIR_KEYWORDS	SH2 domain	1.2e-2
GOTERM_MF_FAT	kinase binding	8.6e-2
SP_PIR_KEYWORDS	membrane	2.0e-2
SP_PIR_KEYWORDS	transmembrane	6.2e-2
BIOCARTA	T Helper Cell Surface Molecules	5.8e-2
SP_PIR_KEYWORDS	Ehler Danlos syndrome	5.6e-2

Supplementary Table 4: Metabolite concentration and alterations between low, intermediate and high NC-EMT signature score, for the main cohort

	Metabolite concentration (mmol/kg wet weight)			P-values [a]		
Signature score	**Low**	**Int**	**High**	**High/Low**	**High/Int**	**Int/Low**
Metabolite	**Median (IQR) (n=32)**	**Median (IQR) (n=31)**	**Median (IQR) (n=32)**			
Alanine	2.23 (1.68-2.82)	2.46 (1.87-3.20)	2.08 (1.65-2.60)	0.407	0.228	0.792
Choline	1.07 (0.68-1.42)	1.03 (0.69-1.92)	1.06 (0.66-1.65)	0.806	0.511	0.792
Citrate	7.31 (5.57-11.56)	6.38 (4.56-11.58)	3.55 (2.08-7.25)	3.38E-04*	0.018*	0.282
Creatine	1.93 (1.39-2.71)	2.32 (2.01-2.67)	1.98 (1.61-2.50)	0.684	0.259	0.592
Ethanolamine	0.00 (0.00-0.29)	0.00 (0.00-0.17)	0.00 (0.00-0.19)	0.881	0.884	0.938
Glucose	0.09 (0.00-0.52)	0.00 (0.00-0.36)	0.00 (0.00-0.25)	0.449	0.884	0.658
Glutamate	4.68 (3.20-6.67)	5.77 (3.85-7.52)	5.24 (4.21-7.51)	0.974	0.905	0.993
Glutamine	2.75 (2.10-3.55)	2.76 (2.27-3.86)	2.80 (2.53-3.49)	0.974	0.884	0.938
GPC	0.98 (0.53-1.36)	0.44 (0.73-1.16)	0.74 (0.50-1.06)	0.407	0.578	0.754
GPEA	0.00 (0.00-0.68)	0.09 (0.00-0.53)	0.10 (0.00-0.58)	0.958	0.884	0.938
Glycine	2.38 (1.57-3.04)	2.48 (1.89-3.43)	2.55 (1.91-3.63)	0.881	0.884	0.754
Isoleucine	0.12 (0.00-0.20)	0.17 (0.12-0.29)	0.19 (0.10-0.32)	0.163	0.884	0.132
Lactate	16.74 (13.66-22.57)	21.86 (15.62-26.41)	20.31 (16.47-25.26)	0.974	0.511	0.425
Leucine	0.42 (0.21-0.58)	0.49 (0.36-0.65)	0.57 (0.33-0.93)	0.684	0.884	0.792
Myo-inositol	8.17 (6.46-10.21)	9.83 (7.43-12.87)	9.55 (8.36-11.65)	0.834	0.497	0.131
Phosphocholine	0.55 (0.28-1.04)	0.87 (0.54-1.17)	0.74 (0.48-1.32)	0.806	0.689	0.282
PEA	2.18 (1.25-2.89)	2.92 (2.23-4.00)	2.88 (2.20-3.79)	0.159	0.511	2.14E-03*
Putrescine	0.04 (0.00-0.35)	0.08 (0.00-0.20)	0.00 (0.00-0.53)	0.601	0.884	0.754
Scyllo-inositol	0.41 (0.32-0.59)	0.44 (0.32-0.64)	0.45 (0.38-0.62)	0.589	0.327	0.792
Spermine	1.55 (1.02-2.36)	1.23 (0.67-2.27)	0.75 (0.39-1.43)	3.38E-04*	0.028*	0.113
Succinate	0.64 (0.42-0.88)	0.61 (0.49-0.94)	0.60 (0.45-0.71)	0.589	0.511	0.965
Taurine	3.84 (2.84-4.94)	4.05 (4.62-6.96)	5.76 (4.03-7.28)	0.131	0.884	0.022*
Valine	0.32 (0.20-0.46)	0.38 (0.24-0.56)	0.41 (0.28-0.59)	0.806	0.689	0.938

ex vivo.

[a] P-values from Linear mixed model adjusted or multiple samples per patient, corrected for multiple testing by the Benjamini and Hochberg procedure.

* Indicates signficant p-values

Abbreviations: Int - Intermediate, IQR - Interquartile range, GPC - Glycerophosphocholine, GPEA - Glycerophosphoethanolamine, PEA - Phosphoethanolamine.

Supplementary Table 5: P-values for metabolite alteration between low and high NC-EMT samples, adjusting for Gleason risk and tissue heterogeneity

	P-values [a]					
Adjusted for:	**Patient**	**Patient, Gleason risk**	**Patient, Stroma**	**Patient, Cancer**	**Patient, Benign epithelium**	**Patient, Luminal space**
Alanine	0.407	0.352	0.623	0.509	0.329	0.480
Choline	0.806	0.839	0.847	0.837	0.818	0.735
Citrate	3.38E-04*	2.42E-03*	4.20E-04*	2.01E-04*	8.61E-04*	2.81E-03 *
Creatine	0.684	0.855	0.650	0.671	0.805	0.735
Ethanolamine	0.881	0.958	0.950	0.837	0.867	0.839
Glucose	0.449	0.657	0.109	0.176	0.515	0.417
Glutamate	0.974	0.839	0.650	0.916	0.856	0.965
Glutamine	0.974	0.839	0.650	0.837	0.867	0.903
GPC	0.407	0.303	0.623	0.509	0.249	0.578
GPEA	0.958	0.958	0.734	0.837	0.933	0.965
Glycine	0.881	0.955	0.623	0.671	0.981	0.965
Isoleucine	0.163	0.143	0.119	0.176	0.109	0.239
Lactate	0.974	0.958	0.650	0.837	0.933	0.965
Leucine	0.684	0.958	0.542	0.587	0.818	0.735
Myo-inositol	0.834	0.839	0.650	0.787	0.832	0.735
Phosphocholine	0.806	0.958	0.623	0.587	0.867	0.735
PEA	0.159	0.143	0.109	0.127	0.221	0.246
Putrescine	0.601	0.397	0.623	0.587	0.805	0.691
Scyllo-inositol	0.589	0.566	0.650	0.587	0.515	0.735
Spermine	3.38E-04*	1.62E-03*	4.20E-04*	2.01E-04*	8.61E-04*	2.81E-03*
Succinate	0.589	0.414	0.767	0.658	0.515	0.691
Taurine	0.131	0.143	0.119	0.127	0.092	0.161
Valine	0.806	0.958	0.623	0.671	0.828	0.843

[a] P-values from Linear mixed modell, corrected for multiple testing by the Benjamini and Hochberg procedure.
* Indicates signficant p-values
Abbreviations: GPC - Glycerophosphocholine, GPEA - Glycerophosphoethanolamine, PEA - Phosphoethanolamine.

Supplementary Table 6: There were no metabolic differences between low, intermediate and high WNT5A gene expression in the main cohort or between low/moderate and high Wnt5a IHC expression in the validation cohort

WNT5A expression/ Metabolite	Main cohort P-values [a]			Validation cohort P-values [b]
	Low/Int	Low/High	Int/High	IHC: Low/High
Alanine	0.502	0.710	0.990	0.670
Choline	0.836	0.826	0.990	0.926
Citrate	0.715	0.826	0.892	0.623
Creatine	0.882	0.826	0.892	0.406
Ethanolamine	0.987	0.240	0.190	0.175
Glucose	0.836	0.882	0.990	0.396
Glutamate	0.836	0.882	0.892	0.755
Glutamine	0.836	0.882	0.990	0.672
GPC	0.715	0.882	0.990	0.433
GPEA	0.744	0.826	0.990	0.666
Glycine	0.638	0.882	0.892	0.746
Isoleucine	0.855	0.882	0.892	0.367
Lactate	0.684	0.882	0.892	0.682
Leucine	0.744	0.826	0.892	0.117
Myo-inositol	0.684	0.826	0.990	0.603
Phosphocholine	0.638	0.882	0.892	0.628
PEA	0.568	0.826	0.892	0.610
Putrescine	0.502	0.710	0.990	0.423
Scyllo-inositol	0.684	0.826	0.990	0.765
Spermine	0.836	0.918	0.990	0.619
Succinate	0.638	0.826	0.990	0.498
Taurine	0.502	0.826	0.892	0.430
Valine	0.684	0.826	0.990	0.313

[a] P-values from Linear mixed model adjusted or multiple samples per patient, corrected for multiple testing by the Benjamini and Hochberg procedure.
[b] P-values from independent samples t-test. Not corrected for multiple testing
Abbreviations: Int - Intermediate, GPC - Glycerophosphocholine,
GPEA - Glycerophosphoethanolamine, PEA - Phosphoethanolamine.

Supplementary Table 7: Immunohistochemistry scoring for staining index (SI)

Score	0	1	2	3
Staining intensity	No detectable staining	Weak staining	Moderate staining	Strong staining
Percentage of positive cells	0%	1-10%	11-50%	>50%
Staining index (SI)	0	1,2	3,4,6	9
Staining classification	Negative	Weak	Moderate	Strong

Staining index (SI) is obtained by multiplying the scores of staining intensity and percentage of positive cells.

Supplementary Table 8: Multivariate cox proportional hazards analyses of biochemical recurrence

Variables	NCWP-EMT model (AIC = 65.61)		Post-operative Gleason score model (AIC = 64.24)	
	Hazard ratio (95% CI)	P-value	Hazard ratio (95% CI)	P-value
Post-operative Gleason score (≤7* and ≥8)	-	-	5.42 (0.86-34.11)	0.072
Pathological T-stage (≤T2c* and ≥T3a)	3.79 (0.91-1.63)	0.088	2.12 (0.35-12.77)	0.41
Preoperative PSA ≤10* and ≥10	2.14 (9.82-17.5)	0.22	2.47 (0.69-8.85)	0.17
NCWP-EMT Continuous score/100 (-4.4-5.4)	1.22 (0.91-1.63)	0.188	-	-

* Indicates the group used as a reference in each analysis.

Paper III

SFRP4 gene expression is increased in aggressive prostate cancer

Elise Sandsmark[1]*, Maria K. Andersen[1], Anna M. Bofin[2], Helena Bertilsson[3,4], Finn Drabløs[4], Tone F. Bathen[1], Morten B. Rye[4,5Δ], May-Britt Tessem[1Δ]*

1. *Department of Circulation and Medical Imaging, Faculty of Medicine and Health Sciences, NTNU - Norwegian University of Science and Technology, Trondheim, Norway*
2. *Department of Laboratory Medicine, Children's and Women's Health, Faculty of Medicine and Health Sciences, NTNU - Norwegian University of Science and Technology, Trondheim, Norway*
3. *Department of Urology, St. Olav's Hospital, Trondheim University Hospital, Trondheim Norway*
4. *Department of Cancer Research and Molecular Medicine, Faculty of Medicine and Health Sciences, NTNU - Norwegian University of Science and Technology, Trondheim, Norway*
5. *Clinic of Surgery, St. Olav's Hospital, Trondheim University Hospital, Trondheim, Norway*

[Δ] These authors contributed equally to this work

*Corresponding authors:
Elise Sandsmark: elise.sandsmark@gmail.com
May-Britt Tessem: may-britt.tessem@ntnu.no

Abstract

Increased knowledge of the molecular differences between indolent and aggressive prostate cancer is urgently needed for improved risk stratification and treatment selection for patients. SFRP4 is a modulator of the cancer-associated Wnt pathway, and previously suggested as a potential marker for prostate cancer aggressiveness. We investigated and validated the association between *SFRP4* gene expression and aggressiveness in nine independent cohorts with follow-up data (total n=2197). By differential expression and combined meta-analysis of all the cohorts, we detected a significantly higher *SFRP4* expression in cancer compared with normal samples, and in high (≥4+3) compared with low (≤3+4) Gleason score samples. *SFRP4* expression was a significant predictor of biochemical recurrence in six of seven cohorts and in the overall analysis, and was a significant predictor of metastatic event in one cohort. In our main cohort, where metabolic information was available, *SFRP4* expression correlated significantly with the concentration of citrate and spermine, two previously suggested biomarkers for aggressive prostate cancer. SFRP4 immunohistochemistry in an independent cohort (n=33) was not associated with aggressiveness. High *SFRP4* gene expression is associated with high Gleason score and recurrent prostate cancer after surgery, and future studies investigating the mechanistic as well as assessing the clinical usefulness are warranted.

Introduction

Prostate cancer is the second most common cancer in men, and the fifth leading cause of cancer related death in men worldwide[1]. The lack of accurate markers to separate aggressive from non-aggressive prostate cancer at an early time point, are causing considerable overtreatment of indolent cancers[2]. Discovery of new biomarkers of aggressiveness, as well as improved understanding of differences between indolent and aggressive prostate cancer, are therefore highly needed.

The family of secreted frizzled-related proteins (SFRP1-5) are extracellular inhibitors of Wnt signalling, a pathway identified for its role in carcinogenesis[3]. The SFRPs are in general regarded as tumour suppressors. However, the SFRPs may also have oncogenic properties, due to implications in other signalling pathways, as well as a suggested biphasic modulation of Wnt signalling[4,5]. SFRP4 is the largest and the most structurally different of the family members[6]. In several types of cancer, SFRP4 follows the tumour suppressor pattern, by epigenetic silencing due to promotor hypermethylation, and reduced gene expression, as reviewed by Pohl et al.[7]. However, for prostate cancer, increased gene expression of *SFRP4* has been detected[8,9], and shown to be a predictor of recurrent disease[10]. Additionally, *SFRP4* has been included in different gene expression signatures linked to prostate cancer aggressiveness and recurrence[10,11], including our previously published signature for non-canonical

Wnt pathway and epithelial-to-mesenchymal transition markers (NCWP-EMT)[12]. Protein levels of SFRP4 measured by immunohistochemistry are discordant in prostate cancer; Horvath et al. reported increased expression of membranous SFRP4 staining to be associated with good prognosis[13,14], while Mortensen et al. reported cytoplasmic expression to be linked to worse prognosis[10]. Overall SFRP4 seems to be a potential biomarker candidate for prostate cancer aggressiveness, and there is a need to validate and clarify the role of *SFRP4* in prostate cancer.

Reprogramming of metabolism is one of the hallmarks of cancer development[15], and in prostate cancer the metabolites citrate and spermine have shown promise as aggressive biomarkers[16,17]. Our previously published NCWP-EMT gene expression signature was associated with reduced concentrations of these metabolites[12], but the correlation between *SFRP4* gene and protein expression levels, and citrate and spermine has not previously been investigated in prostate cancer. Our previously published method enabling integration of gene expression levels, with metabolic data and histopathology of the exact same samples, gives an excellent opportunity to investigate this[18].

The overall aim of this study was to validate *SFRP4* expression in prostate cancer, and its relation to the aggressiveness of the disease. The results were validated in eight independent publically available gene expression prostate cancer cohorts with patient follow-up information. Furthermore, SFRP4

immunohistochemistry was investigated in a separate cohort. Our approach of including several independent patient cohorts gave increased statistical power, and improved accuracy and generalisation of the results.

Results

The main cohort consisted of 156 prostate tissue samples from 41 patients, of which 116 were cancer tissue samples. An additional cohort termed the IHC cohort, included 40 cancer samples. Eight independent prostate cancer validation cohorts were downloaded from gene expression omnibus (GEO) and The Cancer Genome Atlas (TCGA), giving a total number of 2197 samples from 1830 patients. Five of the validation cohorts included normal samples as well as cancer samples. Clinical and histopathological data for all patients included in the study are listed in Table 1.

Table 1. Clinical and histopathological variables of the cohorts.

Clinical variables	Main cohort	IHC cohort	Erho *et al.*	TCGA-PRAD	CAM Ross-Adams *et al.*
Samples (patients)	156(41)	40 (40)	545 (545)	549 (497)	186 (163)
Cancer samples (patients)	116 (41)	40 (40)	545 (545)	497 (497)	112 (112)
Age at diagnosis, years (median, range)	64 (48-69)	61 (48-73)	65.3±6.4	61 (41-78)	61 (41-73)
PSA before surgery, ng/mL (median, range)	9.1 (4.0-45.8)	8.85 (5.2-18)	-	7.4 (0.7-107)	7.8 (3.2-23.7)
Gleason score					
Low (≤3+4)	60 (52%)	19 (47.5%)	334 (61%)[a]	207(42%)	82 (73%)
High (≥4+3)	56 (48%)	21 (52.5%)	211 (39%)[a]	289 (58%)	30 (27%)
Pathological T stage					
pT1	-	-	-	-	-
pT2	25 (60%)	27 (68%)	219 (40%)	187 (38%)	33 (29%)
pT3	40 (35%)	12 (30%)	253 (47%)	293 (59%)	74 (66%)
pT4	-	-		9 (2%)	1 (1%)
No data	6 (5%)	1 (2%)	73 (13%)	8 (1%)	4 (4%)
Follow-up					
Event	BCR	BCR	Metastasis	BCR	Recurrence
Occurred	13 (32%)	16 (40%)	212 (39%)[b]	91 (18%)	19 (17%)
Not occurred	21 (51%)	21 (53%)	333 (69%)[b]	399 (80%)	93 (83%)
No data	7 (17%)	3 (8%)	-	7 (2%)	-

Clinical variable	STK Ross-Adams *et al.*	Wang *et al.*	Sboner *et al.*	Taylor *et al.*	Mortensen *et al.*
Samples (patients)	94 (94)	136 (82)	281 (281)	160 (131)	50 (50)
Cancer samples (patients)	94 (94)	65 (56)	281 (281)	131 (131)	36 (36)
Age years (median, range)		63 (43-77)	74 (51-91)	58 (37-73)	63 (46-71)
PSA before surgery, ng/mL (median, range)	7.95 (1.5-117)	6.62 (1.0-75)		5.92 (1.0-46)	16 (5.0-43)
Gleason score					
Low (≤3+4)	60 (64%)	50 (77%)	162 (58%)	107 (82%)	32 (89%)[a]
High (≥4+3)	34 (36%)	15 (23%)	119 (42%)	24 (18%)	4 (11%)[a]
Pathological T-stage					
pT1	-	1 (2%)	281 (100%)	-	-
pT2	48 (51%)	32 (57%)	-	85 (65%)	19 (53%)
pT3	42 (45%)	20 (35%)	-	40 (30%)	17 (47)
pT4	-	1 (2%)	-	6 (5%)	-
No data	4 (4%)	2 (2%)	-	-	-
Follow-up					
Event	Recurrence	BCR	PCa-death	BCR	BCR
Occurred	45 (48%)	29 (52%)	165 (59%)	27 (21%)	22 (61%)
Not occurred	48 (51%)	27 (48%)	116 (41%)	104 (79%)	14(39%)
No data	1 (1%)	-	-	-	-

BCR – biochemical recurrence, PCa-death – prostate cancer-specific death. [a]In Erho *et al.* and Mortensen *et al.*: Low Gleason score ≤7, high Gleason score ≥8. [b]In Erho *et al.* metastatic progression at 10-year patient follow-up.

SFRP4 expression in cancer. In the main cohort, there was significantly higher *SFRP4* expression (log fold change) in cancer samples compared with normal samples (p<0.001, Figure 1a). This was also true for four of the five independent validation cohorts which included expression data from both cancer and normal samples (Figure 1a). Meta-analysis of all the cohorts gave a significant combined Cohen's *d* of 0.81, which is considered a large effect-size (Figure 1c). Together, this clearly describes upregulation of *SFRP4* in prostate cancer compared with normal prostate tissue.

SFRP4 expression in cancer with high Gleason score. In the main cohort, there was significantly higher *SFRP4* expression (log fold change) in high Gleason score (≥4+3) compared with low Gleason score (≤3+4) cancer samples (p<0.001, Figure 1b), and this was also confirmed in six of the seven validation cohorts (Figure 1b). Meta-analysis of all the analysed cohorts further strengthened this finding, giving a significant combined Cohen's *d* of 0.57 (Figure 1d). The Mortensen et al. cohort was excluded from differential expression analysis between high and low Gleason score due to the low number of high Gleason score samples (n=4).

SFRP4 and patient follow-up. In the main cohort, the continuous value of *SFRP4* expression was a significant predictor of biochemical recurrence after radical prostatectomy, by univariate Cox proportional hazards analysis (Figure 2). This was further confirmed in five of the six validation cohorts with biochemical recurrence as endpoints (Figure 2). Meta-analysis of the six microarray based cohorts gave a significant combined *SFRP4* standardised hazard ratio of 1.70 for prediction of biochemical recurrence (p<0.001, Figure 2). Continuous *SFRP4* expression was not a predictor of prostate cancer-specific death in the watchful waiting Sboner et al. cohort (Figure 2). Furthermore, logistic regression showed SFRP4 expression to be a predictor of metastases after radical prostatectomy in the Erho et al. cohort (Figure 2).

SFRP4 expression and metabolism. In the main cohort, the *SFRP4* expression level was negatively correlated with concentrations of citrate (r=-0.53, p<0.001) and the polyamine spermine (r=-0.49, p<0.001) (Figure 3). These were the highest

correlations to citrate and spermine of all the genes in our previously published NCWP-EMT gene expression signature[12] (Supplementary Table S1).

SFRP4 immunohistochemistry. In the IHC cohort, seven of the 40 samples had to be excluded from further immunohistochemistry analysis due to insufficient or lack or tumour cells in the stained section. We did not detect membranous SFRP4 staining of prostate cancer cells in any samples. However, different staining intensities of cytoplasmic SFRP4 staining, as well as different proportion of positive cancer cells, were identified (Figure 4). Full immunohistochemistry scoring of each sample along with clinical, histopathologic and metabolic data can be found in Supplementary Table S2.

There was no relationship between Gleason score and SFRP4 cytoplasmic staining index (Fisher's exact p=1.0). This was also the case when looking at staining intensities and staining proportions, separately (Fisher's exact p=0.80 and p=0.82, respectively). Furthermore, neither associations between SFRP4 staining and biochemical recurrence (Log-rank: staining index p=0.87, intensity p=0.82, proportion p=0.95), nor any significant correlation between SFRP4 staining index and citrate and spermine concentrations (r=0.13 p=0.47 and r=0.18 p=0.32, respectively) were detected.

Discussion

In this study, we performed analyses of *SFRP4* gene expression, and validated the results in eight independent prostate cancer cohorts. We showed *SFRP4* expression to be increased in prostate cancer, and further increased in high Gleason score compared with low Gleason score cancer. Additionally, *SFRP4* expression was found to be a predictor of worse outcome in prostatectomy treated prostate cancer patients, and the expression level was negatively correlated with citrate and spermine concentrations in the samples. Together, these results underpin *SFRP4* as a biomarker candidate of prostate cancer aggressiveness.

We showed *SFRP4* gene expression to be increased in prostate cancer compared with normal tissue in five of six cohorts, and in the combined meta-analysis of all cohorts. This is in agreement with Luo et al. and Wissmann et al., who investigated matched tumour and normal tissue,
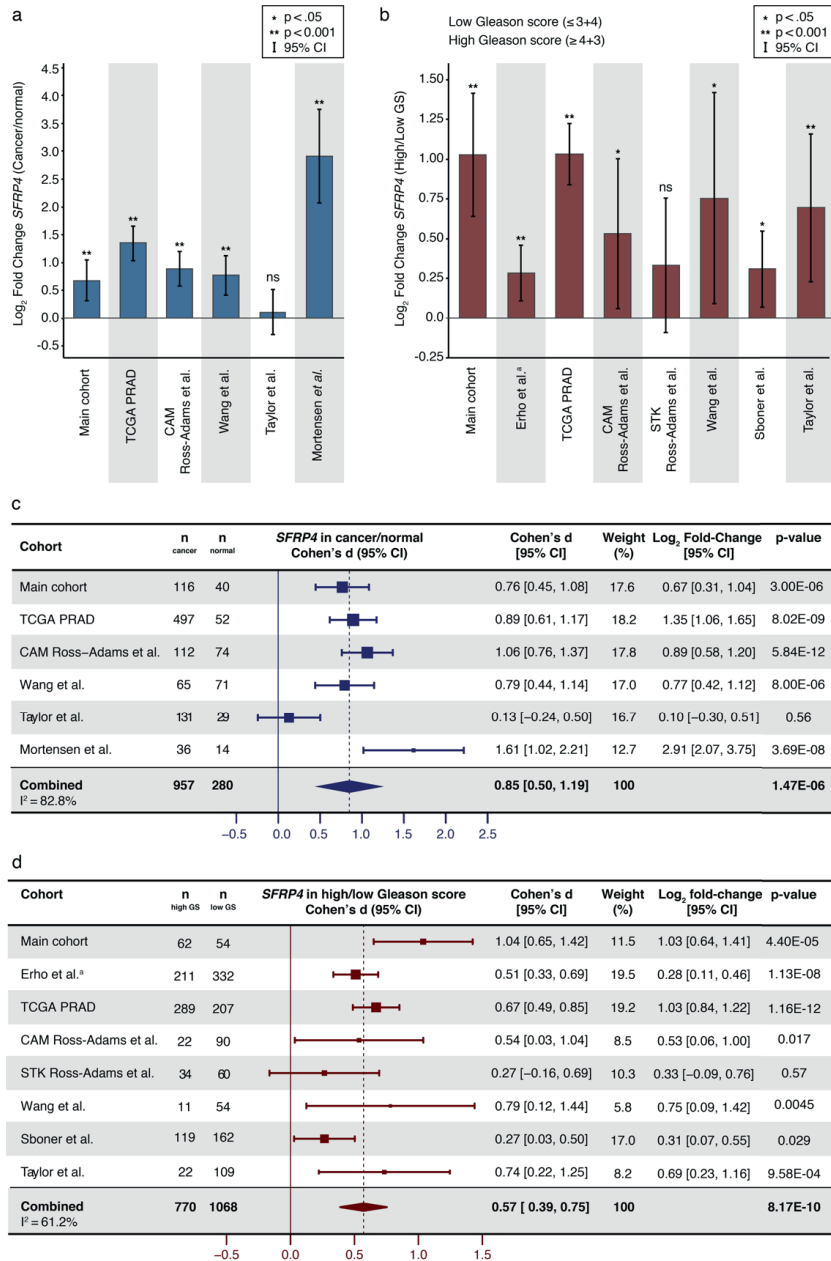
Figure 1. **SFRP4 gene expression in prostate cancer.** (**a**) Log_2 fold change of *SFRP4* expression in cancer compared to normal samples (**b**) Log_2 fold change of *SFRP4* expression in high Gleason score ($\geq$4+3) compared with low Gleason score ($\leq$3+4) samples (**c**) Forest plot and meta-analysis of *SFRP4* expression in prostate cancer compared with normal prostate samples. (**d**) Forest plot and meta-analysis of *SFRP4* expression in high Gleason score ($\geq$4+3) compared with low Gleason score ($\leq$3+4) prostate cancer samples. Fieller's method was used to obtain confidence interval (CI) for the fold changes. [a]In the Erho *et al.* cohort low Gleason score was defined as $\leq$7, and high Gleason score as $\geq$8

samples from 16 and 56 prostate cancer patients, respectively[8,9]. García-Tobilla et al. did not find significantly different expression levels of *SFRP4* between normal and prostate cancer tissue, however, the study suffered from small sample size (normal n=4, cancer n=11)[19]. In a previous paper, we also showed increased *SFRP4* when balancing for stroma content in the samples[12]. Interestingly, two studies detected an increase in *SFRP4* expression in prostate cancer tissue compared with benign prostate hyperplasia[19,20], but this approach was not possible to pursue in our study. To summarise, previous studies have in general reported increased *SFRP4* gene expression prostate cancer compared with normal prostate, but have been carried out on small cohorts. The result of the present study adds substantial validation for *SFRP4* expression to be increased in prostate cancer.

We showed increased expression of *SFRP4* in high Gleason score (≥4+3) compared with low Gleason score (≤3+4) samples, and an association between *SFRP4* expression and risk of biochemical recurrence and metastasis after radical prostatectomy. *SFRP4* gene expression has previously been linked to more aggressive prostate cancer. Luo et al. showed increased expression of *SFRP4* in tissue samples from prostate cancer patients with pathological stage T3a-b compared with pathological stage T2b. Mortensen et al. found *SFRP4* to be a part of two aggressive gene expression clusters, as well as an independent predictor of recurrence after prostatectomy in the Nakagawa et al. cohort[10]. Our previously published NCWP-EMT gene expression signature included SFRP4 as one of 15 genes, and was shown to be associated with biochemical recurrence and metastasis after prostatectomy[12]. Furthermore, Oncotype DX® for prostate cancer, a commercially available gene expression signature, includes *SFRP4* as one of 17 genes, which has been associated with clinical recurrence of prostate cancer after prostatectomy[11]. Our analyses of multiple independent cohorts in the current study, further support high *SFRP4* expression to be associated with more aggressive prostate cancer. To conclude, several studies[8,10-12], including the current study, support *SFRP4* gene expression to be upregulated in aggressive compared with less aggressive prostate cancer.
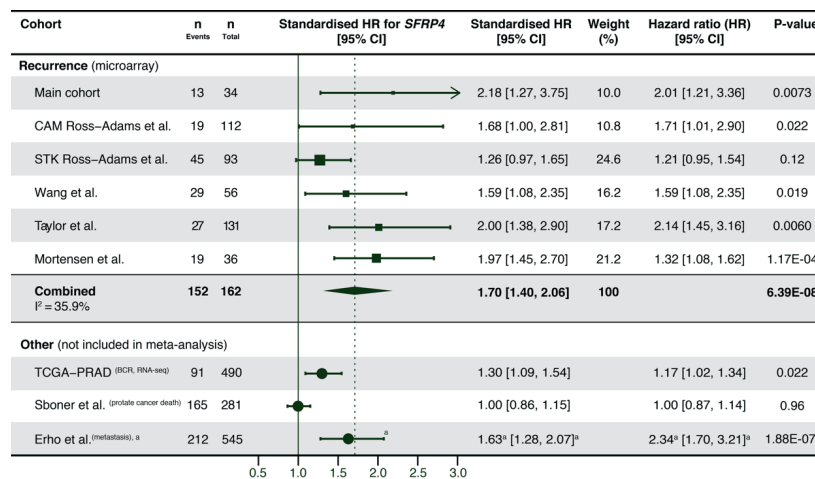

Cohort	n Events	n Total	Standardised HR for *SFRP4* [95% CI]	Standardised HR [95% CI]	Weight (%)	Hazard ratio (HR) [95% CI]	P-value
Recurrence (microarray)							
Main cohort	13	34		2.18 [1.27, 3.75]	10.0	2.01 [1.21, 3.36]	0.0073
CAM Ross–Adams et al.	19	112		1.68 [1.00, 2.81]	10.8	1.71 [1.01, 2.90]	0.022
STK Ross–Adams et al.	45	93		1.26 [0.97, 1.65]	24.6	1.21 [0.95, 1.54]	0.12
Wang et al.	29	56		1.59 [1.08, 2.35]	16.2	1.59 [1.08, 2.35]	0.019
Taylor et al.	27	131		2.00 [1.38, 2.90]	17.2	2.14 [1.45, 3.16]	0.0060
Mortensen et al.	19	36		1.97 [1.45, 2.70]	21.2	1.32 [1.08, 1.62]	1.17E-04
Combined $I^2 = 35.9\%$	152	162		**1.70 [1.40, 2.06]**	100		6.39E-08
Other (not included in meta-analysis)							
TCGA–PRAD (BCR, RNA-seq)	91	490		1.30 [1.09, 1.54]		1.17 [1.02, 1.34]	0.022
Sboner et al. (prostate cancer death)	165	281		1.00 [0.86, 1.15]		1.00 [0.87, 1.14]	0.96
Erho et al. (metastasis), a	212	545		1.63[a] [1.28, 2.07][a]		2.34[a] [1.70, 3.21][a]	1.88E-07[a]

Figure 2 Univariate cox proportional hazard analysis of *SFRP4* expression and follow-up endpoints. *SFRP4* gene expression was used as a continuous variable in the analyses. Meta-analyses were performed on the cohorts with microarray based *SFRP4* gene expression data and biochemical recurrence as endpoint. One sample per patient was selected randomly for the cohorts with multiple samples per patients (main and Wang et al. cohort). CI – confidence interval. [a]The Erho et al. cohort was analysed by logistic regression, with odds ratio as the effect size.
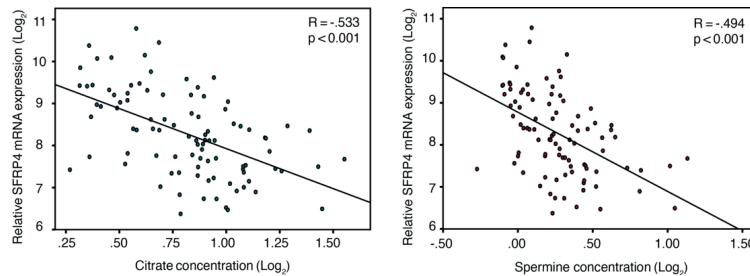
Figure 3. Correlations with metabolism. Linear Pearson correlations between *SFRP4* gene expression and citrate and spermine in the main cohort. All variables were log_2 transformed.

SFRP4 is classified as an inhibitor of Wnt signalling, a pathway implicated in carcinogenesis[3]. Consequently, SFRP4 is expected to be a tumour suppressor, and to be downregulated in aggressive cancer. As reviewed by Pohl et al., DNA hypermethylation of the *SFRP4* promotor and reduced *SFRP4* gene expression have been detected in many types of cancers, including, but not limited to, endometrial, ovarian, bladder, and oesophageal cancer[7]. Although *SFRP4* expression in prostate cancer tissue seems to deviate from this, two prostate cell line studies have supported tumour suppressor properties of SFRP4 in prostate cancer. In the first study, Horvath et al. detected that PC3 and LNCaP cell lines modified to overexpress SFRP4 proteins had reduced cellular proliferation compared to controls[13,14]. García-Tobilla et al. showed reduced gene expression of *SFRP4* in prostate cancer cell lines (LNCaP, PC3, DU145 and 22Rv1) compared with control cells (PREC)[19]. However, they did not detect DNA hypermethylation at the *SFRP4* promotors in any of the cell lines that could explain this downregulation[19]. Absence of *SFRP4* gene hypermethylation was also shown by Perry et al. in both prostate cancer cell lines and tumour tissue[20]. In contrast to García-Tobilla et al., and in coherence with human prostate cancer tissue studies, Perry et al. also detected upregulation of SFRP4 in all prostate cancer cell lines (LNCaP, PC3, DU145 and 22Rv1) compared with controls (PWR-1, RWPE1)[20]. Interestingly, in the two latter mentioned studies, DNA hypermethylation of SFRP2, SFRP3 and SFRP5 was detected in both cell lines and human prostate cancer[19,20]. This is in agreement with findings in colorectal cancer, where Suzuki et al. suggested that SFRP4 may not be an important inhibitor of the Wnt signalling pathway due to lower frequency of DNA hypermethylation and weaker

Wnt signalling inhibition compared with other *SFRP* family members[21]. This may be translatable to prostate cancer, and could explain why SFRP4 is not downregulated in prostate cancer. However, more mechanistic studies of how SFRP4 regulate the Wnt signalling pathway in prostate cancer are necessary before a conclusion can be drawn.

In the current study, we detected an association between SFRP4 expression and development of metastases after prostatectomy in the Erho et al. cohort. Bones are the most frequent site for haematogenous metastases for prostate cancer[22]. Interestingly, SFRP4 has been suggested to have an important role in bone homeostasis[23,24]. However, to our knowledge, the function of SFRP4 in bone metastases has not been specifically investigated. A hypothesis to explain the association between *SFRP4* gene expression and high Gleason score, as well as recurrence and metastasis after prostatectomy, could therefore be that SFRP4 increases the cancer cell's ability to metastasise to bone. Future studies investigating the role of SFRP4 in prostate cancer bone metastases would consequently be of interest.

For patient follow-up in this study, we used the surrogate endpoints of biochemical recurrence and metastases, in all except one cohort, Sboner et al., in which prostate cancer-specific death was used. Such surrogate endpoints are commonly used in prostate cancer studies, due to a natural long survival time of patients. Unexpectedly, we did not see any association between *SFRP4* gene expression and cancer-specific death in the Sboner et al. cohort. This cohort did, however, differ substantially from the other analysed cohorts. Whereas the cancer samples in all other cohorts were from patients undergoing radical prostatectomy, Sboner et al. was a watchful waiting cohort of patients classified with stage
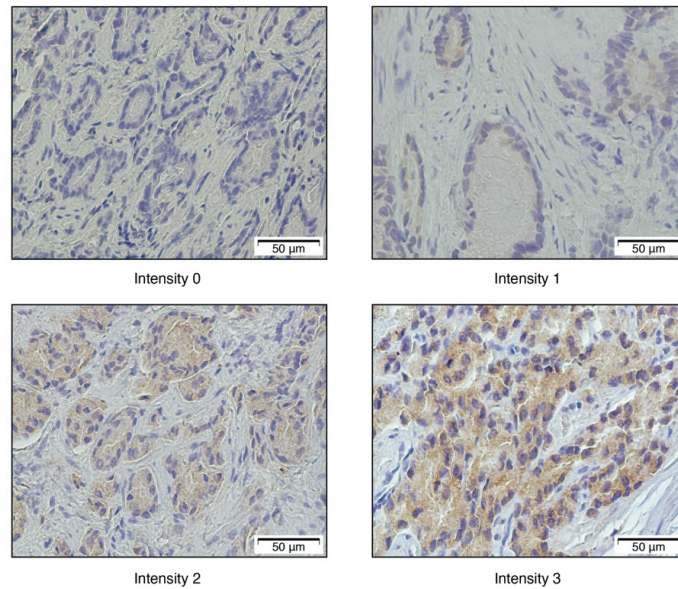
8

Figure 4. Immunohistochemistry of SFRP4. The figure shows examples of the staining intensities 0 to 3.

T1a-T1b, NX, M0 disease. These patients had incidental prostate cancer discovered by trans-urethral resection of the prostate (TURP) due to symptomatic benign prostate hyperplasia. The samples used for gene expression were from the same TURP procedure. Although most prostate cancers arise from the peripheral zone, resection performed by TURP represents the transition zone, and is likely to detect a higher rate of transition zone prostate cancers. Substantial differences in gene expression between the Sboner et al. TURP cohort and a radical prostatectomy cohort has previously been observed[25], and was related to the different zonal origins of the tumours[25]. This may limit the future clinical use of *SFRP4* expression for risk stratification in patients with transitional zone prostate cancers, and potentially also in patients with very early stage prostate cancer, and this should be further investigated.

Changes in metabolism is regarded as one of the hallmarks of cancer[15]. In prostate cancer, the concentrations of the metabolites citrate and spermine are found to be reduced in cancer compared with normal tissue[26,27], and further reduced in aggressive prostate cancer[16]. A recent study has also shown citrate and spermine to be

predictors of prostate cancer biochemical recurrence in three independent cohorts[17]. The high negative correlation between *SFRP4* expression and spermine and citrate in the main cohort of the current study thus further supports *SFRP4* expression to be associated with aggressive cancer. One of the normal functions of prostate cells is production of citrate and the polyamine spermine for the prostatic fluid, and reduced concentration of these metabolites may signify loss of normal cell function. However, whether these metabolic mechanisms are directly related to *SFRP4* expression was not investigated in the current study.

We did not find any association between immunohistochemistry staining of SFRP4 and histopathological, metabolic and follow-up data in the IHC cohort in this study. Our cohort only included tissue samples from 33 patients, as it was originally part of a demanding integrated analysis of metabolomics, histopathology and patient follow-up[12,17,28]. The small samples size limits the interpretation of our immunohistochemistry results. There are only four previous studies including immunohistochemistry of SFRP4 in prostate cancer, and there are no standardised protocols for staining or scoring. Three of these studies were based on the

9

same cohort and staining of tissue microarray (TMA) samples from 229 radical prostatectomy patients[13,14,29], where membranous SFRP4 staining was detected to be associated with good prognosis[13]. In the current study, we did not detect any membranous staining of SFRP4. The lack of membranous staining is in accordance with a previous study of Mortensen et al., which included TMA sections from 517 radical prostatectomy patients[10]. Our IHC cohort was stained by the same antibody and dilution as used in the Mortensen et al. study[10], which may explain the similar staining pattern. The use of different antibodies compared with the Horvath et al. study[13] may be a possible cause of the observed disparity of membranous staining. In addition, the relatively weak staining of SFRP4 in the current study (Figure 4) could have hidden membranous expression. In contrast to the TMA sections used in both the Mortensen et al. and Horvath et al. studies, our IHC cohort consisted of sections from needle biopsy samples. Biopsy sections are larger than TMA section, and this increases the challenges of intensity scoring due to increased heterogeneity within each sample. Additionally, the biopsies in the current study were not necessarily from the most aggressive part of the tumour, and may consequently not be representative of the lesion. As mentioned, there are limitations to the immunohistochemistry evaluation of SFRP4 in the current study, and no certain conclusion can be made based on our results. Nevertheless, we have demonstrated a few issues that are important to address before immunohistochemistry of SFRP4 can have a role in prostate cancer risk stratification. These include the lack of standardised staining and evaluation protocols, and the uncertain impact of staining heterogeneity and sampling bias.

In the current study, we did not look into possible clinical application of *SFRP4* expression, and this should be investigated in future studies. Absolute quantification of SFRP4 mRNA by real time PCR in biopsies may have a role for risk stratification and treatment selection for prostate cancer patients, including selection of patients for active surveillance, as well as patients in need of adjuvant treatment. Another interesting possibility for further studies, are investigation of the SFRP4 gene and protein expression levels in less invasive liquid biopsies such as serum, urine, prostatic fluid and seminal fluid.

In this study, we have validated the presence of increased *SFRP4* gene expression in

10

prostate cancer. We detected, and validated, higher SFRP4 expression in high Gleason score prostate cancer compared with low Gleason score cancer. We further showed that SFRP4 expression was as a predictor of the patient follow-up endpoints recurrence and metastases after prostatectomy. Finally, we showed a negative correlation between *SFRP4* expression and the metabolic markers, citrate and spermine. To conclude, *SFRP4* expression is associated with more aggressive disease, and SFRP4 deserves further attention in prostate cancer studies as a promising marker of aggressiveness.

Methods

Ethics statement. The study was approved by the central regional committee for medical and health research ethics, case numbers 010-04, 4.2007.1890, and 2009/1161(4.2007.1654). All patients in the main cohort and the IHC cohort signed a written informed consent.

Patients and samples. Samples in the main and IHC cohort are from patients diagnosed with localised or locally advanced prostate cancer, treated with radical prostatectomy at St. Olav's Hospital, Trondheim University Hospital, between 2007 to 2010. None of the patients received prostate cancer treatment prior to surgery. Samples in the main cohort were harvested from fresh-frozen prostatectomy specimens in a highly standardised method previously described by Bertilsson et al.[18]. The samples in the IHC cohort were collected as needle biopsies after prostatectomy, and snap frozen within minutes.

Follow-up. At least 5 years' follow-up data were collected for the patients in the main cohort and the IHC cohort as previously described by Braadland et al.[17]. Biochemical recurrence was defined as serum PSA levels of at least 0.2 ng/mL in two independent measurements.

Histopathology. For histopathological evaluation, a cryosection from each tissue sample in the main cohort and two formalin-fixed paraffin-embedded sections of each sample in the IHC cohort were used. All sections were evaluated by an experienced pathologist as previously described[12]. The reproducibility of the histopathological evaluation has previously been assessed in the main cohort, by an independent pathologist, blinded for previous

evaluation, where high interrater agreement was reported[12,28]. Patient post-operative Gleason score was obtained from whole-mount prostate sections according to the clinical criteria for prostate cancer. Samples and patients were divided into two groups of low Gleason score ($\leq 3+4$) and high Gleason score ($\geq 4+3$).

Metabolomics. The samples in the main cohort and IHC cohort were analysed by proton high-resolution magic angle spinning magnetic resonance spectroscopy (HR-MAS MRS) using a Bruker Avance DRX600 Spectrometer (Bruker Biopsin, Germany). LCModel was used for absolute quantification of 23 metabolites from the spectra. More details on the HR-MAS MRS acquisition and metabolite quantification have been described by Giskeødegård et al. for the main cohort[16] and Hansen et al. for the IHC cohort[28].

Microarray gene expression. Gene expression analysis was performed on the tissue samples in the main cohort after HR-MAS MRS. Illumina TotalPrep RNA amplification Kit (Ambion Inc.) and Illumina Human HT-12v4 Expression Bead Chip (Illumina) were used to measure relative gene expression as previously described by Bertilsson et al.[30].

Immunohistochemistry. In the IHC cohort, immunohistochemistry was performed using 4μm thick, formalin fixed and paraffin embedded tissue sections. Rabbit polyclonal antibody against SFRP4 (Protein Tech catalogue: 15328-1-AP) was used in a 1:200 dilution with a pH of 9. The sections were counterstained with Haematoxylin. Every section was evaluated for SFRP4 staining location (membranous or cytoplasmic). Based on the staining intensities described by Mortensen et al.[10], the samples were scored from 0-3 in regards to their most common cancer staining intensity (Figure 4). Additionally, the percentage of positive cancer cells was scored from 0-3, and was multiplied by the intensity score to obtain a staining index (0-9). For statistical analyses, the staining index was divided into three groups (0, 1-3 and 4-9). Further details of the scoring are given in Supplementary Table S3. One pathologist experienced in immunohistochemistry in addition to one physician scored all sections. When scoring differed, consensus was reached.

Validation cohorts. For validation, the following seven prostate cancer cohorts with available microarray gene expression and follow-up data were downloaded from GEO: Erho et al. (GSE46691)[31,32], CAM (Cambridge) Ross-Adams et al. (GSE70768)[33], STK (Stockholm) Ross-Adams et al. (GSE70769)[33], Wang et al. (GSE8218)[34-36], Sboner et al. (GSE16560)[37], Taylor et al. (GSE21035/32)[38], and Mortensen et al. (GSE46602)[10]. In addition, a RNA sequencing cohort of prostate adenocarcinomas, TCGA PRAD, was downloaded from TCGA[39,40]. Cancer samples for all cohorts were from radical prostatectomy specimens, except Sboner et al. which was from a watchful waiting patient cohort of incidental prostate cancer discovered by transurethral resection of the prostate. Normal samples in Mortensen et al. were from surgical prostate specimens from patients with bladder cancer, four of the normal prostate samples in Wang et al. were autopsy samples from normal subjects, the rest and the other cohorts were adjacent normal prostate tissue from prostatectomy specimens. Biochemical recurrence was the follow-up endpoint in Wang et al., Taylor et al., Mortensen et al., and TCGA PRAD. In addition to biochemical recurrence, CAM and STK Ross-Adams et al. included salvage treatment in the criterion for their recurrence endpoint. Metastasis was the end point in Erho et al., and prostate cancer-specific death was the endpoint in Sboner et al. Clinical and histopathological data of the cohorts are listed in Table 1, and an overview table of the cohorts is included as Supplementary Table S4.

Statistical analysis. When more than one probe for *SFRP4* existed in a cohort, the probe with the highest variance was chosen for statistical analyses. For all analyses, *SFRP4* gene expression data were $\log_2$ transformed if not previously performed. For the gene expression cohorts, independent sample t-tests (two-tailed) were used for comparisons between two groups. Q-Q plots were used to check the normality assumption; small deviations were accepted due to the robustness of the test. Equal variance assumption was tested by Levene's test, and corrected for when applicable. Fieller's method was used to obtain pooled confidence interval for the $\log_2$ fold changes. To obtain Cohen's *d,* a standardised effect size for meta-analyses, the difference between two means (cancer and normal, and high and low Gleason score) were divided by their pooled standard deviation.

Meta-analyses by random-effect model were performed using the *metafor* package in R[41].

In the two cohorts with multiple samples per patients (the main cohort and Wang et al.), one sample per patient was randomly selected for survival analyses. Univariate Cox proportional hazard regression analyses were performed on the continuous *SFRP4* expression. The proportional hazard assumption was tested using the *survival* package in R[42]. Standardised hazard ratios were obtained by multiplying the natural logarithm of the hazard ratio (beta) by its standard deviation[43]. Cohorts with microarray based gene expression data, and biochemical recurrence as endpoint were included in a random-effect model meta-analysis, which was performed in R using the *metafor* package[41]. Due to unavailable data for time-points of event in the Erho et al. cohort, logistical regression was used for the follow-up analyses of this cohort.

Pearson correlation coefficients (two-tailed) were used to test the correlations between gene expression and $\log_2$ transformed concentrations of the metabolites citrate and spermine in the main and IHC cohort. Fisher exact tests (two-tailed) were used to examine the relationship between immunohistochemistry staining and Gleason score, and log-rank statistics were used to investigate the relationship between SFRP4 staining and time to biochemical recurrence.

For all statistical tests the significant level was set at $p=0.05$. When mentioned, analyses were performed in R (R foundation for statistical computing v3.3.1), all other analyses were performed in SPSS (IBM SPSS Statistics v24.0).

Author Contributions

E.S., M.K.A., F.D., T.F.B. M.B.R., and M-B.T. contributed to the conception and design of the study. E.S., H.B., M.B.R and M-B.T. developed methods and performed experiments. E.S and A.M.B performed immunohistochemistry scoring. Data analysis was performed by E.S. with consult from M.B.R and M-B.T. The paper was written by E.S, and all authors edited and approved the final manuscript.

References

1   Torre, L. A. *et al.* Global cancer statistics, 2012. *CA: a cancer journal for clinicians* **65**, 87-108, doi:10.3322/caac.21262 (2015).
2   Loeb, S. *et al.* Overdiagnosis and overtreatment of prostate cancer. *European urology* **65**, 1046-1055, doi:10.1016/j.eururo.2013.12.062 (2014).
3   Nusse, R. & Varmus, H. E. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell* **31**, 99-109 (1982).
4   Bovolenta, P., Esteve, P., Ruiz, J. M., Cisneros, E. & Lopez-Rios, J. Beyond Wnt inhibition: new functions of secreted Frizzled-related proteins in development and disease. *Journal of cell science* **121**, 737-746, doi:10.1242/jcs.026096 (2008).
5   Uren, A. *et al.* Secreted frizzled-related protein-1 binds directly to Wingless and is a biphasic modulator of Wnt signaling. *The Journal of biological chemistry* **275**, 4374-4382 (2000).
6   Jones, S. E. & Jomary, C. Secreted Frizzled-related proteins: searching for relationships and patterns. *BioEssays : news and reviews in molecular, cellular and developmental biology* **24**, 811-820, doi:10.1002/bies.10136 (2002).
7   Pohl, S., Scott, R., Arfuso, F., Perumal, V. & Dharmarajan, A. Secreted frizzled-related protein 4 and its implications in cancer and apoptosis. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **36**, 143-152, doi:10.1007/s13277-014-2956-z (2015).
8   Luo, J. H. *et al.* Gene expression analysis of prostate cancers. *Molecular carcinogenesis* **33**, 25-35 (2002).
9   Wissmann, C. *et al.* WIF1, a component of the Wnt pathway, is down-regulated in prostate, breast, lung, and bladder cancer. *The Journal of pathology* **201**, 204-212, doi:10.1002/path.1449 (2003).

10   Mortensen, M. M. *et al.* Expression profiling of prostate cancer tissue delineates genes associated with recurrence after prostatectomy. *Scientific reports* **5**, 16018, doi:10.1038/srep16018 (2015).

11   Klein, E. A. *et al.* A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. *European urology* **66**, 550-560, doi:10.1016/j.eururo.2014.05.004 (2014).

12   Sandsmark, E. *et al.* A novel non-canonical Wnt signature for prostate cancer aggressiveness. *Oncotarget*, doi:10.18632/oncotarget.14161 (2016).

13   Horvath, L. G. *et al.* Membranous expression of secreted frizzled-related protein 4 predicts for good prognosis in localized prostate cancer and inhibits PC3 cellular proliferation in vitro. *Clinical cancer research : an official journal of the American Association for Cancer Research* **10**, 615-625 (2004).

14   Horvath, L. G. *et al.* Secreted frizzled-related protein 4 inhibits proliferation and metastatic potential in prostate cancer. *The Prostate* **67**, 1081-1090, doi:10.1002/pros.20607 (2007).

15   Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).

16   Giskeodegard, G. F. *et al.* Spermine and citrate as metabolic biomarkers for assessing prostate cancer aggressiveness. *PloS one* **8**, e62375, doi:10.1371/journal.pone.0062375 (2013).

17   Braadland, P. R. *et al. Ex vivo metabolic fingerprinting identifies biomarkers predictive of prostate cancer recurrence following radical prostatectomy* (Submitted: British Journal of Cancer 2017, 2016).

18   Bertilsson, H. *et al.* A new method to provide a fresh frozen prostate slice suitable for gene expression study and MR spectroscopy. *The Prostate* **71**, 461-469, doi:10.1002/pros.21260 (2011).

19   Garcia-Tobilla, P. *et al.* SFRP1 repression in prostate cancer is triggered by two different epigenetic mechanisms. *Gene* **593**, 292-301, doi:10.1016/j.gene.2016.08.030 (2016).

20   Perry, A. S. *et al.* Gene expression and epigenetic discovery screen reveal methylation of SFRP2 in prostate cancer. *International journal of cancer* **132**, 1771-1780, doi:10.1002/ijc.27798 (2013).

21   Suzuki, H. *et al.* Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. *Nature genetics* **36**, 417-422, doi:10.1038/ng1330 (2004).

22   Bubendorf, L. *et al.* Metastatic patterns of prostate cancer: an autopsy study of 1,589 patients. *Human pathology* **31**, 578-583 (2000).

23   Haraguchi, R. *et al.* sFRP4-dependent Wnt signal modulation is critical for bone remodeling during postnatal development and age-related bone loss. *Scientific reports* **6**, 25198, doi:10.1038/srep25198 (2016).

24   Simsek Kiper, P. O. *et al.* Cortical-Bone Fragility--Insights from sFRP4 Deficiency in Pyle's Disease. *The New England journal of medicine* **374**, 2553-2562, doi:10.1056/NEJMoa1509342 (2016).

25   Sinnott, J. A. *et al.* Molecular differences in transition zone and peripheral zone prostate tumors. *Carcinogenesis* **36**, 632-638, doi:10.1093/carcin/bgv051 (2015).

26   Swanson, M. G. *et al.* Quantitative analysis of prostate metabolites using 1H HR-MAS spectroscopy. *Magnetic resonance in medicine* **55**, 1257-1264, doi:10.1002/mrm.20909 (2006).

27   van der Graaf, M. *et al.* Proton MR spectroscopy of prostatic tissue focused on the detection of spermine, a possible biomarker of malignant behavior in prostate cancer. *Magma (New York, N.Y.)* **10**, 153-159 (2000).

28   Hansen, A. F. *et al.* Presence of TMPRSS2-ERG is associated with alterations of the metabolic profile in human prostate cancer. *Oncotarget*, doi:10.18632/oncotarget.9817 (2016).

29   Yip, P. Y. *et al.* Low AZGP1 expression predicts for recurrence in margin-positive, localized prostate cancer. *The Prostate* **71**, 1638-1645, doi:10.1002/pros.21381 (2011).

30   Bertilsson, H. *et al.* Changes in gene transcription underlying the aberrant citrate and choline metabolism in human prostate cancer samples. *Clinical cancer research : an official journal of the American Association for Cancer Research* **18**, 3261-3269, doi:10.1158/1078-0432.ccr-11-2929 (2012).

31   Erho, N. *et al.* Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PloS one* **8**, e66855, doi:10.1371/journal.pone.0066855 (2013).

32   Zhao, S. G. *et al.* The Landscape of Prognostic Outlier Genes in High-Risk Prostate Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **22**, 1777-1786, doi:10.1158/1078-0432.ccr-15-1250 (2016).

33   Ross-Adams, H. *et al.* Integration of copy number and transcriptomics provides risk stratification in prostate cancer: A discovery and validation cohort study. *EBioMedicine* **2**, 1133-1144, doi:10.1016/j.ebiom.2015.07.017 (2015).

34   Wang, Y. *et al.* In silico estimates of tissue components in surgical samples based on expression profiling data. *Cancer research* **70**, 6448-6455, doi:10.1158/0008-5472.can-10-0021 (2010).

35   Jia, Z. *et al.* Diagnosis of prostate cancer using differentially expressed genes in stroma. *Cancer research* **71**, 2476-2487, doi:10.1158/0008-5472.can-10-2585 (2011).

36   Chen, X. *et al.* An accurate prostate cancer prognosticator using a seven-gene signature plus Gleason score and taking cell type heterogeneity into account. *PloS one* **7**, e45178, doi:10.1371/journal.pone.0045178 (2012).

37   Sboner, A. *et al.* Molecular sampling of prostate cancer: a dilemma for predicting disease progression. *BMC medical genomics* **3**, 8, doi:10.1186/1755-8794-3-8 (2010).

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

39   The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011-1025, doi:10.1016/j.cell.2015.10.025 (2015).

40   *The Cancer Genome Atlas (TCGA),* http://www.cbioportal.org/study?id=prad_tcga - clinical *Accessed: 05.10.2016.*

41   Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J Stat Softw* **36**, 1-48 (2010).

42   Therneau, T. *A Package for Survival Analysis in S. R package version 2.38, URL:* https://cran.r-project.org/package=survival (2015).

43   Crager, M. R. Generalizing the standardized hazard ratio to multivariate proportional hazards regression, with an application to clinical~ genomic studies. *Journal of Applied Statistics* **39**, 399-417 (2012).

Supplementary to:

SFRP4 gene expression is increased in aggressive prostate cancer

Elise Sandsmark[1], Maria K. Andersen[1], Anna M. Bofin[2], Helena Bertilsson[3,4], Finn Drabløs[4], Tone F. Bathen[1], Morten B. Rye[4,5], May-Britt Tessem[1]

1. Department of Circulation and Medical Imaging, Faculty of Medicine and Health Sciences, NTNU - Norwegian University of Science and Technology, Trondheim, Norway
2. Department of Laboratory Medicine, Children's and Women's Health, Faculty of Medicine and Health Sciences, NTNU - Norwegian University of Science and Technology, Trondheim, Norway
3. Department of Urology, St. Olav's Hospital, Trondheim University Hospital, Trondheim Norway
4. Department of Cancer Research and Molecular Medicine, Faculty of Medicine and Health Sciences, NTNU - Norwegian University of Science and Technology, Trondheim, Norway
5. Clinic of Surgery, St. Olav's Hospital, Trondheim University Hospital, Trondheim, Norway

Supplementary Table S1. Correlations between citrate and spermine concentrations and gene expression of the genes in the NCWP-EMT gene expression signature.

Gene	Citrate		Spermine	
	Pearson's ρ	p-value	Pearson's ρ	P-value
SFRP4	-0.533	<0.001	-0.494	<0.001
FZD2	-0.421	<0.001	-0.350	<0.001
SFRP2	-0.354	<0.001	-0.31	0.002
LEF1	-0.35	0.002	-0.296	0.004
PLCB2	-0.343	<0.001	-0.265	0.01
CDH11	-0.342	<0.001	-0.258	0.012
CDH2	-0.339	<0.001	-0.236	0.021
SFRP1	-0.324	<0.001	-0.343	<0.001
FYN	-0.297	0.003	-0.246	0.008
VIM	-0.281	0.006	-0.222	0.03
NKD2	-0.270	0.008	-0.235	0.022
TCF4	-0.265	0.009	-0.225	0.028
MMP9	-0.185	0.073	-0.118	0.257
CDH3	-0.071	0.495	-0.155	0.133
WNT5A	0.051	0.627	0.067	0.521

Supplementary Table S2. SFRP4 immunohistochemistry evaluation, Gleason score, follow-up and metabolite concentrations of the samples/patients in the IHC cohort.

Patient	Immunohistochemistry SFRP4			Gleason score	Biochemical recurrence		Metabolites (mmol/kg wet weight)	
	Intensity	Percentage	Staining index		Status	Time (months)	Citrate	Spermine
1	1.00	2.00	2.00	3+4=7	1	27.57	1.16	0.09
2	1.00	3.00	3.00	4+3=7	1	14.59	2.51	0.54
3	2.00	3.00	6.00	3+4=7	1	1.44	5.03	0.52
4	2.00	2.00	4.00	4+3=7	1	32.43	7.98	0.62
5	ND	ND	ND	3+4=7	ND	ND	4.04	0.44
6	1.00	2.00	2.00	3+4=7	ND	ND	11.61	0.82
7	0.00	0.00	0.00	4+3=7	0	28.30	1.95	0.37
8	0.00	0.00	0.00	3+4=7	0	82.89	8.04	0.78
9	1.00	3.00	3.00	3+3=6	0	82.95	2.97	0.37
10	3.00	2.00	6.00	4+5=9	0	81.90	13.17	0.98
11	1.00	3.00	3.00	3+4=7	0	84.30	7.11	0.57
12	0.00	0.00	0.00	4+3=7	0	83.18	19.48	1.79
13	1.00	2.00	2.00	4+5=9	1	16.03	1.25	0.13
14	ND	ND	ND	3+4=7	0	82.23	11.17	1.11
15	0.00	0.00	0.00	3+4=7	0	68.92	6.98	0.60
16	1.00	3.00	3.00	4+4=8	1	31.31	5.24	0.43
17	1.00	2.00	2.00	3+3=6	0	63.67	6.05	0.38
18	ND	ND	ND	3+3=6	0	80.39	13.89	1.16
19	1.00	2.00	2.00	4+4=8	1	7.44	4.69	0.30
20	2.00	3.00	6.00	3+4=7	0	59.25	6.56	0.71
21	1.00	2.00	2.00	4+4=8	1	3.21	2.22	0.24
22	ND	ND	ND	3+4=7	0	72.00	5.33	0.62
23	1.00	1.00	1.00	3+4=7	1	43.05	3.91	0.52
24	2.00	2.00	4.00	3+4=7	0	71.21	3.88	0.37
25	1.00	2.00	2.00	4+3=7	0	71.80	11.67	0.71
26	1.00	3.00	3.00	4+3=7	0	73.77	6.52	0.65
27	1.00	2.00	2.00	3+3=6	0	35.74	3.96	0.29
28	2.00	3.00	6.00	5+5=10	1	1.15	4.15	0.77
29	ND	ND	ND	4+3=7	0	71.84	16.26	2.09
30	1.00	3.00	3.00	3+4=7	ND	ND	0.77	0.00
31	2.00	2.00	4.00	4+4=8	1	53.11	2.22	0.38
32	1.00	2.00	2.00	5+4=9	1	32.33	4.32	0.71
33	ND	ND	ND	3+3=6	0	72.30	8.93	1.31
34	0.00	0.00	0.00	3+4=7	1	1.61	2.85	0.13
35	1.00	2.00	2.00	3+4=7	0	72.85	1.91	0.08
36	0.00	0.00	0.00	4+4=8	1	1.28	1.17	0.12
37	2.00	3.00	6.00	4+5=9	1	62.03	8.74	1.40
38	1.00	2.00	2.00	3+4=7	0	70.07	2.06	0.52
39	ND	ND	ND	4+3=7	0	64.07	8.25	0.81
40	1.00	1.00	1.00	4+3=7	1	55.31	3.69	0.39

ND – no data/excluded. For immunohistochemistry, samples were excluded because of low tumour content. For biochemical recurrence data, some patients were excluded due to lack of follow-up data. Gleason score represents the samples, not the patients. Metabolite data were quantified with LCModel.

Supplementary Table S3. Scoring of SFRP4 immunohistochemistry.

SI score = Intensity * Index for percentage positive cells				
Staining intensity Cytoplasmic (highest intensity)	**0** No detectable signal	**1** (weak signal seen only at intermediate to high power)	**2** (moderate signal seen at low to intermediate power)	**3** (strongest signal seen at low power)
Percentage of positive cancer cells	**0** (<1%)	**1** (<10 %)	**2** (10-50 %)	**3** (>50 %)

Supplementary Table S4. Overview of the gene expression cohorts.

Cohort	Access number	Gene expression method	Cancer samples	Normal samples	Follow-up Endpoint
Main cohort	E-MTAB-1021	Microarray, Illumina HT 12v4	RP	Same patients	BCR
TCGA-PRAD	TCGA PRAD	RNA Sequencing	RP	Same patients	BCR
CAM Ross-Adams et al.	GSE70768	Microarray, Illumina HT 12v4	RP	Matched benign tissue	BCR or salvage treatment
STK Ross-Adams et al.	GSE70769	Microarray, Illumina HT 12v4	RP	-	BCR or salvage treatment
Wang et al.	GSE8218	Microarray, Affymetrix gene chips U133A	RP	4 Autopsy, biopsies smaller.	BCR
Sboner et al.	GSE16560	Microarray, Illumina DASL Assay	TURP from watchful waiting cohort	-	Prostate cancer specific death
Taylor et al.	GSE21034	Microarray, Affymetrix Human Exon 1.0 ST	RP	From RP of PCa patients	BCR
Mortensen et al.	GSE46602	Microarray, Affymetrix U133 Plus 2.0	RP	Surgical specimens of prostate from cystectomy of bladder cancer patients	BCR
Erho et al.	GSE46691	Microarray, Affymetrix Human Exon 1.0 ST GeneChips	RP	-	Metastatic progression

RP – Radical prostatectomy, BCR – biochemical recurrence, TURP – Transurethral resection of the prostate