

Department of Oncology-Pathology
Cancer Center Karolinska
Karolinska Institutet, Stockholm, Sweden

**Identification, validation and clinical
application of a three-gene signature for
accurate prognosis prediction and treatment
selection of newly diagnosed prostate cancer**

Zhuochun Peng

彭卓醇



**Karolinska
Institutet**

Stockholm 2015

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by AJ E-print AB, Stockholm, Sweden

© Zhuochun Peng, 2015

zhuo-chun.peng@ki.se

ISBN 978-91-7676-064-2

Identification, validation and clinical application of a three-gene signature for accurate prognosis prediction and treatment selection of newly diagnosed prostate cancer

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Zhuochun Peng

彭卓醇

Principal Supervisor:

M.D., Ph.D., Chunde Li
Karolinska Institutet
Department of Oncology-Pathology
Cancer Center Karolinska

Co-supervisor(s):

Professor Sten Nilsson
Karolinska Institutet
Department of Oncology-Pathology
Cancer Center Karolinska

Professor Monica Nistér
Karolinska Institutet
Department of Oncology-Pathology
Cancer Center Karolinska

Opponent:

Professor Jack A. Schalken
Radboud University Nijmegen
Department of Medical Center

Examination Board:

Professor Annika Lindblom
Karolinska Institutet
Department of Molecular Medicine and Surgery

Associate Prof. Weng-Onn Lui
Karolinska Institutet
Department of Oncology-Pathology
Cancer Center Karolinska

Associate Prof. Annika Håkansson
Linköping University
Department of Clinical and Experimental
Medicine (IKE), Clinical Oncology

ABSTRACT

This thesis presents a complete, comprehensive and stepwise approach of translational molecular research. Starting from whole-genome bioinformatics analyses based on the embryonic stem (ES) cell hypothesis, a three-gene signature was identified, validated with the goal of clinical application in order to optimize treatment decision based on improvement of overall survival estimation.

In Paper I, we hypothesized that gene signatures of embryonic stem (ES) cells may have prominent importance to determine the tumor subtypes and may be associated with the prognosis of various cancers including prostate cancer (PCa). Using published microarray datasets, 641 embryonic stem cell gene predictors (ESCGPs) were identified. Using gene expression patterns of these 641 ESCGPs tumor subtypes of different cancers can be stratified, particularly for prostate cancer. We further analyzed the gene expression levels of selected ESCGP genes in fresh-frozen fine needle aspiration biopsy samples taken from a Swedish cohort of 189 prostate cancer patients. The registry follow-up period for these patients was up to 18 years, where 97.9% patients had overall and cancer-specific survival data. As a result, a three-gene signature (VGLL3, IGFBP3 and F3) was identified sufficient to categorize the patients into high-risk, intermediate-risk and low-risk subtypes directly correlated with the overall and cancer-specific survival.

Currently, formalin-fixed paraffin embedded (FFPE) prostate core needle biopsy material is the most common sample material available in clinical practice, on which Gleason grading for prostate cancer diagnosis is usually conducted. Since each patient typically has multiple biopsy samples, and since Gleason grading is an operator dependent procedure known to be difficult, the impact of the operator's choice of biopsy was evaluated in paper II. We analyzed expression levels of the three-gene signature identified in paper I, using a four multiplex one-step RT-qPCR kit specially designed and optimized for measuring the three-gene expression signature in 127 FFPE prostate core needle biopsy samples taken from 43 patients. Our results show that the assessment of expression levels of two highly expressed genes (IGFBP3 and F3) in prostate cancer tissue is independent of Gleason patterns. These findings indicate that the impact of operator's choice of biopsy is low.

In paper III, we carried out a new cohort study including 241 prostate cancer patients with 6-9 years of registry follow up in order to verify the prognostic value of the three-gene expression signature in FFPE prostate core needle biopsy tissue samples. The cohort consisted of four patient groups with different survival times and cause of death (COD). We observed that supplementing readily available clinical data with gene expression levels of IGFBP3 and F3 in FFPE PCa biopsy tissues could improve survival prediction for PCa patients at time of diagnosis.

Based on the above work, a so-called Prostatype test system has been industrially designed and developed for clinical application. It integrates a robust multiplex RT-qPCR kit to measure expression levels of the three-gene signature and, a database of reference patients

with accurate clinical documentation using a kNN-algorithm called CPMA (Classification of Prostatic Malignancy Algorithm). The survival prediction in relation to different treatment modalities can greatly assist both clinicians and patients to make an individualized treatment decision.

The flowchart in Figure 1 summarizes the present thesis.

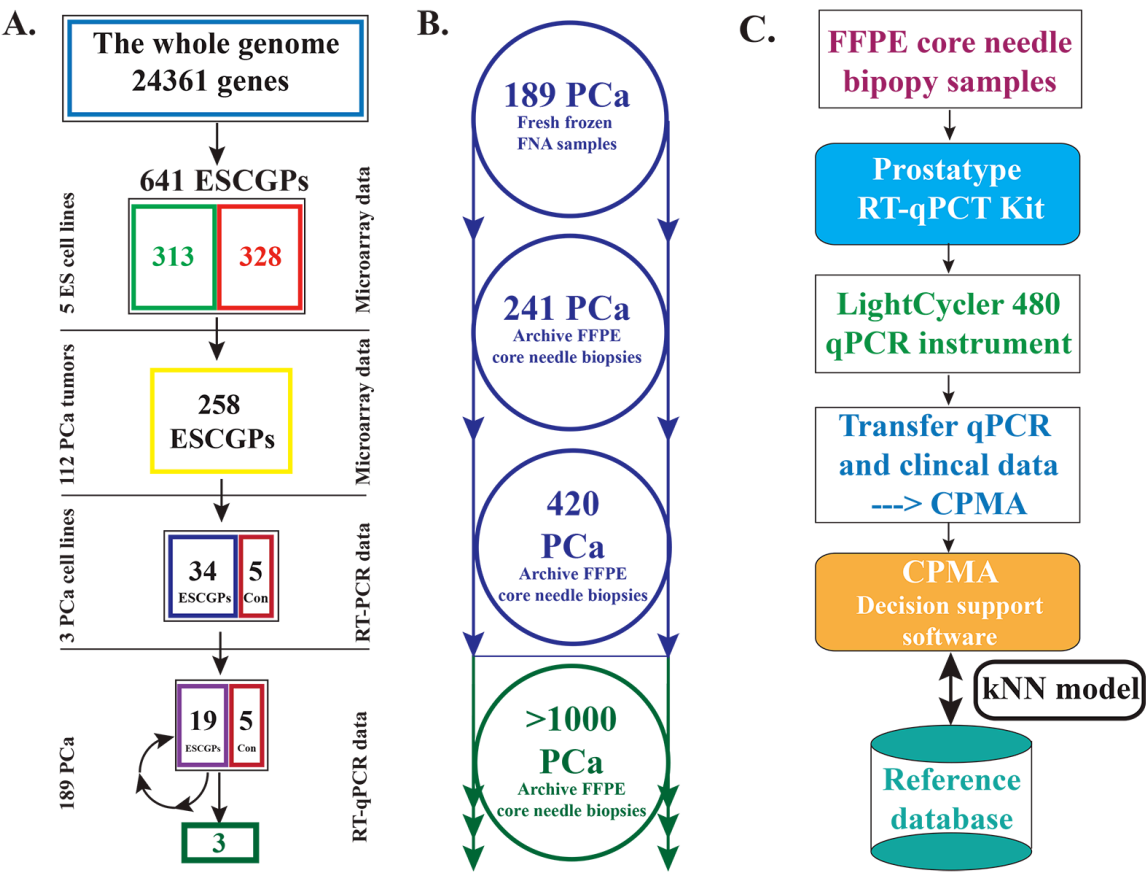


Figure 1. Graphic summary of the present thesis. A. A step-wise gene selection process, starting from bioinformatics analyses of the whole genome expression data of 24361 genes derived from 5 human embryonic stem cell lines, identified 641 Embryonic stem cell gene predictors (ESCGPs), until the identification of a three-gene signature. (Right vertical description: methodologies used in each step; left vertical description: materials used in each step of study). B. Cohorts represented in the order in which they were used for the pilot study (Paper I), the validation study (Paper II, III) and studies for filling CPMA reference database. The blue cycles are ready completed and ongoing cohort studies; the green cycle shows cohort studies planned in the near future. C. An industrially developed Prostatype Test System works according to the presented workflow. The system is composed of two parts: Prostatype RT-qPCR kit and CPMA software. It can provide prognostic statements assisting individualized/personalized estimation of overall survival time and decision of treatment for newly diagnosed prostate cancer patients.

LIST OF SCIENTIFIC PAPERS

- I. **Zhuochun Peng**, Lambert Skoog, Henrik Hellborg, Gustaf Jonstam, Inga-lill Wingmo, Marie Hjälm-Eriksson, Ulrika Harmenberg, Gabriella Cohn Cedermarck, Karl Andersson, Lars Ährlund-Richter, Setia Pramana, Yudi Pawitan, Monica Nistér, Sten Nilsson and Chunde Li.
An Expression Signature at Diagnosis to Estimate Prostate Cancer Patients Overall Survival. Prostate Cancer Prostatic Dis. 2014 Mar;17(1):81-90.
- II. **Zhuochun Peng**, Karl Andersson, Johan Lindholm, Inger Bodin, Setia Pramana, Yudi Pawitan, Monica Nistér, Sten Nilsson and Chunde Li.
Operator dependent choice of prostate cancer biopsy has limited impact on a gene signature analysis for the highly expressed genes IGFBP3 and F3 in prostate cancer epithelial cells. PLoS One. 2014 Oct 8;9(10):e109610.
- III. **Zhuochun Peng**, Karl Andersson, Johan Lindholm, Olga Hrydziusko, Setia Pramana, Yudi Pawitan, Monica Nistér, Sten Nilsson and Chunde Li.
Improving the prediction of prostate cancer overall survival by supplementing readily available clinical data with gene expression levels of IGFBP3 and F3 in formalin-fixed paraffin embedded core needle biopsy material. (Submitted)

TABLE OF CONTENTS

Introduction	1
1 Prostate cancer- a major clinical challenge.....	1
1.1 Discrepancy between incidence and mortality of prostate cancer	1
1.2 Heterogeneity of prostate cancer.....	3
2 Prosate cancer diagnosis.....	3
2.1 Gleason score (GS).....	4
2.2 Prostate cancer staging.....	5
3 Prostate cancer prognosis	6
3.1 Tumor risk factors	6
3.2 Patient risk factors	8
3.3 Life expectancy estimation	8
4 Prostate cancer treatments, clinical outcomes and endpoints	8
4.1 ADT.....	9
4.2 Curative treatment.....	9
4.3 AS	9
4.4 Clinical outcomes and endpoints.....	10
5 Inaccurate prognosis prediction.....	11
6 Swedish patient cohorts.....	11
7 Cancer biomarkers	12
7.1 PSA.....	12
7.2 Genomic biomarkers.....	13
7.3 Prognostic biomarkers.....	14
8 Embryonic stem (ES) cells	15
9 Sample type.....	16
10 Sample collection.....	17
10.1 Digital scanning	17
10.2 Automated sample collecting system	18
11 Prostatype Test System.....	18
11.1 Prostatype RT-qPCR kit	20
11.2 CPMA	20
Aim of the thesis	24
Results and discussion.....	25
Future studies and perspective	31
Acknowledgements.....	33
12 References	37

LIST OF ABBREVIATIONS

PCa	Prostate cancer
PSA	Prostate-specific antigen
DRE	Digital rectal exam
PAD	Pathological anatomical diagnosis
GS	Gleason Score
HGPIN	High grade prostatic intraepithelial neoplasia
ISUP	International society of urological pathology
TNM	Tumor, lymph node, metastasis staging system
AJCC	The American Joint Committee on Cancer
CT	Computerized tomography
MRI	Magnetic resonance imaging
LE	Life expectancy
NCCN	National Comprehensive Cancer Network
EAU	European Association of Urology
US SSA	US Social Security Administration
AUC	Area under the curve
COD	Cause of death
RCC	Regional Cancer Center
BPH	Benign prostatic hyperplasia
PHI	Prostate health index
p2PSA	Prostate-specific antigen isoform [-2] proPSA
PCA3	Prostate cancer antigen 3
AS	Active surveillance
RP	Radical prostatectomy
PLND	Pelvic lymph node dissection
RT	Radiation therapy
EBRT	External beam radiation therapy
ADT	Androgen deprivation therapy
HDR	High dose rate
LHRH	Luteinizing hormone-releasing hormone
CRPC	Castration resistant prostate cancer
mCRPC	Metastatic castration resistant prostate cancer
BCR	Biochemical recurrence
PIVOT	Prostate Cancer Intervention Versus Observation Trial
SPCG-4	Prostate Cancer Group Study Number 4

TMPRSS2	Transmembrane protease serine 2
ERG	v-ets erythroblastosis virus E26 homolog (avian)
PTEN	Phosphatase and tensin homolog on chromosome 10
EVT1	Ets variant 1(Homo sapiens)
FFPE	Formalin fixed paraffin embedded
GPS	Genomic Prostatic Score
CAPRA	Cancer of the Prostate Risk Assessment
CCP	Cell cycle progression
ES cell	Embryonic stem cell
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
iPS cell	Induced pluripotent stem cell
ESCGP	Embryonic stem cell gene predictor
H&E	Hematoxylin and eosin
kNN algorithm	K-nearest neighbor algorithm

INTRODUCTION

1 PROSTATE CANCER- A MAJOR CLINICAL CHALLENGE

The prostate gland fulfills important functions in male reproductive system. Despite its small size in proportion to the whole body, malignant disease affecting this organ, prostate cancer (PCa), is the most common type of cancer in adult men. Since the introduction of the Prostate-specific antigen (PSA) test and ultrasound-guided multicore prostate biopsy, PCa is now easily detected and diagnosed at early stage. However, it is a major clinical challenge to decide whether a patient should be radically treated or not. Because some PCa tumors progress rapidly to life-threatening conditions, while the majority of early detected PCa are less aggressive and not life-threatening. However radical treatments often lead to life-long complications or severe side effects.

1.1 DISCREPANCY BETWEEN INCIDENCE AND MORTALITY OF PROSTATE CANCER

Over the past two decades, the incidence of PCa in developed countries has dramatically increased, while mortality rates remain largely unaltered (Figure 2) (1). In Sweden, for example, the number of newly diagnosed PCa patients increased from 4 000 up to 10 000 from 1970 to 2010, while maintaining approximately 2 400 deaths annually (Figure 3).

Per 100 000 men

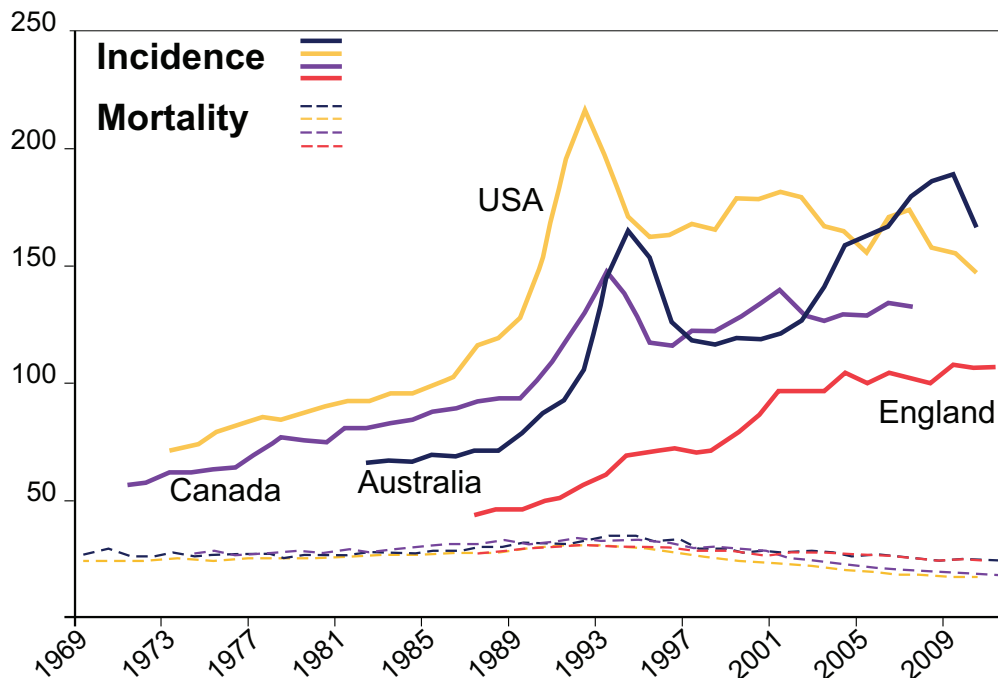


Figure 2. Reported age-standardized prostate cancer incidence and mortality rates per 100 000 men in USA, Canada, England and Australia. Feletto E, Bang A, Cole-Clark D, Chalasani V, Rasiyah K, Smith DP. An examination of prostate cancer trends in Australia, England, Canada and USA: Is the Australian death rate too high? World journal of urology. 2015 Feb 20. PubMed PMID: 25698456.

This is mainly due to the factor that prostate-specific antigen (PSA) was introduced as a first blood biomarker in the 1990s and has been used in the PCa screening since then. A review study reported a dramatically increased number of diagnosed PCa cases in most of developed countries, particularly in USA, Canada and Australia, with so-called ‘PSA-screening peaks’ (Figure 2) (1). In Europe, England and Sweden, similar increasing patterns were observed although high detection peaks were occurred slightly later (Figure 2 and Figure 3). PSA screening tests have led to a higher detection rate of early-stage PCa patients including those with aggressively growing tumors. These patients with rapidly growing tumors benefit from early detection and treatment initiation. However, the PSA test also detects a large number of patients with indolent tumors due to its limited specificity to distinguish aggressive tumors at early stage from indolent ones (2, 3).

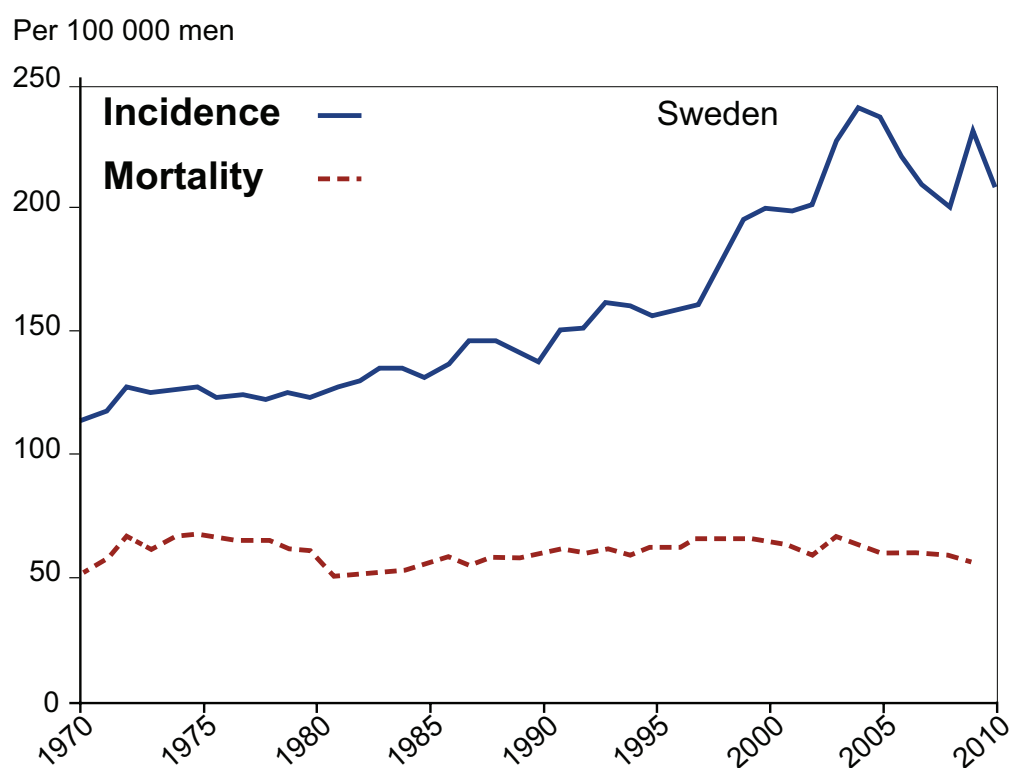


Figure 3. Age-standardized PCa incidence and mortality rates per 100 000 men in Sweden. Swedish data recourse: the National Board of Health and Welfare, www.socialstyrelsen.se

Based on these observations, two big nationwide studies investigating the benefit from PSA screening test were carried out in the first decade of the 21st century. However, the recently reported results of these two studies with long-term follow-up suggested, that the survival benefit due to the PSA screening for PCa diagnosis is controversial (4, 5). Even though European and Swedish data show that the PSA screening has substantial and significant benefits regarding PCa-specific mortality after 13 years of follow-up, there was a clear

evidence that over-diagnosis and overtreatment are major adverse effects of PSA screening (6).

1.2 HETEROGENEITY OF PROSTATE CANCER

PCa is known for its heterogeneity in epidemiological, genetic, pathological, biological and clinical observations. The above big discrepancy between incidence and mortality rates of PCa is a typical example.

Clinically, a group of PCa tumors could rapidly progress to metastatic disease and becoming lethal, while fairly large numbers of PCa tumors often grow slowly and indolently. As a major treatment for clinically advanced and metastatic PCa, hormone treatment or androgen deprivation therapy (ADT) can almost achieve complete biochemical and clinical responses for most patients but castration resistance would be eventually developed. Metastatic castration resistant prostate cancer (mCRPC) is the major cause of cancer-specific death. However, the time of progressing mCRPC varies dramatically in different patients with a range of from a few months up to several years. Similarly there is the pathological heterogeneity, represented by different GS in different patients (or prostates) and different Gleason grading patterns in coexisting multiple foci in the same prostate.

It has been speculated that the above clinical and pathological heterogeneity is determined by the profound genetic or genomic heterogeneity. Previous researches, including the whole genome analyses by gene expression microarrays (7) and next generation sequencing, have identified different constitutional genetic loci, chromosomal rearrangements (8-10), gene expression and mutation patterns being associated with different clinical and pathological outcomes (11-13). Therefore, molecular biomarkers might have the potential to fundamentally distinguish various tumor subtypes with different biological aggressiveness that consequentially determining different clinical outcomes, thus further improving the prognosis estimation accuracy and treatment decision.

Clearly this heterogeneity between tumors in different patients presents the urgent need for the identification and application of prognostic and predictive biomarkers. However, the heterogeneity of cancer cells within the same patient might cause difficulties for the selection of representative samples for the process to develop a clinically applicable biomarker test. That is to say, a clinically applicable and reliable biomarker test should be able to confront and solve the inconvenience caused by the cancer cell heterogeneity derived from the same patient (prostate), also able to distinguish between different tumor subtypes existing in different patients (prostates).

2 PROSTATE CANCER DIAGNOSIS

Prostate cancer develops from the prostate glandular epithelium. The glandular cells produce PSA, a protein with protease function, which constitutes an important component in the semen. Normally, PSA does not enter the blood stream. Under disease conditions however, and particularly in prostate cancer, PSA can enter the blood circulation and is thereby

detectable in blood samples. This is primarily due to the disruption of local blood-tissue barriers by cancer cell invasion or metastasis. In the clinical routine, an abnormally increased PSA value or a suspicious digital rectal exam (DRE) result is usually indicative for prostate core needle biopsies to confirm the existence of malignant cells and thus the diagnosis of PCa (14). With seldom exceptions of patients with significantly high PSA and metastatic disease, the diagnosis of PCa, in particular for patients with localized disease, the diagnosis of PCa can only be made or confirmed by histopathological diagnosis presented as a pathological anatomical diagnosis (PAD) report.

2.1 GLEASON SCORE (GS)

Based on the histopathological morphology of cancer tissue in biopsies, a Gleason grading can be conducted as the Gleason score (GS) reflecting the aggressiveness of a tumor. The GS is composed of two values: (1) the primary Gleason pattern representing the most predominant pattern observed in core needle biopsies, and (2) the secondary Gleason pattern referring to the secondary most common pattern (15).

The Gleason grading system was firstly invented by Dr. Donald F. Gleason in 1966 to assign prostatic carcinoma different grades of severity, based on the histopathological morphology pattern of tumor under fairly low magnification microscopic examination (16). Thus, Gleason grading is focused on the morphological structures of the cancer tissue instead of grading the dedifferentiation of individual cancer cells. The Gleason grading system classifies tumors into five grades ranging from grade 1 to 5 with increasing tumor aggressiveness as compared to normal prostatic glandular architectures. In order to solve the inconvenience caused by the histopathological heterogeneity within the same prostate, Gleason score (GS) is composed of two grades. The predominant pattern is identified as the primary Gleason grade and the second common pattern as the secondary Gleason grade, and the GS is generated as the sum of the primary grade plus the secondary grade. For example, a tumor with a primary Gleason pattern of grade 3 and the secondary Gleason pattern of grade 4 would equal, $GS=3+4=7$.

Since the 1970s the Gleason grading system has been the gold standard when diagnosing prostatic carcinoma, however changes of many factors over the years have affected the grading results. For instance, due to early detection, many patients are asymptomatic at time of diagnosis. Consequently, high GS is much less common in current patients, instead a large number of patients have a GS of 6 or 7 with less distinguishing precision. Since the 1980s, thinner 18-gauge needles have been used to take sextant multiple biopsies in contrast to the original thicker and less number of core biopsies on which the Gleason grading system was developed. With the introduction of immunohistochemical staining of prostatic basal cell markers, Gleason grade 1 is now only classified as atypical adenomatous hyperplasia instead of carcinoma, and the cribriform high grade prostatic intraepithelial neoplasia (HGPIN) should be classified as the cribriform like ductal adenocarcinoma. Thus, the international society of urological pathology (ISUP) had the consensus meetings in 2005 in order to modify and update Gleason grading system trying to adapt to the current clinical situation. Consequently some modifications and adjustments have been made (Figure 4) (15).

Despite the above efforts, there is still a significant number of cases showing mis-grading or up/under-grading of GS due to the fact that the Gleason grading system is highly operator-dependent (17). For genomic biomarker tests which highly rely on the Gleason patterns, it is practically problematic to select the representative sample from multiple core biopsies. Consequently, it needs to be evaluated whether the significance of biomarkers is dependent on the GS derived from different patients (prostates), and if the significance is dependent on the Gleason pattern of cancerous cells taken from the same patient (prostate). In paper II and III, we addressed those questions. Ultimately, an operator independent pathological grading system, e.g. a digital imaging-based algorithm that could automatically generate a better and more objective GS, would be a promising diagnostic tool that could aid pathologists in improving diagnostic results in the future.

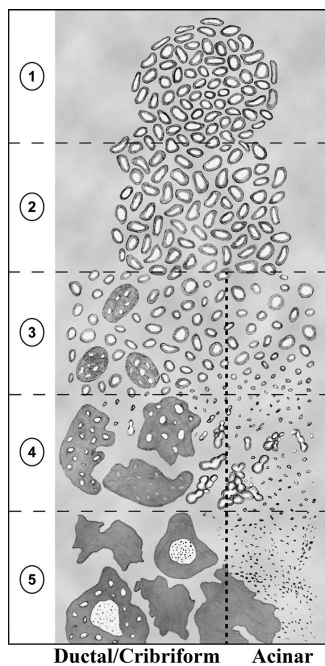


Figure 4. Schematic diagram of modified Gleason grading system. Epstein JI, Allsbrook WC, Jr., Amin MB, Egevad LL, Committee IG. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. The American journal of surgical pathology. 2005 Sep;29(9):1228-42. PubMed PMID: 16096414.

2.2 PROSTATE CANCER STAGING

Apart from PCa diagnosis, disease staging also plays an important role in determining treatment decision-making. There are different staging modalities to define and present the extent of local invasion and distant metastasis, such as clinical staging by DRE, imaging staging by ultrasound, magnetic resonance imaging (MRI) and bone scan to distinguish between localized, locally advanced and metastatic disease, pathological staging of radically removed prostate to assist the identification of capsule or seminal vesicle invasion, or positive surgical margin.

The TNM staging system proposed by the American Joint Committee on Cancer (AJCC) is the most frequently used clinical and pathologic staging system (Table 1) (18). Bone scan, computerized tomography (CT) or MRI is required to confirm potential metastasis of PCa tumors.

<p>Primary Tumor (T)</p> <p>CLINICAL</p> <p>TX Primary tumor cannot be assessed</p> <p>T0 No evidence of primary tumor</p> <p>T1 Clinically inapparent tumor neither palpable nor visible by imaging</p> <p>T1a Tumor incidental histologic finding in 5% or less of tissue resected</p> <p>T1b Tumor incidental histologic finding in more than 5% of tissue resected</p> <p>T1c Tumor identified by needle biopsy (for example, because of elevated PSA)</p> <p>T2 Tumor confined within prostate¹</p> <p>T2a Tumor involves one-half of one lobe or less</p> <p>T2b Tumor involves more than one-half of one lobe but not both lobes</p> <p>T2c Tumor involves both lobes</p> <p>T3 Tumor extends through the prostate capsule²</p> <p>T3a Extracapsular extension (unilateral or bilateral)</p> <p>T3b Tumor invades seminal vesicle(s)</p> <p>T4 Tumor is fixed or invades adjacent structures other than seminal vesicles, such as external sphincter, rectum, bladder, levator muscles, and/or pelvic wall</p>	<p>Pathologic (pT)</p> <p>pT2 Organ confined</p> <p>pT2a Unilateral, one-half of one side or less</p> <p>pT2b Unilateral, involving more than one-half of side but not both sides</p> <p>pT2c Bilateral disease</p> <p>pT3 Extraprostatic extension</p> <p>pT3a Extraprostatic extension or microscopic invasion of bladder neck³</p> <p>pT3b Seminal vesicle invasion</p> <p>pT4 Invasion of rectum, levator muscles, and/or pelvic wall</p>
<p>Regional Lymph Nodes (N)</p> <p>CLINICAL</p> <p>NX Regional lymph nodes were not assessed</p> <p>N0 No regional lymph node metastasis</p> <p>N1 Metastasis in regional lymph node(s)</p>	<p>Regional Lymph Nodes (N)</p> <p>PATHOLOGIC</p> <p>pNX Regional nodes not sampled</p> <p>pN0 No positive regional nodes</p> <p>pN1 Metastases in regional node(s)</p>
<p>Distant Metastasis (M)</p> <p>M0 No distant metastasis</p> <p>M1 Distant metastasis</p> <p>M1a Nonregional lymph node(s)</p> <p>M1b Bone(s)</p> <p>M1c Other site(s) with or without bone disease</p>	
<p>Notes</p> <p>¹ Tumor found in one or both lobes by needle biopsy, but not palpable or reliably visible by imaging, is classified as T1c.</p> <p>² Invasion into the prostatic apex or into (but not beyond) the prostatic capsule is classified not as T3 but as T2.</p> <p>³ There is no pathologic T1 classification.</p> <p>⁴ Positive surgical margin should be indicated by an R1 descriptor (residual microscopic disease).</p> <p>⁵ When more than one site of metastasis is present, the most advanced category is used. pM1c is most advanced.</p>	

Table 1. PCa staging. Leslie H. Sobin (Editor) MKGE, Christian Wittekind (Editor). International Union Against Cancer (UICC): TNM Classification of Malignant Tumours, 7th Edition. 7th Edition ed: Wiley-Blackwell; 2009 November.

3 PROSTATE CANCER PROGNOSIS

Since prostate cancer is a heterogeneous disease with different outcomes, an accurate prediction of prognosis is critically important prior to making a treatment decision. Currently, such a prognosis is based on PSA value, GS, clinical stage, age and comorbidity status. Based on the clinical diagnosis workflow (NCCN 2014 updated version, Figure 5) (19), we find that the current clinical guidelines for PCa management are mainly based on the estimation of life expectancy (LE) and classification by tumor risk factors.

3.1 TUMOR RISK FACTORS

The first and still most widely used tumor risk classification system is the D'Amico classification system, which stratifies patients with different risks of biochemical recurrence after prostatectomy or external radiation therapy (20-22). The D'Amico risk system stratifies non-metastatic PCa patients into low-risk, intermediate-risk and high-risk subgroups based on initial PSA value, biopsy GS and clinical stage (22). The European Association of Urology (EAU) almost directly uses this system in the PCa management guidelines (Table 1) (19, 23).

For staging, PCa can be classified as localized (T1-2N0M0), locally advanced (T3-T4N0M0) and metastatic (anyTN1anyM or AnyTanyNM1).

	Low-risk	Intermediate-risk	High-risk	
Definition	PSA < 10 ng / mL and GS < 7 and cT1-2a	PSA 10-20 ng /mL or GS 7 or cT2b	PSA > 20 ng / mL or GS > 7 or cT2c	any PSA any GS cT3-4 or cN+
	Localised			Locally advanced

Table 1. The EAU guidelines for tumor risks definition, 2014 updated version. N. Mottet (Chair) JB, E. Briers (Patient Representative), R.C.N. van den Bergh (Guidelines Associate) MB, N.J. van Casteren (Guidelines Associate) PC, S. Culine,, S. Joniau TL, M.D. Mason, V. Matveev, H. van der Poel,, T.H. van der Kwast OR, T. Wiegel. EAU Guidelines on Prostate Cancer-updated March 2015. 2015. English.

The National Comprehensive Cancer Network (NCCN) guidelines classify tumors into very low, low, intermediate and high risk as clinically localized PCa, very high risk as locally advanced PCa, and metastatic PCa (Figure 5). The risk classification is based on the tumor risk factors as PSA value and density, GS, clinical stage, number of positive biopsies, cancerous area of core biopsies.

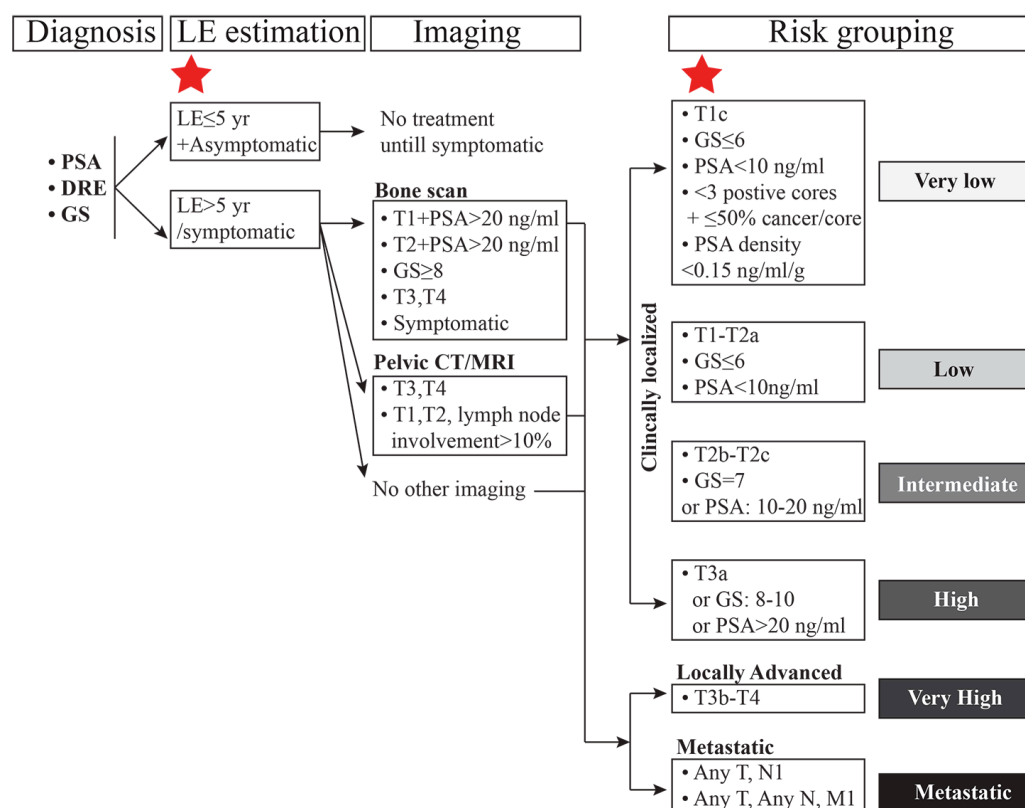


Figure 5. Clinical diagnosis flowchart for prostate cancer. (NCCN clinical practice guidelines for prostate cancer, updated 2014). Mohler JL, Kantoff PW, Armstrong AJ, Bahnson RR, Cohen M, D'Amico AV, et al. Prostate cancer, version 2.2014. Journal of the National Comprehensive Cancer Network : JNCCN. 2014 May;12(5):686-718. PubMed PMID: 24812137.

3.2 PATIENT RISK FACTORS

Besides the tumor risk factors, patient's risk factors also contribute to the mortality of PCa. Age of patient at diagnosis is one of the most predominant risk factors for PCa mortality (24). Additionally, patient's comorbidities and physical performance at diagnosis contribute to the non-prostate cancer-specific mortality (25, 26).

3.3 LIFE EXPECTANCY ESTIMATION

An estimation of life expectancy (LE) of newly diagnosed PCa patient determines the choice of primary treatment. In other words, LE estimation is crucial to determine whether the patient would benefit from active surveillance (AS) program, curative treatments such as prostatectomy and curative radiation therapy, or observation until symptoms for palliative treatments.

Due to the slow progression of most PCa, several studies suggest a cutoff of 10-year LE to distinguish between aggressive and non-aggressive tumors. The 10-year LE rule is still the golden threshold for clinicians to select the treatment for PCa patients. On the other hand, it is difficult to estimate LE with satisfactory accuracy for treatment decision. There are other factors beyond tumor risk factors that cause the complexity of LE estimation, among which patient risk factors such as "biological age" and health status. Alternative LE estimation models based on the age and comorbidity have been reported with improved prediction accuracy of overall survival, particularly for non-cancer-specific survival (27).

NCCN guidelines recommend the use of a nation-wide life-table analysis, such as the US Social Security Administration (SSA) life tables, when estimating LE (Figure 5 and 6). SSA life tables are defined according to national social security data representing a population-based general life expectancy. The estimation is fairly accurate (AUC=0.68 for 66 years old men), and independent of clinician's expertise or experience (28). Comparatively, some LE prediction statistic models adjusted by comorbidities were reported to have more personalized survival estimation (29, 30). However, there is not sufficient evidence that these models would be more accurate than the SSA life tables (31). As pointed out previously, most of these studies mainly focus on the cancer-specific mortality or the effects contributed by cancer. There are only a limited number of studies investigating effects or parameters contributing to the non-cancer-specific mortality despite the observation that there are large numbers of PCa patients died of other diseases or diseases related to treatments. Investigating the specific causes of death (COD) causing non-cancer mortality is important to identify potential human risk factors, which is however very difficult to quantify. Therefore, design a biomarker research that could possibly establish a simple method to improve the accuracy of individualized LE estimation based on common tumor and patient risk factors and molecular biomarkers.

4 PROSTATE CANCER TREATMENTS, CLINICAL OUTCOMES AND ENDPOINTS

The selection of different treatment modalities is based on a balanced consideration of tumor risk factors, patient risk factors and LE estimation. Based on the most updated evidence and in line with the NCCN guidelines, Figure 6 presents different treatment options in relation to LE and different risk groups (32, 33). In NCCN guidelines (Figure 6), LE is considered as the conditional factor prior to choosing treatment due to the fact that elder patients with more comorbidities often benefit much less from any further active treatments. Instead, observation until obvious symptoms for palliative treatment is recommended for patients with short LE.

4.1 ADT

As mentioned earlier, clinically advanced very high risk PCa and metastatic PCa often undergo palliative hormone treatments or androgen deprivation therapy (ADT). ADT could be accomplished by surgical castration or by pharmaceutical agents such as anti-androgens, LH-analogs or antagonists (chemical castration). Because the growth and proliferation of prostate cancer cells requires the androgens, such as testosterone, ADT can lead to growth arrest and apoptosis of PCa cells, and even stromal fibrosis of PCa tissues (34). Despite a rapid response at the beginning, most cases will eventually progress and become metastatic CRPC (mCRPC) with median time to castration resistance of about 18 months (35). For mCRPC patients, many palliative treatments show modestly prolonged survival such as next generation of hormone therapy, chemotherapy including Docetaxel and Cabazitaxel, radiopharmaceutical therapy (Radium 223), and cancer immunotherapy. It is noteworthy that recent studies reported that chemohormonal therapy (ADT + chemotherapy) as the first line treatment for high risk PCa as well as metastatic hormone sensitive PCa could significantly prolong survival time (from 44.0 to 57.6 months) (32, 33, 36, 37).

4.2 CURATIVE TREATMENT

Curative RP (including laparoscopic or robotic-assisted prostatectomy) is the dominant treatment offered for patients with clinically localized cancer of low and intermediate risk, despite its modest clinical survival benefit (38). Curative RT is an effective treatment when combined with neo-adjuvant and adjuvant ADT for patients with high risk and locally advanced cancer (39-43). It is even recommended as a treatment option for patients with localized cancer of low to intermediate risk. For these patients, RT may have the same curative effect but different profile of complications and side effects as compared with RP. A recent study reported that even elder PCa patients could benefit more from curative RP or RT (44). For curative treatment, being cured is the most favorable clinical outcome that can be defined as after a sufficient long time of follow-up e.g. 10 years, there is no sign of recurrence.

4.3 AS

The AS program is preferably recommended for very low risk and low risk PCa patients, and is an effective way of compensating the disadvantage of early detection of PCa resulting in overtreatment currently.

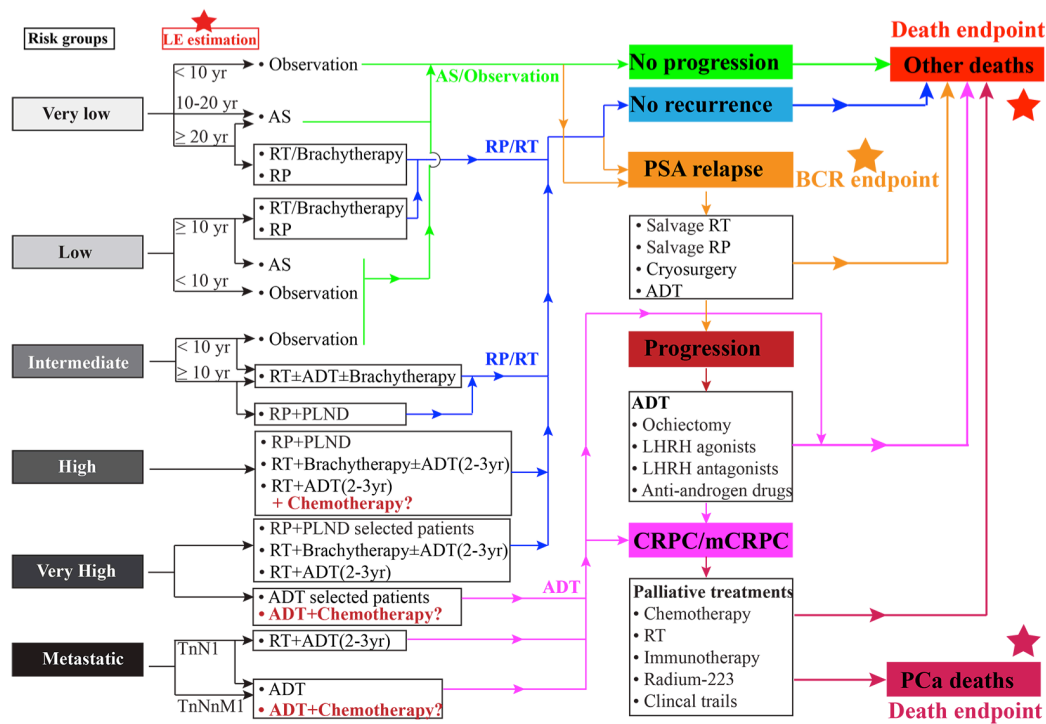


Figure 6. PCa management and clinical outcomes. RT: Radiation therapy, RP: Radical prostatectomy, AS: Active surveillance, ADT: Androgen deprivation therapy, PLND: Pelvic lymph node dissection, BCR: Biochemical recurrence, CRPC: Castration resistant prostate cancer, mCRPC: metastatic CRPC. Risk groups are adopted from the NCCN clinical practice guidelines for prostate cancer, updated 2014 (<http://www.nccn.org/>)

4.4 CLINICAL OUTCOMES AND ENDPOINTS

Figure 6 also presents different progression steps and outcomes of PCa after the diagnosis. No matter which treatment, death would be the final outcome caused either by PCa disease or by other diseases or events. For AS, the most desirable outcome is the absence of progression, and thus no need for any further intervention. After radical treatment, some patients can suffer from recurrence by initially presenting an increase of PSA only, which is defined as the biochemical recurrence (BCR) or PSA relapse. Using BCR event as the endpoint of follow-up in clinical studies, the survival analysis can be defined as BCR free survival analyses. After BCR, a proportion of patients (with range of approx.15%-50% varied in different cohorts) can still be cured by salvage RT (45). Uncured patients will receive ADT treatment; however, despite a rapid response at the beginning, most cases will eventually progress and become metastatic CRPC and mCRPC is the major cause of cancer death for PCa patients. This means that overall mortality or survival is the utmost clinical outcome endpoint for PCa clinical studies. Time from diagnosis to death (overall survival time) is the most important variable to measure the treatment effect.

Currently most PCa clinical studies, particularly those aiming at the identification of prognostic biomarkers, use surrogate endpoints such as adverse pathology, BCR or disease progression. These endpoints are clinically relevant but not like the utmost real clinical outcome endpoint: mortality with COD annotated and overall survival time. Thus biomarkers

identified based on surrogate endpoints would predict less relevant clinical outcomes instead of mortality with different CODs. The quality of life is gaining more attention than ever before since the radical treatments have become predominant in the past decades, and their severe side effects are dramatically affecting men's normal life for the rest of their lives postoperatively.

5 INACCURATE PROGNOSIS PREDICTION

Obviously, a significant proportion of patients with low risk and perhaps even intermediate risk cancer have been treated by radical treatments unnecessarily (46). As stated above, the majority of indolent PCa patients are affected radical overtreatment according to the Prostate Cancer Intervention Versus Observation Trial (PIVOT), a randomized clinical trial. The study points out that patients underwent RP do not benefit significantly from their treatment compared to the watchful waiting group during a 12-year follow-up period (38). Another large randomized clinical trial, the Scandinavian Prostate Cancer Group Study Number 4 (SPCG-4), indicates however, that death numbers are less in the RP group (200/347) compared to watchful waiting group (247/348) after 23 years of follow-up (47). Nevertheless, clearly the majority of PCa patients appear over-treated (200/247=80.9% of RP treated PCa patients would not benefit more from their treatment than watchful waiting).

These clinical challenges are fundamentally caused by inaccurate survival prediction at time of diagnosis, however the current clinical parameters just simply can not do more to get better prognosis prediction. A good cancer biomarker would offer a potential to further improve the current situation. The major reason of inaccurate prediction is that only using PSA, DRE and prostate biopsy, some actually high-risk cancers at early stage are mis-classified as low or intermediate-risk cancer, undergo AS program and thereby miss the best time window to be cured (48). It is not seldom that preoperative low GS and localized cancer becomes up-graded or up-staged postoperatively. Secondly, elder patients with high-risk cancer may be undertreated due to an inaccurate LE estimation (44). Furthermore, high-risk and locally advanced cancer needs extensive treatment more than just radical prostatectomy or local radiation (36). An improved prognosis prediction by integrating molecular biomarkers together with current common clinical parameters may lead to an optimized and individualized treatment selection, perhaps even for selection of suitable adjuvant hormone or chemotherapy in the near future.

6 SWEDISH PATIENT COHORTS

Sweden is a small country with a population of 9.5 million, the majority of them are Caucasian with a highly homogeneous ethnical background. Each person that is registered a resident in Sweden is assigned a unique personal number. By using this number, the clinical record of cancer patients is registered to one of the numerous nationwide cancer registries. These Regional Cancer Centers (RCC) are responsible for semiannual, regional registration and follow-up of clinical data of cancer patients. Each patient's clinical record can be accessed by each individual clinical unit (local clinics, big hospitals, other clinical related

organization, etc.) by using the patient's personal number. Thereby, the RCC has registered and automatically followed the clinical records for each cancer patient in this country for more than half a century. Moreover, there is a bio-bank law in Sweden that manages the patient samples. Assuming that patients have signed permits allowing the use of their material for current or future research, these samples could potentially contribute to many clinical studies. Every specimen shall be stored and kept in a proper way according to the bio-bank law.

The complete registry data from RCC enables many translational researches to access one of the best clinically followed cohorts in the world, and more feasible accessed and well-maintained human specimen bio-banks allow these studies could access to high-quality sample materials. Our studies, particularly the study in paper I, used fresh frozen fine needle aspiration (FNA) biopsy samples with up to 18 years clinically follow-up and nearly 90% of patients had been deceased until the endpoint. Consequently, these well documented and long-term followed clinical patient data allowed our study to access the ultimate clinical outcome: mortality with COD.

7 CANCER BIOMARKERS

The understanding of cancer at the molecular level has been deepened since the past decades, in particular due to the introduction of DNA, RNA and protein analyses in genome scale. It is generally accepted that cancer is a phenotype of genetic/genomic changes/mutations in the cells chromosomes (49). From phenotypic associations, molecular changes detected by whole genome analyses can be classified as: random mutations/changes, causative mutations, driving mutations/changes and disease associated mutations/changes. Except random mutations/changes, all other types have the potential to be further characterized as different types of biomarkers. Causative and driving mutations or changes can be further characterized as molecular targets for developing new effective treatments.

A cancer biomarker is defined as a substance detected from tissue, blood, or other body fluids, which might indicate a sign of cancer, a degree of tumor aggressiveness, or the effect of drug pharmacodynamics (49). Cancer biomarkers help to distinguish different tumor subtypes of different aggressiveness or treatment sensitivities, and are therefore called predictive and prognostic biomarkers. The identification and application of prognostic biomarkers at the time of diagnosis would guide and improve treatment decision making for cancer management.

7.1 PSA

The prostate specific antigen (PSA) was identified as the first diagnostic biomarker for PCa in the early 1990s, and is one of the prognostic biomarkers used for risk classification in current clinical guidelines. As a classical blood diagnostic marker, a serum total PSA value of 4.0ng/ml is the cutoff for PCa diagnosis (50). However PSA is prostate specific but not cancer specific, which leads to its limited specificity for diagnosing prostate carcinoma due to

hardly distinguishing from the PSA increase caused by benign prostatic hyperplasia (BPH), infection or chronic inflammation. The two major clinical trials reported recently that the potential of PSA screening for decreasing PCa mortality is becoming more controversial (4, 5).

As mentioned above, an increase of PSA value is one of mainstreamed biochemical recurrence biomarkers as the end-point of follow-up. It is a very convenient and practical biomarker for measuring the health status of the prostate gland since it can be assessed by a simple blood test. However there are some demerits to use an increase of PSA levels as the end-point for clinical studies in terms of LE estimation. Firstly PSA is only prostate-specific but not cancer-specific, which means the increase of PSA value does not necessarily related to a cancer event. Because its abnormal increased value in serum could possibly be caused by benign prostatic hyperplasia (BPH), infection or chronic inflammation. It is also known that poorly or non-differentiated prostate cancer doesn't produce much PSA. Furthermore, the specificity of the PSA test itself is limited. Its value is also easily affected by prostate volume, age of patient and the dedifferentiation degree of the cancer cells. Several optimized PSA tests such as the prostate health index (PHI) test measuring three forms of PSA: total PSA, free PSA and p2PSA, would be a promising solution to provide more specific results (51). However, PSA recurrence (BCR) as an endpoint surrogate for real survival is unreliable. As described in the part of outcome and endpoints, treatments and outcomes after BCR can vary dramatically in different patients, some patients with BCR after surgery can even be cured by salvage RT. Using final end-point of follow-up such as 'death of patient' to generate overall survival rates for investigating LE prediction is more clinically reliable since it is also considering mortality due to other diseases or treatment related effects.

7.2 GENOMIC BIOMARKERS

Also in the 1990s, PCA3 was identified and is considered as the most PCa specific biomarker. It is a segment of non-coding RNA transcribed by chromosome 9q21-22 and can distinguish benign conditions from prostatic carcinoma with > 90% accuracy. PCA3 can be easily measured in urine samples. Unlike PSA, PCA3 measurement is not affected by prostate volume, patient's age or any other prostatic disease such as prostatitis (52, 53). In 2012, a commercial PCA3 urine test was approved by the US Food Drugs Administration (FDA); currently the test is recommended in European PCa management guidelines for guiding re-biopsy after the initial negative biopsy (23, 54). Its prognostic value in monitoring tumor progression status for those patients undergoing active surveillance (AS) program is still requiring more evidence to be confirmed (55). The use of PCA3 in combination with other biomarkers might improve its prognostic significance.

A fusion protein consisting of transmembrane protease serine 2 (TMPRSS2) fused with the v-ets erythroblastosis virus E26 homolog (avian) (ERG) gene (TMPRSS2-ERG) can be detected in approximately 40-80% of PCa urine samples (56). Using TMPRSS2-ERG to aid PCa diagnose have been investigated intensively recently, but its prognostic value in distinguishing aggressive from non-aggressive PCa remains to be established. Interestingly in

a prospective study, in combination with the urine biomarker PCA3, TMPRSS2-ERG appears to have increased prediction accuracy to distinguish tumors with higher GS>7 or clinically more significant tumors according to the Epstein criteria (57). The Epstein criteria identify insignificant or more significant tumors based on PSA density, Gleason score, number of positive cores and percentage of cancer area in each positive core (58).

Deletion of a tumor suppressor gene, the phosphatase and tensin homolog on chromosome 10 (PTEN), is widely associated with poorly differentiated prostate tumors. Loss of PTEN is associated with higher Gleason grade, risk of tumor progression and recurrence after treatment (59). Deleted PTEN status is further associated with higher metastatic rates and cancer-specific mortality, particularly in combination with ERG/ETV1 rearrangement status (60).

7.3 PROGNOSTIC BIOMARKERS

A PCa predictive biomarker could guide and assist clinicians in the treatment decision-making. This has been obtaining more attention over the years due to increased numbers of early detected and clinically insignificant prostate tumors, resulting in a large number of over-treated patients. Moreover, due to the often uncertainty of choice making about preferable treatment for these early detected indolent PCa patients, better prognostic biomarkers providing a more accurate prediction as an improved supplement for current clinical standards are needed. Personalized genetic information extracted by biomarkers would aid clinicians to make treatment options with better survival prediction accuracy. Such biomarkers should be identified from studies with more relevant clinical end point and with longer period of follow-up, externally be validated and easily measured. There are a number of commercially available genetic prognostic tests that have been developed to guide treatment decision making for clinicians.

Oncotype DX is a gene signature expression test based on a panel of 17 genes. The test measures mRNA expression levels of this gene panel using FFPE prostate core needle biopsy samples. The test has developed a new score called ‘Genomic Prostatic Score (GPS)’, which stratifies indolent prostate cancer into very low-, low-, low-intermediate risks of subgroups with improved accuracy compared to the Cancer of the Prostate Risk Assessment (CAPRA) scores or the NCCN risks (61, 62). Both CAPRA and NCCN are scoring systems based on clinical parameters such as GS, pre-treatment PSA, clinical stage, proportion (%) of positive cores, and patient age at diagnosis (63). Thereby, the GPS provides genetic information in addition to conventional clinical parameters. A second study verified the prognostic value of the 17-gene signature in 402 patients including approximately 20% African American men. The test uses RT-qPCR, a standardized, precise and easily performed method to measure gene expression signatures. However the cohorts defined in both studies use BCR, i.e., increasing PSA levels after prostatectomy, as the follow-up endpoint, which is less clinically reliable in terms of overall or cancer-specific mortality.

The Prolaris gene test is a gene signature expression test measuring the mRNA levels of 31 cell cycle genes, also called the cell cycle progression (CCP) signature. Measuring the expression levels of CCP generates a CCP score to predict tumor progression status in terms of BCR occurrence following prostatectomy. The test has also been used to predict the mortality rate in a cohort of patients diagnosed by transurethral resection of the prostate (TURP) (64). A recent core needle biopsy cohort study with 558 patients shows that the CCP score is a strong predictor for PCa-specific 10-year mortality (65). In this study, the CCP score provides an independent additional prognostic value compared to other clinical scores such as CAPRA scores.

It is a promising genomic biomarker test to predict 10 years PCa-specific mortality rate due to its clinical relevant follow-up endpoint of cohort and external clinical studies validated. The test could provide better guidance for clinician if more validation studies confirm its prognostic value in the future. However the test system is rather complicated as it is based on a panel of 31 genes requiring more advanced statistical data analyses compared to CAPRA scores. Although the CCP score showed comparable results compared to CAPRA score, and using the CCP score in combination with the CAPRA score can help to generate further improved prediction accuracy instead of using any of alone. Additionally, the test can only be performed in certified laboratories owned by the provider. Consequently, pathologists would be required to deliver the patient samples to a limited number of the provider certified laboratories from worldwide. Thereby, the cost of performing this test per patient poses a significant economic burden.

As mentioned above the clinical endpoint is the most important aspect to evaluate the clinical relevance of prognostic biomarkers. Different endpoints could be used for different clinical indications such as adverse pathology, BCR after radical treatment, metastatic progression and death with specified COD (overall survival, cancer-specific or non-cancer specific survival time). Oncotype DX is based on results from clinical studies mainly using BCR or adverse pathology as the final endpoints, and Prolaris uses death with specified COD as endpoint (overall and cancer-specific mortalities) but not for non-cancer-specific mortality, which comprises the majority of deaths (Figure 6). PCa heterogeneity and more studies with clinically relevant outcome endpoints allow for the identification of better cancer biomarkers to improve overall survival estimation.

8 EMBRYONIC STEM (ES) CELLS

As introduced in an earlier section, genetic or genomic heterogeneity of PCa is the key feature to potentially identify more fundamental cancer biomarkers reflecting PCa tumor heterogeneity with different aggressiveness that consequentially determine different clinical outcomes.

Embryonic stem (ES) cells are a type of undifferentiated cells capable of self-renewal and differentiating pluripotent. ES cells can be isolated from blastocysts, which are clusters of 50-150 cells that develop 4-5 days post fertilization, at the so-called blastocyst stage (66). The

pluripotency of ES cells allows them to differentiate into any type of organ cells. During the organ forming stage, ES cells normally differentiate into tissue stem cells or tissue progenitors, eventually differentiated into terminal tissue cells.

ES cell differentiation is a process of temporary loss of self-renewal and pluripotency, nowadays the induced pluripotent stem (iPS) cells technology has demonstrated that terminally differentiated cells can re-gain ‘stemness’ ability by artificially introducing a set of key iPS cell genes (67). Thereby the cell’s ‘stemness’, characterized by pluripotency and self-renewal, indicates a potential plasticity. Previous studies reported that the gene expression patterns of ES cells could stratify tumor subtypes with different aggressiveness in terms of clinical outcome and therapy efficacy (68). Identification and isolation of a very small proportion of prostate cancer stem cells (CSCs) in order to identify the gene expression signature reflecting the ‘stemness’ is experimentally challenging. Instead, we aimed at finding the origin of ‘stemness’ from ES cells in our study I.

Therefore we hypothesized that: 1) Genes that are important in maintaining ESC status and regulating cell differentiation are also important in abnormal differentiation (dedifferentiation). 2) Genes that show consistently high or consistently low expression levels across various ES cell lines are equally important in maintaining ESC status. Different expression patterns of these genes determine the development of different normal or cancer tissues. These genes are here named as ESCGPs (embryonic stem cell gene predictors). 3) These ESCGPs may be expressed in cancer cells and their expression levels can be measured by RT-PCR. 4) Different expression patterns of these ESCGPs measured in the cancer tissues can reflect cancer’s biological aggressiveness, and predict the efficacy of treatment and patient survival.

9 SAMPLE TYPE

Biomarkers can be measured in several types of samples, including blood, urine and tissue. Currently the most accessible sample types are blood and urine samples, such as PCA3 measured in urine, PSA measured in blood. These types of samples are easy to collect and can be measured freshly no matter which type of molecules (i.e. nucleic acids or proteins). The disadvantage of these types of samples lies in the fact that they cannot be easily stored for longer period of time in contrast to FFPE samples. The FFPE method has been used for over a century to store tissue samples. Formalin fixation and paraffin embedding maintain tissue structure over decades, however the genetic materials (RNA and DNA) within FFPE tissue are dramatically damaged right after the fixation process. Although there are large amounts of FFPE tissue materials with good clinical follow-up data accessible for researchers compared other freshly frozen blood or tissue samples, FFPE tissue samples have not been utilized for molecular testing until recently. In recent times, updated technologies have been introduced such as Taqman® probe based RT-qPCR technology, specially optimized DNA/RNA sequencing technologies for FFPE tissue, and customized RNA/DNA extraction kits for FFPE tissue.

In a straightforward approach, molecular genomic tests using prostate FFPE tissue samples can extract genomic information directly from the prostate gland. As earlier mentioned, the Oncotype Dx and the Prolaris gene expression tests exploit prostate cancer tissue material from prostate FFPE core needle biopsy samples. For prostate needle biopsy samples, there is a practical issue that usually needs to be addressed. Needle biopsy samples often contain two types of Gleason patterns, e.g. Gleason grad 3 or 4. For experimental staff, it would be advantageous to know which histopathological type of cancer tissue should be taken counting into one sample measurement for single patient. Knowing that patients with different GSs show different survival rates, it would be practically relevant to evaluate whether gene signature expression levels are affected by different Gleason or other pathological patterns of cancer cells from the same patient.

10 SAMPLE COLLECTION

Prostate FFPE core needle biopsy samples often contain one or multiple prostate tissue cores taken by 18-gauge needles, which is a very limited amount of tissue. Routinely paraffin blocker shall be prepared into 1-10 μ m thick of FFPE sections, which are then stained with hematoxylin and eosin stain (H&E) or by immunohistochemical methods for cancer diagnosis. Apart from these sections that are used for diagnosis, the amount of remaining tissue that can be used for molecular genomic testing often is limited. Consequently genomic tests using prostate core needle biopsy tissue samples shall not require large amounts of tissue input.

Prostatic cancerous cells easily grow in a cluster-like pattern presented in core needle biopsy samples, which allows manual isolation of cancer cells from FFPE sections become feasible. However even a single core biopsy also could contain multi-focal cancer cells, with relatively massive cells gathering together within each focus. In this case, neither manual scraping of cancer cells method nor laser micro-dissection method would make sample collection easier and efficient. The manual method is less precise while the micro-dissection method is time consuming for this mini-scale of tissue collection. A more precise, faster and cheaper method of sample collecting method would be favorable to do mini-scale of FFPE tissue collection.

10.1 DIGITAL SCANNING

The digital scanning technology can scan the H&E stained FFPE slides to obtain digital images containing a large range of magnifications from 1 up to 40 times. The quality of these digitally scanned, high-resolution images is favorable compared to traditional microscopy and meets the requirements for pathological diagnosis. The images can be uploaded and stored in a cloud-based server in an encrypted manner, which allows pathologists all over the world to practice their jobs only by logging in to this server, independent of their geographical location assuming that they have access to the internet. Digital images do not require large spatial volumes to store these archive stained slides. These advantages could dramatically decrease the cost of transporting histochemically stained slides, and allow pathologists across large distances in order obtain confirmed diagnosis. Furthermore, and the cost of storing archived slides can be greatly reduced. Moreover the digitalized imaging method would allow for a

more precise quantification of the cancer area compared to the traditional calculation method using length of cancer area in core needle biopsies. In our studies II and III, samples were taken manually guided by digitally scanned images.

10.2 AUTOMATED SAMPLE COLLECTING SYSTEM

After investigation in our studies, we are trying to introduce a digital image-guided and automated sample collecting system to collect the cancer area from prostate core needle biopsy FFPE sections (Figure 7). From each FFPE block, one H&E stained slide is prepared and serving as a ‘map’, and sequential FFPE sections are used for RNA extraction. The H&E stained slide can be scanned to generate digital images, which can be uploaded to an internet server in an encrypted manner, allowing pathologists to perform cancer cell marking and annotation with precise area calculations independent of their geographical location. An automated mini-dissection system, developed by an external provider, can use the digitally scanned image serving as a ‘digital map’ to guide its instrument to efficiently match and dissect the sequential FFPE sections with high precision and speed. This method allows sample taking of cancer cells from FFPE tissue sections with well-controlled precision, efficiency and automaticity.

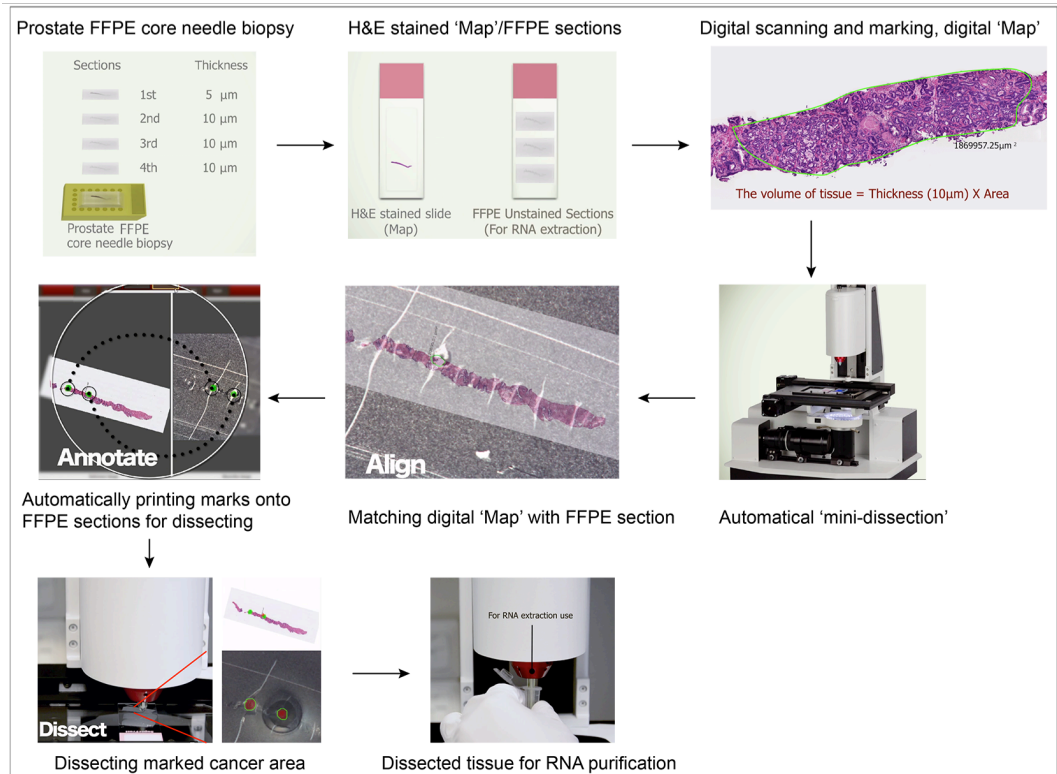


Figure 7. A digital image-guided and automated sample-collecting system.

11 PROSTATYPE TEST SYSTEM

There is need for a better prognostic biomarker indicating the utmost relevant clinical outcomes, such as estimating overall survival, PCa-specific survival and non-PCa specific survival. We are also aware of the significant contributions in terms of survival prediction from currently used clinical parameters such as GS, clinical stage, age at diagnosis, and PSA

value. Thus we performed a biomarker identification study based on ES cell theory in our first study, as the result, a three-gene signature was found showing independent prognostic significance in estimating overall survival time. Study II and III also verified the improved survival estimation accuracy by complementing gene expressions with conventional clinical parameters together in prostate FFPE core needle biopsy sample materials. Based on this finding, we transferred this invention into an industrialized project to develop an *in vitro* diagnostic (IVD) kit according to industrial regulatory standards such as ISO13485:2003 (69). The regulatory rules require standardized and comprehensive risk analyses, verification and validation plans in order to develop robust and to receive mandatory conformity marking approval in European area (CE marking) for the IVD test (Figure 8). The test should be qualified in many aspects such as transport stability, in-use stability, interfering substances, robustness, testing analytical performance and precision, tissue input, and reproducibility on three individual test sites (Figure 8).

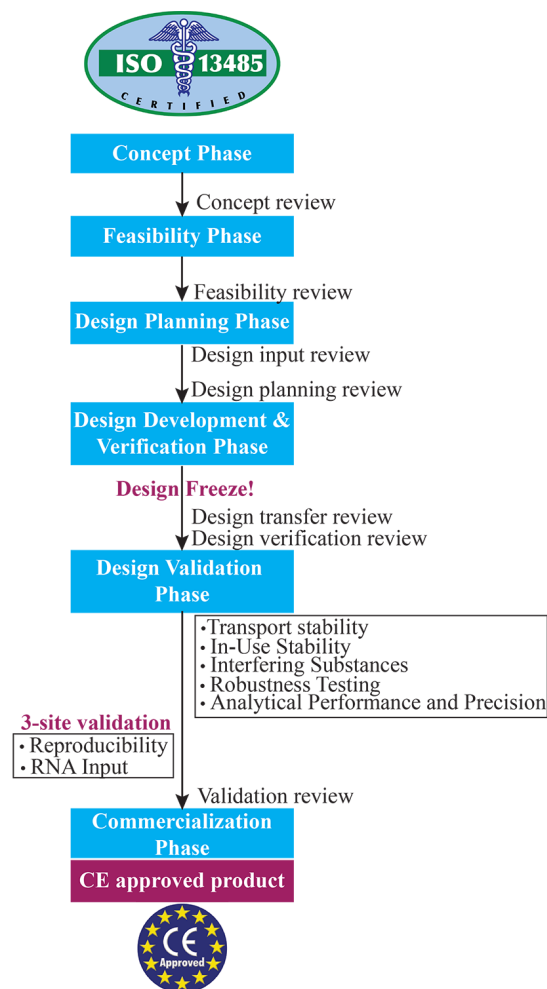


Figure 8. ISO13485:2003 regulatory procedures standardize the development of a industrial medical device product. Standardization of ISO13485:2003, Medical devices- Quality management systems-Requirements for regulatory purposes 2003. Available from: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=36786.

11.1 PROSTATYPE RT-QPCR KIT

The Prostatype Test System is composed of two parts. The first part is the Prostatype RT-qPCR kit, a four multiplex one-step RT-qPCR kit, which can simultaneously measure expression levels of three-gene signatures and one housekeeping gene in one reaction well. Purified total RNA from tissue samples containing a minimum of 2/3 of cancer cells is the input material for this kit. Within 1.5 hours, the expression levels of four genes can be measured.

11.2 CPMA

The second part of the Prostatype Test System is the Classification of Prostatic Malignancy Algorithm (CPMA), which is a software containing a large database. Based on the multiplex RT-qPCR reactions, the measured Ct values can be typed into the CPMA software together with four typical clinical parameters - age at diagnosis, pretreated PSA value, clinical stage and Gleason score - in order to generate the prognostic statement. The CPMA software contains a database that includes a large number of authentic, historical PCa patients' clinical outcomes such as gene signature data, data regarding the four classic clinical parameters, and information on the treatment and overall survival time. In contrast to randomized selection, those authentic historical PCa patients included in the CPMA were selected aiming at the largest possible variety of clinical characteristics including death, COD, survival time, treatment, age, GS, clinical stage, and PSA value. The patients mainly originate from Scandinavian and Swiss populations with at least 7-11 years of follow-up. The CPMA is being extended continuously and will be updated in specific time intervals.

CPMA uses a K-nearest neighbor (kNN) algorithm to calculate the similarity distance of a newly diagnosed patient compared to any of neighbors in its reference database. The kNN is a non-parametric statistical model used for classification or regression (70). As a result, the CPMA, which uses the measured gene data in combination with the clinical data of a newly diagnosed PCa patient, could find the most similar three authentic historical PCa patients in the database. These three patients are presented together with information regarding their treatments and overall survival time. Instead of inventing a totally new scoring system which would require intensive correlation studies to current risk scores such as the NCCN risk score, the CPMA utilizes its large database to directly present the clinical records of the most relevant historical patients (living or deceased) in terms of treatment options and survival time. Additionally the information is supplemented with weighted gene signature data resulting in improved overall survival estimation accuracy. Over time, the CPMA database would be expanded and thereby provide even better accuracy and precision by including patients from more various ethnicities, different PCa subtypes from different geographical regions. In a near future, a comprehensive CPMA database with big-data scale of genomic data and clinical outcome data would be a promising prognostic tool providing urologists, pathologists and oncologists an informative guidance to improve cancer management for PCa patients (Figure 9).

Compared to the two another prognostic biomarker tests, the Prostatype Test System is developed in a slightly less conventional manner (Table 3). Usually methods using FFPE cancer tissue to measure gene signature expression levels contain comprehensive steps in the test's workflow. Conventionally these tests are developed as a laboratory-based service instead of a ready-to-use IVD test kit or system, which can be performed at any molecular laboratory equipped with a real time quantitative PCR machine. This product type also dramatically decreases the additional transport costs of patient samples, and a digitalized CPMA software analysis method reduces time and manpower requirement. Thereby this test system has fundamental advantages for public health economy compared to the other two commercially available tests (Table 3).

In summary, the content of this thesis consists of (1) hypothesis-driven searching and identification of novel genomic biomarkers, (2) optimization of sample collection, (3) validation and verification of clinical indication of novel biomarkers, and potential integration of the industrially developed application into clinical situations.

		Oncotype ProstateDX®	Prolaris®	Prostatype®
Developer		Genomic Health Inc	Myriad Inc	Chundsell Medicals AB
Country		USA	USA	Sweden
Signature	Genes	12 genes from 4 pathways	31 cell cycle genes	3 embryonic stem cell gene predictors
	Housekeeping genes	5	15	1
	Total genes	17	46	4
	Measuring method	Two steps RT-qPCR	Two steps RT-qPCR arrays	One-step 4-plex RT-qPCR
	Results calculating method	Manually statistical calculation	Manually statistical calculation	CPMA database software
	Results	GPS score (Ref to NCCN risk scores)	CCP score (Ref to CAPRA risk scores)	3 closest authentic cases (Treatment/Overall survival time)
Sample	Type	FFPE core needle biopsy	FFPE core needle biopsy	FFPE core needle biopsy
	Minimum input (n x thickness x length x width)	6 x 5µm x 1.0 mm x 0.75mm	5 x 5µm x 2.0 mm x 0.75mm	6 x 10µm x 1.0 mm x 0.75mm
	Minimum input (mm3)	0.0225	0.0375	0.05
	Gleason pattern of cancer cells	Not mentioned	Not mentioned	Any type of prostatic cancer cells
	Cancer cells percentage	Not mentioned approx.>95%	Not mentioned approx.>95%	Verified >67%
	Sample drop-out rate	7%	15%	17%
	Test turnaround time	Approx. 2-4 working days	Approx. 2-4 working days	8 hours
Clinical study	Study patients number	382	585	241 + 420 + n
	Studies Endpoint	Biochemical recurrence	PCa-specific death	Overall death, PCa/Non-PCa specific death
Clinical application	Prediction	BCR free survival	PCa-specific survival	Overall survival, PCa/non-PCa specific survival
	Treatment indication	Ref to NCCN	Ref to CAPRA	Historical treatments of 3 authentic cases
	Risk factors	Tumor risk factors	Tumor risk factors	Tumor risk factors & Patient risk factors
	Suitable patients	Low risk PCa	Any PCa	Any PCa
Product	Business model	Lab based service	Lab based service	CE marked Kit and CPMA software
	Patent status	US pending	US pending	Sweden approved, PCT phase
	Laboratory requirement	Genomic Health's Labs	Myriad's Labs	Any pathology Labs
	Price/Patient	Approx. 4000 USD	Approx. 3000 USD	Approx. 10 000 SEK

Table 3. Comparison of three prostate cancer biomarker tests.

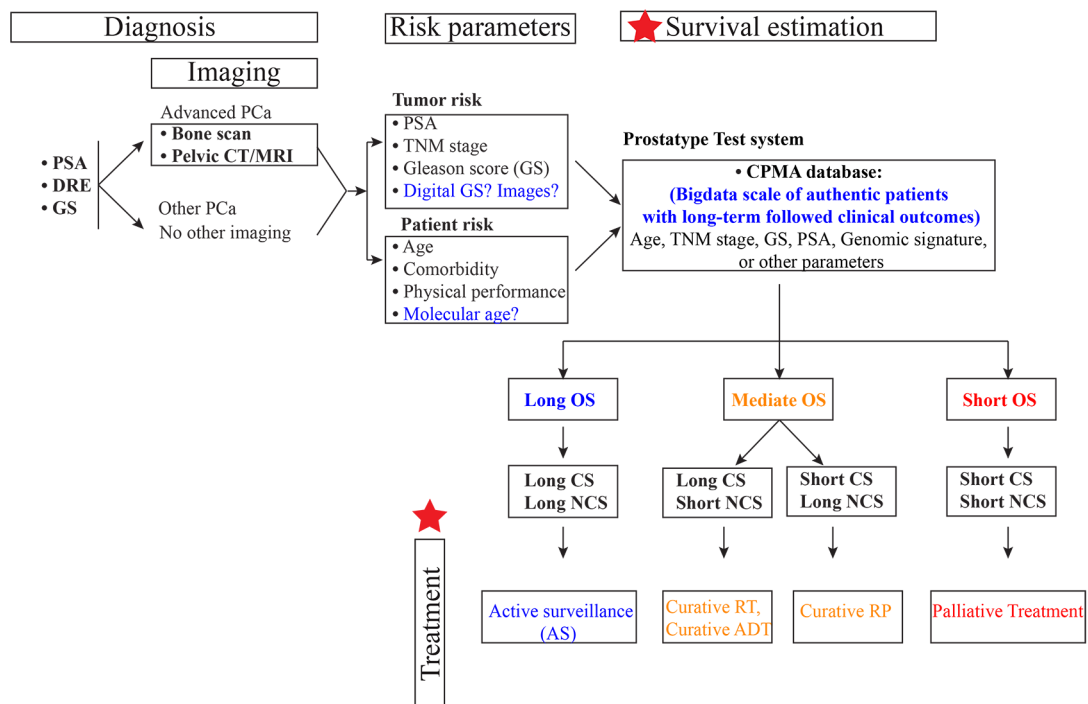


Figure 9. Potential clinical application of the Prostatype Test System in the future. OS: overall survival, CS: cancer-specific survival, NCS: non-cancer-specific survival.

AIM OF THE THESIS

Specific aims

In this thesis work, we aimed at identifying genetic biomarker bases on the gene expression signature of ES cells, in order to stratify prostate tumor subtypes with different aggressiveness in terms of relevant clinical outcomes at time of PCa diagnosis. We further pursued validation of the prognostic value of the identified biomarkers in an external cohort with relevant clinical end-points and long-term follow-up. Moreover our goal is to implement the verified biomarkers into a clinical application, which could provide guidance for clinicians in the development of improved and personalized treatment decision for PCa patients.

Paper I

To identify genetic biomarker candidates from the ES cells based on their gene expression signature, the embryonic stem cell gene predictors (ESCGPs), to verify the certain key ESCGPs can classify tumor subtypes with overall survival differences in a long-term follow-up Swedish fine needle aspiration biopsy cohort.

Paper II

To evaluate the identified gene signature (paper I) expression differences effect derived from the operator dependent choice of prostate cancer biopsy in FFPE core needle biopsy tissue material.

Paper III

To validate the prognostic value of the identified gene signature from paper I in an external population-based Swedish cohort with prostate FFPE core needle biopsy material, and to discuss the most optimal manner for implementing the newly identified biomarkers into clinical application.

RESULTS AND DISCUSSION

Paper I

An Expression Signature at Diagnosis to Estimate Prostate Cancer Patients Overall Survival

We hypothesized that the gene expression signature of ES cells could determine the tumor subtypes with different levels of aggressiveness. To test this hypothesis, we developed and identified the embryonic stem cell gene predictors (ESCGPs) concept. The ESCGPs concept is based on the following assumptions: 1) Embryonic stem cells are the origin of tissue differentiated cells and tissue stem cells. 2) Genes that are important in maintaining ES cell status and regulating differentiation are also important in abnormal differentiation (dedifferentiation). 3) Genes with significant expression variations among different ES cell lines are not relevant in this respect. 4) Genes that show consistently high or consistently low expression levels across various ES cell lines are equally important in maintaining ES cell status. Different expression patterns of these genes determine the development of different normal or cancer tissue. These genes are here refereed to as ESCGPs. 5) These ESCGPs may be expressed not only in CSCs but also in cancer cells and their expressions can be measured by microarray, RT-PCR or qPCR. 6) Different expression patterns of these ESCGPs measured in the tumor tissues can reflect the cancer's biological aggressiveness, and predict the efficacy of treatment as well as patient survival.

The study results are presented as the following steps:

Step 1: Identification of candidate ESCGPs

From the Stanford Microarray Database (SMD), we retrieved previously published datasets of whole-genome cDNA microarrays of five human ES cell lines (71). These were normalized using the datasets of 115 human normal tissues including various organs (72). A data subset with the whole-genome expression profile of 24361 genes in the ES cell lines was isolated. A single-class SAM (Significance Analysis of Microarrays) (73) was performed, whereby all genes were ranked according to the consistency (without significant variations) of their expression levels across the ES cell lines. As a result, SAM analysis of this data identified 328 genes with consistently high levels of expression and 313 genes with consistently low levels of expression in ES cells i.e. 641 ESCGPs in total.

Step 2: Selection of candidate ESCGPs in prostate cancer

An independent dataset (7) with 112 prostate tissue samples was used to verify the ESCGP findings and to select ESCGPs associated with PCa. The list of genes in the published dataset was matched to the list of the candidate ESCGPs identified in Step 1. The ability of the 641 genes to classify tumor subtype was verified on an independent dataset of 112 PCa samples. In this analysis, the clustering result was almost identical compared to the complete original

data set of 5513 genes and the 258 PCa related ESCGPs isolated from the same original data set.

Step 3: Refining ESCGPs selection using RT-PCR and multiplex qPCR analyses of three prostate cancer cell lines.

A 4-plex qPCR method was optimized for the quantification of these genes by using RNAs from three prostate cancer cell lines (LNCaP, DU145, PC-3). Among the 258 verified prostate cancer ESCGPs, the 34 genes of highest-ranking order in the SAM analyses performed in step 2, were selected for follow-up analysis. In addition, 5 reported genes based on previously published studies were included in the same set. The 19 ESCGPs and 5 reported genes were included in an optimization of the 4-plex qPCR using RNAs from prostate cancer cell lines, and ready to use for analysis of FNA samples taken from prostate cancer patients.

Step 4: Establishing of the clinical relevance

A Swedish cohort composed of 189 PCa patients diagnosed between 1986 and 2001 was studied to evaluate clinical relevance of the previous findings. Patient samples were collected by fine needle aspiration (FNA) cytology smear samples at the time of diagnosis, freshly frozen and stored until the time of analysis in 2008. The cohort was followed from diagnosis until December 31, 2008. At the end point of the study, 22 of the original 189 patients were still alive, 163 were deceased, and 4 could not be found in the registries. In a step-wise manner, the cohort's gene profile was analyzed divided accordingly into three subsets of patients.

Out of the 25 gene expression markers that were measured, 10 (F3, WNT5B, VGLL3, CTGF, IGFBP3, c-MAF-a, c-MAF-b, AMACR, MUC1 and EZH2) were significantly correlated with either overall or PCa-specific survival. Of more than 120 gene signature combinations derived from these 10 significant genes, a gene signature of three genes - IGFBP3, F3 and VGLL3 - showed the best stratification ability of tumor subtypes. For 87 patients, all clinical parameters were available and according to the expression of three-gene signature these patients could be categorized into three subtypes. The median overall survival time was 3.23 years for patients with the high-risk subtype, 4.00 years for the intermediate-risk subtype and 9.85 years for the low-risk subtype, and these values corresponded to hazard ratios of 5.86 (95% CI 2.91-11.78, $P<0.001$) for the high-risk subtype and 3.45 (95% CI 1.79-6.66, $P<0.001$) for the intermediate-risk subtype compared to the low-risk subtype.

The kNN classification algorithms were developed using the training set to estimate the overall survival (74). The performance of the kNN model using only clinical parameters was similar to the random model, whereas all kNN models including the selected ESCGP genes were significantly ($P<0.04$) better than the random model. Compared to the prediction model that used only the clinical parameters, when the combining three-gene signature and clinical parameters, the area under the curve (AUC) value was increased from 0.755 to 0.815 in

overall survival prediction, from 0.726 to 0.793 in PCa-specific survival prediction, and from 0.730 to 0.793 in Non-PCa specific survival prediction, respectively.

In summary, the three-gene ESCGP signature is a promising biomarker combination suitable for estimating the survival of PCa patients. After validation in an independent cohort study, it would provide an important and orthogonal complement to the current clinical parameters that are routinely used in the process of treatment decision for individual patients, in particular for patients diagnosed with early-stage PCa.

Paper II

Operator dependent choice of prostate cancer biopsy has limited impact on a gene signature analysis for the highly expressed genes IGFBP3 and F3 in prostate cancer epithelial cells

FFPE tissue material derived from either biopsy material or surgically removed tumors constitutes an appropriate and easily accessible sample. For PCa, FFPE core needle biopsies on which Gleason grading for diagnosis has been conducted, are readily available in the clinical routine pathology laboratories and suitable for such analyses. Typically, multiple biopsy samples were collected from each patient. Since Gleason grading is an operator dependent procedure known to be difficult, the impact of the operator's choice of biopsy needs to be evaluated.

Multiple biopsy samples from 43 patients were evaluated using the previously reported gene signature of IGFBP3, F3 and VGLL3 for their potential prognostic value in estimating overall survival at diagnosis of prostate cancer. A four multiplex one-step RT-qPCR test kit, designed and optimized for measuring this three-gene signature in FFPE core needle biopsy samples was used. Concordance of gene expression levels between primary and secondary Gleason tumor patterns, as well as benign tissue specimens was analyzed.

We found gene expression levels of IGFBP3 and F3 in prostate cancer epithelial cell-containing tissue representing the primary and secondary Gleason patterns were high and consistent. On the contrary, VGLL3 was expressed at markedly lower levels and showed a higher extent of variation in its expression levels.

In summary the assessment of IGFBP3 and F3 gene expression levels in prostate cancer tissue is independent of Gleason patterns. Thereby, we can conclude that the impact of operator's choice of biopsy is low.

Paper III

Improving the prediction of prostate cancer overall survival by supplementing readily available clinical data with gene expression levels of IGFBP3 and F3 in formalin-fixed paraffin embedded core needle biopsy material.

In a previous study performed by our laboratory, we showed that measurement of the expression levels of a three-gene signature (IGFBP3, F3 and VGLL3) in fresh frozen FNA cytology samples provides a reliable estimate of the overall survival time for PCa patients at diagnosis. The gene signature provided additional prediction power in terms of patients' survival compared to the standard clinical parameters, such as age at diagnosis, cytology WHO grade, tumor stage and PSA value. Gleason score (GS) cannot be determined for FNA samples.

In this work, we carried out a new cohort study with 241 prostate cancer patients diagnosed between 2004 and 2007 with a follow-up exceeding 6 years in order to verify the prognostic value of gene expression signature in FFPE prostate core needle biopsy tissue samples. The cohort consisted of four patients groups with different survival times and death causes. There were two groups of deceased cases, prostate cancer death within 5 years and death due to other diseases within 5 years. We used two control groups, one matched alive group where GS and age were matched to the deceased groups, and one randomly selected alive patient group. The main purpose of the study was to determine whether there are any differences in expression levels of IGFBP3 and F3 within these different patient groups with significantly different survival time. We also attempted to verify whether the gene signature combined together with current clinical parameters can provide higher prediction accuracy in terms of patients' survival time, compared to the prediction solely based on clinical parameters.

Our study presented in the paper II found the effect that operator dependent choice of FFPE core needle biopsy based Gleason pattern of epithelial cancer cells as basis for measurement of expression levels of IGFBP3 and F3 had limited impact on the results, when using the Prostatype RT-qPCR kit. The effect on VGLL3 measurements could not be estimated in that study due to limited tissue input. We only analyzed the gene expression levels of IGFBP3 and F3, measured only on the primary Gleason pattern tissue samples.

Survival time predictions only based on the current clinical parameters, such as age at diagnosis, Gleason score, PSA value and tumor stage, were compared to survival estimations considering a combination of clinical parameters and expression levels of IGFBP3 and F3. The results of k Nearest Neighbor (kNN) analysis, multiple linear regression analysis and nominal logistic regression modeling showed that when combined with currently used clinical parameters, the gene expression levels of IGFBP3 and F3 could improve the prediction accuracy of survival time compared to using clinical parameters.

In summary, the assessment of IGFBP3 and F3 gene expression levels in FFPE prostate cancer tissue could provide an improved survival prediction for prostate cancer patients at the time of diagnosis. We provide evidence, that expression levels of IGFBP3 and F3 in combination with clinical parameters such as Gleason score most probably play an important role in the stratification of newly diagnosed prostate cancer patients. The results reported in this study warrants initiation of further investigations to evaluate the use of gene expression as a complement to clinical parameters to improve prediction accuracy of PCa prognosis. Studies with larger cohorts and survival follow-up exceeding 10 years would be required to further improve survival prediction and treatment choice. This could be particularly relevant for patients who could be safely assigned to active surveillance.

FUTURE STUDIES AND PERSPECTIVE

As an industrial sponsored PhD student associated both with an academic research institute (Karolinska Institutet) and a company (Chundsell Medicals AB), my research has always been performed with considerations of two perspectives: scientific and industrial. Industrial medical research is more focused on how to transfer and integrate scientific findings into routine clinical settings using applicable methodologies. Therefore my future research will continue to work on these projects that have both scientific and industrial aspects.

Part I. Expanding the reference database of CPMA

In paper III, we validated the prognostic value of two of three genes that were identified from in the cohort in paper I, even though it was challenging since the quality of the sample material was considerably much lower. More cohorts of clinical studies are needed to further validate the prognostic value of these two genes in a much larger number of patients, and patients from different countries should be included. At the same time, CPMA uses the kNN model to calculate the similarity of each parameter of each patient, in which the modeling has higher noise bias comparing to the conventional survival analyses. This means more patients filled in the reference database, more accurate kNN modeling prediction with lower noise bias.

To expand the CPMA reference database, approximately 420 additional Swedish patients, diagnosed from 2004 to 2008, were selected and their samples are currently being tested. Some archive samples collected from patients are on the way to be filled in the CPMA reference database in the near future: about 200 from Sweden, 200 from Switzerland, and more than 200 from Germany.

Part II. Developing the second version of the CPMA

In paper II and III, VGLL3 has been excluded for further analyses and discussion, mainly due to the fact that average expression levels of this gene were relatively low in prostatic cancerous epithelial cells compared to IGFBP3 and F3. This causes higher noise of Ct values resulting in less reliable data, and further contributing to increased difficulty and complexity of analysis. However, interestingly, our unpublished data indicates that VGLL3 expression levels in prostatic benign cells were associated with survival time, which we did not publish in the previous studies. If this finding can be validated, the process of sample collection could be dramatically improved rendering the test system user-friendlier. Currently the test system requires that a minimum of 67.7% of cancer cells used for RNA isolation. If VGLL3 expression in benign prostatic cells can be used to estimate survival time, this restriction might not be required anymore. In that way, samples containing cancer cells and benign cells at any ratio or even benign cells alone could be utilized to generate relevant results.

Part III. Prospective clinical validation of the Prostatype Test System

Ultimately, prospective clinical studies will be initiated once the Prostatype Test System is CE-marked according to regulatory rules in order to clinically validate the prediction accuracy when using the Prostatype test system in a prospective manner. After at least 5 years of follow-up, we could evaluate whether the initial survival prediction by the test system is accurate as compared with the real survival outcomes. With prospective clinical studies further validating the improvement of the Prostatype Test System, a shifting in clinical PCa management diagrams would be possible. It can be speculated that the proportion of patients who are recommended for an active surveillance (AS) program is going to increase (Figure 10).

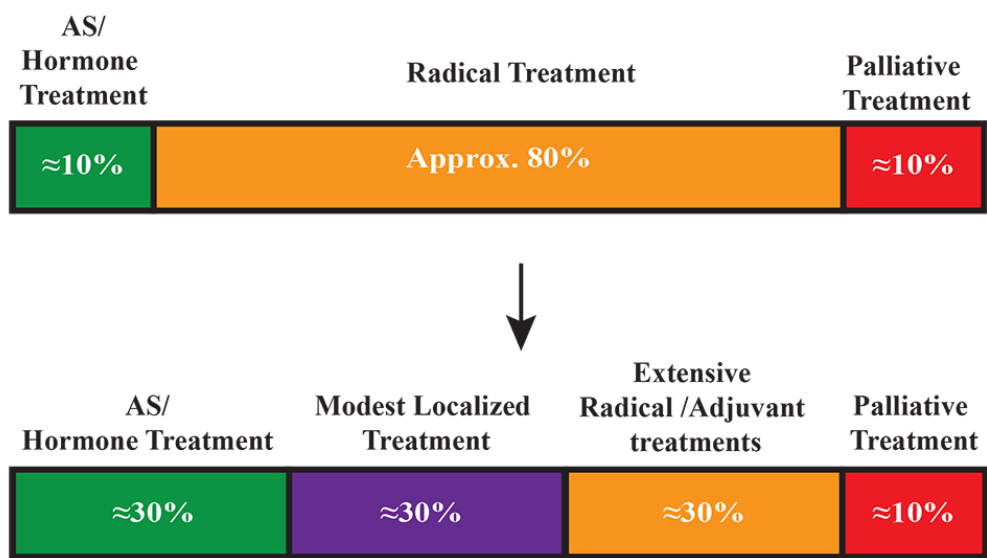


Figure 10. Shifting of PCa treatments in the future.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to everyone who has been supporting me with guidance, encouragement and accompany during my PhD study.

Dr. Chunde Li, my main supervisor also a close friend, I deeply thank you for all efforts that you have made to open me a new window in the scientific world. Moreover your kindness as a great person has convinced me helping people always weights more than gaining from them. Apart from scientific inspirations, far more importantly, what you have inspired me is a new way of living, learning and giving. Thank you!

Professor Monica Nistér and **professor Sten Nilsson**, my dear co-supervisors, thank you for all your supports as always! **Monica**, thank you for your nearly flawless details when revising for our manuscripts. **Sten**, thank you for always being strongly and solidly backing us up every time Chunde and I really need help. I sincerely appreciate your help!

To all my scientific collaborators: **Professor (Associate) Johan Lindholm**, thanks for your priceless supports to our projects, you presented me with a whole new world in urological pathology research. Hopefully we can continue to explore it and make it absolutely better. **Professor Yudi Pawitan**, thanks you for your profound statistic help and guidance! You are extremely professional and serious. Hope that we can keep working together! **Professor Lambert Skoog**, thank you for the contribution of clinical samples with one of the best quality in the whole world. **Proferssor Lars Ährlund-Richter**, my mentor, thanks for your guidance and wisdom outside of science.

To lab colleagues and friends: **Yuanjun, Yanling, Xiaobing, Xinming, Salah, Toota, Inga, Rona, Urica, Sanna, Mikael, Karl, Kaveh, Anna, Johanna, Teresita, Mingqi, Jian, Lena**, for your accompanies and help for so many years in lab. **Lidi, Ran, Linjing, Xiaonan, Xin, Qiang, Mei, Dudi, Emarn, Sussane**, for enjoyable lunch chatting in the kitchen over years! **Min Wan, Li Yu, Jikui, Qiaoling, Jinzhe, Hongshi, Xiaoqun** for your kind help when I firstly came to Sweden, I'm lucky to know you guys since then. **Ziming, Qin Li and Qing Cheng**, thanks for your kindness and being friends with me.

Thanks those who ensure CCK always run: **Sören, Erica, Eva-Lena** and **Juan** for your efficient and fully supportive help! To **Ann-Brit** and **Erica** who always keep the admin up-to-date, and extremely efficient. **Inger**, thanks for your expert help in the histochemical Lab!

To all my industrial colleagues, collaborators and friends: **Jan Erik Collin**, my mentor in diagnostic industry, you have been trying to teach me nearly everything you have known over years about this industry. You successfully transformed me from a scientific researcher into an industrial developer if I may say so. You even managed to teach me how to do down hill skiing in black slopes! Plus you are a great boss and person! I sincerely appreciate what you have done for me! **Professor Sune Rosell**, thanks for your positive supports! Your knowledge and experience both in academics and industry have been always inspiring me! **Professor (associate) Karl Andersson**, your kindness and expertise ultimately helped me to

accomplish the PhD study much more efficiently! You have always been a brilliant problem-solver. Without your help I cannot imagine Chundsell could develop so quickly in such a short period of time. Thank you indeed!

To my dear colleagues in Chundsell Medicals: **Annika Bolind Bågenholm**, we have been working together to develop the Prostatype Test System since the early beginning of the project. You are very kind, helpful and smart! And you contribute great efforts for this project! Hope both of us can work another 50% time more in Chundsell after my PhD defense! It has been really happy time to work with you! **Göran Sällström**, you are an elegant, brilliant, outstanding salesman! You are also a master of enjoying and embracing life from all exclusive perspectives. I admire you always being honest to the others and especially to yourself. And I heard you are a great dancer! Wish we would have more fun adventures together in the near future! **Jennie Andersson**, thanks for your caring reminding always according to the QA standards! Without your suggestion of time plan for PhD defense, I would be desperate now for sure... Thanks indeed! **Rong Cao**, you are the queen of Chundsell's Lab! The 'mama' to all instruments and the Prostatype RT-qPCR kits! You take care of them all! Thanks for your ubiquitous caring to everyone! **Jennifer Honek**, the newest member of us, thanks for your absolutely beautiful English editing for this thesis! (OBS! this chapter has never been edited by you, I ensure that any misuse of English wordings and writings in this chapter has nothing to do with you!) Your talent and potential is unbelievable just like your 100km running! I'm lucky to work with you!

To all my collaborators and friends in Berlin: **Oliver Hasinger, Reimo Terzner, Gunter Weiss, Uwe Staub, Thomas Taapken**, for your excellent and extremely professional expertise in developing our product! I have learnt a lot from all of you! Our collaboration will continue!

To all collaborators and friends in Switzerland: **Professor Peter Wild**, for your enthusiastic supports for our collaborations, without your help Chundsell would not get anywhere like today. **Dr. Christine Fritz**, thanks for your excellent and meticulous work in our collaborations!

To all talented artists that I know: **Mats Bergman**, for your absolutely outstanding skills in writing manuscript and directing films! Without your help, we would be still struggling to find out how to explain the CPMA to our customers with endless wordings and ended up useless! **Lin Chen**, you are unbelievably talented and with great potential to develop your art career in a whole new level in the future! Your work for the Prostatype Test System is our video-version of lab protocol! I admire you being always brave and grateful to find your own way to understand and interpret everything in this world, and absolutely enjoy and embrace the life that you can explore. I'm lucky to meet you!

To all the friends over years: **Christer Ericsson**, you have been always a really decent and reliable friend, also my mentor giving me so much philosophy guidance and wisdom! Our friendship is becoming closer over the years even after a bit longer Swedish 'warming-up'

time (absolutely less than 5 years) to be started. I have been always a fan of your black humors! I really appreciate that you being my friend, and you are a great one! **Jinfeng Shen**, my buddy! A great friend! You have been always supportive and helpful to me over years! Our friendship was derived from Sweden, now becomes a long-distance one! Not easy, but so worthy! **Tong Liu, Huiyuan Zhen, Olle Olsson, Yang Liu**, for your supports and tolerance to my endless talking, no matter what I am talking about. Thank you my dear friends! **Claudia Queiroz**, you are a very kind and warm person! It has been always entertaining and fun when we spent time together! Let's go get tanned together outside of your beach house in a near future!

To anyone I have missed out, sorry and thank you!

To my **grandma (奶奶)** and **mama, papa (妈妈, 爸爸)**! 谢谢你们! Thanks for having me and giving me a chance to explore this wonderful world, for example, such a long-term, comprehensive and rewarding PhD study!

12 REFERENCES

1. Feletto E, Bang A, Cole-Clark D, Chalasani V, Rasiah K, Smith DP. An examination of prostate cancer trends in Australia, England, Canada and USA: Is the Australian death rate too high? *World journal of urology*. 2015.
2. Guinan P, Bhatti R, Ray P. An evaluation of prostate specific antigen in prostatic cancer. *The Journal of urology*. 1987;137(4):686-9.
3. Stamey TA, Yang N, Hay AR, McNeal JE, Freiha FS, Redwine E. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. *The New England journal of medicine*. 1987;317(15):909-16.
4. Andriole GL, Crawford ED, Grubb RL, 3rd, Buys SS, Chia D, Church TR, et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *Journal of the National Cancer Institute*. 2012;104(2):125-32.
5. Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Zappa M, Nelen V, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet*. 2014;384(9959):2027-35.
6. Hayes JH, Barry MJ. Screening for prostate cancer with the prostate-specific antigen test: a review of current evidence. *Jama*. 2014;311(11):1143-9.
7. Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, Montgomery K, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101(3):811-6.
8. Bergerheim US, Kunimi K, Collins VP, Ekman P. Deletion mapping of chromosomes 8, 10, and 16 in human prostatic carcinoma. *Genes, chromosomes & cancer*. 1991;3(3):215-20.
9. Li C, Larsson C, Futreal A, Lancaster J, Phelan C, Aspenblad U, et al. Identification of two distinct deleted regions on chromosome 13 in prostate cancer. *Oncogene*. 1998;16(4):481-7.
10. Li C, Berx G, Larsson C, Auer G, Aspenblad U, Pan Y, et al. Distinct deleted regions on chromosome segment 16q23-24 associated with metastases in prostate cancer. *Genes, chromosomes & cancer*. 1999;24(3):175-82.
11. Shendure J, Ji H. Next-generation DNA sequencing. *Nature biotechnology*. 2008;26(10):1135-45.
12. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, et al. The genomic complexity of primary human prostate cancer. *Nature*. 2011;470(7333):214-20.
13. Oszlak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nature reviews Genetics*. 2011;12(2):87-98.
14. Catalona WJ, Richie JP, Ahmann FR, Hudson MA, Scardino PT, Flanigan RC, et al. Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men. *The Journal of urology*. 1994;151(5):1283-90.

15. Epstein JI, Allsbrook WC, Jr., Amin MB, Egevad LL, Committee IG. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *The American journal of surgical pathology*. 2005;29(9):1228-42.
16. Gleason DF. Classification of prostatic carcinomas. *Cancer chemotherapy reports Part 1*. 1966;50(3):125-8.
17. Sooriakumaran P, Lovell DP, Henderson A, Denham P, Langley SE, Laing RW. Gleason scoring varies among pathologists and this affects clinical risk in patients with prostate cancer. *Clinical oncology*. 2005;17(8):655-8.
18. Leslie H. Sobin (Editor) MKGE, Christian Wittekind (Editor). *International Union Against Cancer (UICC): TNM Classification of Malignant Tumours, 7th Edition*. 7th Edition ed: Wiley-Blackwell; 2009 November.
19. Mohler JL, Kantoff PW, Armstrong AJ, Bahnson RR, Cohen M, D'Amico AV, et al. Prostate cancer, version 2.2014. *Journal of the National Comprehensive Cancer Network : JNCCN*. 2014;12(5):686-718.
20. D'Amico AV, Whittington R, Malkowicz SB, Cote K, Loffredo M, Schultz D, et al. Biochemical outcome after radical prostatectomy or external beam radiation therapy for patients with clinically localized prostate carcinoma in the prostate specific antigen era. *Cancer*. 2002;95(2):281-6.
21. D'Amico AV, Whittington R, Malkowicz SB, Fondurulia J, Chen MH, Kaplan I, et al. Pretreatment nomogram for prostate-specific antigen recurrence after radical prostatectomy or external-beam radiation therapy for clinically localized prostate cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 1999;17(1):168-72.
22. D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *Jama*. 1998;280(11):969-74.
23. N. Mottet (Chair) JB, E. Briers (Patient Representative),, R.C.N. van den Bergh (Guidelines Associate) MB, N.J. van Casteren (Guidelines Associate) PC, S. Culine,, S. Joniau TL, M.D. Mason, V. Matveev, H. van der Poel,, T.H. van der Kwast OR, T. Wiegel. *EAU Guidelines on Prostate Cancer-updated March 2015*. 2015.
24. Salinas CA, Tsodikov A, Ishak-Howard M, Cooney KA. Prostate cancer in young men: an important clinical entity. *Nature reviews Urology*. 2014;11(6):317-23.
25. Chamie K, Daskivich TJ, Kwan L, Labo J, Dash A, Greenfield S, et al. Comorbidities, treatment and ensuing survival in men with prostate cancer. *Journal of general internal medicine*. 2012;27(5):492-9.
26. Hoffman RM, Koyama T, Albertsen PC, Barry MJ, Daskivich TJ, Goodman M, et al. Self-Reported Health Status Predicts Other-Cause Mortality in Men with Localized Prostate Cancer: Results from the Prostate Cancer Outcomes Study. *Journal of general internal medicine*. 2015;30(7):924-34.
27. Daskivich TJ, Fan KH, Koyama T, Albertsen PC, Goodman M, Hamilton AS, et al. Effect of age, tumor risk, and comorbidity on competing risks for survival in a U.S. population-based cohort of men with prostate cancer. *Annals of internal medicine*. 2013;158(10):709-17.

28. Leung KM, Hopman WM, Kawakami J. Challenging the 10-year rule: The accuracy of patient life expectancy predictions by physicians in relation to prostate cancer management. *Canadian Urological Association journal = Journal de l'Association des urologues du Canada*. 2012;6(5):367-73.
29. Albertsen PC, Fryback DG, Storer BE, Kolon TF, Fine J. The impact of co-morbidity on life expectancy among men with localized prostate cancer. *The Journal of urology*. 1996;156(1):127-32.
30. Cowen ME, Halasyamani LK, Kattan MW. Predicting life expectancy in men with clinically localized prostate cancer. *The Journal of urology*. 2006;175(1):99-103.
31. Sammon JD, Abdollah F, D'Amico A, Gettman M, Haese A, Suardi N, et al. Predicting Life Expectancy in Men Diagnosed with Prostate Cancer. *European urology*. 2015.
32. Sweeney CJ, Chen YH, Carducci M, Liu G, Jarrard DF, Eisenberger M, et al. Chemohormonal Therapy in Metastatic Hormone-Sensitive Prostate Cancer. *The New England journal of medicine*. 2015.
33. Fizazi K, Faivre L, Lesaunier F, Delva R, Gravis G, Rolland F, et al. Androgen deprivation therapy plus docetaxel and estramustine versus androgen deprivation therapy alone for high-risk localised prostate cancer (GETUG 12): a phase 3 randomised controlled trial. *The Lancet Oncology*. 2015;16(7):787-94.
34. Harris WP, Mostaghel EA, Nelson PS, Montgomery B. Androgen deprivation therapy: progress in understanding mechanisms of resistance and optimizing androgen depletion. *Nature clinical practice Urology*. 2009;6(2):76-85.
35. Loriot Y, Eymard JC, Patrikidou A, Ileana E, Massard C, Albiges L, et al. Prior long response to androgen deprivation predicts response to next-generation androgen receptor axis targeted drugs in castration resistant prostate cancer. *European journal of cancer*. 2015;51(14):1946-52.
36. Howard M, Sandler CH, Seth A, Rosenthal, Oliver Sartor, Leonard G. Gomella, Mahul Amin, James Purdy, Jeff M. Michalski, Mark Garzotto, Nadeem Pervez, Alexander G. Balogh, George Rodrigues, Luis Souhami, M. Neil Reaume, Scott G. Williams, Raquibul Hannan, Eric M. Horwitz, Adam Raben, Rebecca Paulus, William U. Shipley; Department of Radiation Oncology, Cedars-Sinai Medical Center, Los Angeles, CA; NRG Oncology Statistics and Data Management Center, Philadelphia, PA; Sutter Cancer Center, Sacramento, CA; Tulane University School of Medicine, New Orleans, LA; The Sidney Kimmel Cancer Center at Thomas Jefferson University, Philadelphia, PA; Cedars-Sinai Medical Center, Los Angeles, CA; UC Davis, Sacramento, CA; Washington University in St. Louis, St. Louis, MO; Portland VAMC, Portland, OR; Cross Cancer Institute, Edmonton, AB, Canada; Tom Baker Cancer Center, Calgary, AB, Canada; Department of Radiation Oncology, London Regional Cancer Program, London, ON, Canada; Department of Radiation Oncology, McGill University Health Centre, Montreal, QC, Canada; Ottawa Hospital Cancer Center, Ottawa, ON, Canada; Peter MacCallum Cancer Centre, East Melbourne, Australia; The University of Texas Southwestern Medical Center, Dallas, TX; Fox Chase Cancer Center, Philadelphia, PA; Helen F Graham Cancer Ctr, Newark, DE; Radiation Therapy Oncology Group, Statistical Center, Philadelphia, PA; Massachusetts General Hospital, Harvard Medical School, Boston, MA. A phase III protocol of androgen suppression (AS) and 3DCRT/IMRT versus AS and 3DCRT/IMRT followed by chemotherapy (CT) with docetaxel and prednisone for localized, high-risk prostate cancer (RTOG 0521). *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2015;33(2015 (suppl; abstr LBA5002)).

37. James ND, Spears MR, Clarke NW, Dearnaley DP, De Bono JS, Gale J, et al. Survival with Newly Diagnosed Metastatic Prostate Cancer in the "Docetaxel Era": Data from 917 Patients in the Control Arm of the STAMPEDE Trial (MRC PR08, CRUK/06/019). *European urology*. 2015;67(6):1028-38.
38. Wilt TJ, Brawer MK, Jones KM, Barry MJ, Aronson WJ, Fox S, et al. Radical prostatectomy versus observation for localized prostate cancer. *The New England journal of medicine*. 2012;367(3):203-13.
39. Pilepich MV, Winter K, Lawton CA, Krisch RE, Wolkov HB, Movsas B, et al. Androgen suppression adjuvant to definitive radiotherapy in prostate carcinoma--long-term results of phase III RTOG 85-31. *International journal of radiation oncology, biology, physics*. 2005;61(5):1285-90.
40. D'Amico AV, Chen MH, Renshaw AA, Loffredo M, Kantoff PW. Androgen suppression and radiation vs radiation alone for prostate cancer: a randomized trial. *Jama*. 2008;299(3):289-95.
41. Roach M, 3rd, Bae K, Speight J, Wolkov HB, Rubin P, Lee RJ, et al. Short-term neoadjuvant androgen deprivation therapy and external-beam radiotherapy for locally advanced prostate cancer: long-term results of RTOG 8610. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2008;26(4):585-91.
42. Bolla M, Van Tienhoven G, Warde P, Dubois JB, Mirimanoff RO, Storme G, et al. External irradiation with or without long-term androgen suppression for prostate cancer with high metastatic risk: 10-year results of an EORTC randomised study. *The Lancet Oncology*. 2010;11(11):1066-73.
43. Denham JW, Steigler A, Lamb DS, Joseph D, Turner S, Matthews J, et al. Short-term neoadjuvant androgen deprivation and radiotherapy for locally advanced prostate cancer: 10-year data from the TROG 96.01 randomised trial. *The Lancet Oncology*. 2011;12(5):451-9.
44. Bratt O, Folkvaljon Y, Hjalmar Eriksson M, Akre O, Carlsson S, Drevin L, et al. Undertreatment of Men in Their Seventies with High-risk Nonmetastatic Prostate Cancer. *European urology*. 2015;68(1):53-8.
45. Stephenson AJ, Scardino PT, Kattan MW, Pisansky TM, Slawin KM, Klein EA, et al. Predicting the outcome of salvage radiation therapy for recurrent prostate cancer after radical prostatectomy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2007;25(15):2035-41.
46. Wilt TJ, Brawer MK, Barry MJ, Jones KM, Kwon Y, Gingrich JR, et al. The Prostate cancer Intervention Versus Observation Trial:VA/NCI/AHRQ Cooperative Studies Program #407 (PIVOT): design and baseline results of a randomized controlled trial comparing radical prostatectomy to watchful waiting for men with clinically localized prostate cancer. *Contemporary clinical trials*. 2009;30(1):81-7.
47. Bill-Axelsson A, Holmberg L, Garmo H, Rider JR, Taari K, Busch C, et al. Radical prostatectomy or watchful waiting in early prostate cancer. *The New England journal of medicine*. 2014;370(10):932-42.
48. Klotz L, Zhang L, Lam A, Nam R, Mamedov A, Loblaw A. Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2010;28(1):126-31.

49. Sawyers CL. The cancer biomarker problem. *Nature*. 2008;452(7187):548-52.
50. Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level ≤ 4.0 ng per milliliter. *The New England journal of medicine*. 2004;350(22):2239-46.
51. Catalona WJ, Partin AW, Sanda MG, Wei JT, Klee GG, Bangma CH, et al. A multicenter study of [-2]pro-prostate specific antigen combined with prostate specific antigen and free prostate specific antigen for prostate cancer detection in the 2.0 to 10.0 ng/ml prostate specific antigen range. *The Journal of urology*. 2011;185(5):1650-5.
52. Bussemakers MJ, van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HF, Schalken JA, et al. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer research*. 1999;59(23):5975-9.
53. Hessels D, Klein Gunnewiek JM, van Oort I, Karthaus HF, van Leenders GJ, van Balken B, et al. DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. *European urology*. 2003;44(1):8-15; discussion -6.
54. site. UFaDAW. Progenssa PCA3 assay 2012. Available from: <http://www.fda.gov/MedicalDevices/ucm294907.htm>.
55. Tosoian JJ, Loeb S, Kettermann A, Landis P, Elliot DJ, Epstein JI, et al. Accuracy of PCA3 measurement in predicting short-term biopsy progression in an active surveillance program. *The Journal of urology*. 2010;183(2):534-8.
56. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*. 2005;310(5748):644-8.
57. Leyten GH, Hessels D, Jannink SA, Smit FP, de Jong H, Cornel EB, et al. Prospective multicentre evaluation of PCA3 and TMPRSS2-ERG gene fusions as diagnostic and prognostic urinary biomarkers for prostate cancer. *European urology*. 2014;65(3):534-42.
58. Epstein JI, Walsh PC, Carmichael M, Brendler CB. Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer. *Jama*. 1994;271(5):368-74.
59. Chaux A, Peskoe SB, Gonzalez-Roibon N, Schultz L, Albadine R, Hicks J, et al. Loss of PTEN expression is associated with increased risk of recurrence after prostatectomy for clinically localized prostate cancer. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 2012;25(11):1543-9.
60. Reid AH, Attard G, Ambrosine L, Fisher G, Kovacs G, Brewer D, et al. Molecular characterisation of ERG, ETV1 and PTEN gene loci identifies patients at low and high risk of death from prostate cancer. *British journal of cancer*. 2010;102(4):678-84.
61. Cullen J, Rosner IL, Brand TC, Zhang N, Tsiatis AC, Moncur J, et al. A Biopsy-based 17-gene Genomic Prostate Score Predicts Recurrence After Radical Prostatectomy and Adverse Surgical Pathology in a Racially Diverse Population of Men with Clinically Low- and Intermediate-risk Prostate Cancer. *European urology*. 2015;68(1):123-31.
62. Klein EA, Cooperberg MR, Magi-Galluzzi C, Simko JP, Falzarano SM, Maddala T, 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*. 2014;66(3):550-60.

63. Cooperberg MR, Pasta DJ, Elkin EP, Litwin MS, Latini DM, Du Chane J, et al. The University of California, San Francisco Cancer of the Prostate Risk Assessment score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy. *The Journal of urology*. 2005;173(6):1938-42.
64. Cuzick J, Swanson GP, Fisher G, Brothman AR, Berney DM, Reid JE, 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*. 2011;12(3):245-55.
65. Cuzick J, Stone S, Fisher G, Yang ZH, North BV, Berney DM, et al. Validation of an RNA cell cycle progression score for predicting death from prostate cancer in a conservatively managed needle biopsy cohort. *British journal of cancer*. 2015;113(3):382-9.
66. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, et al. Embryonic stem cell lines derived from human blastocysts. *Science*. 1998;282(5391):1145-7.
67. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006;126(4):663-76.
68. Glinsky GV. "Stemness" genomics law governs clinical behavior of human cancer: implications for decision making in disease management. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2008;26(17):2846-53.
69. Standardization tIOF. ISO13485:2003, Medical devices-Quality management systems-Requirements for regulatory purposes 2003. Available from: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=36786.
70. Altman NS. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician*. 1992;Vol. 46(No. 3. (Aug., 1992)):11.
71. Sperger JM, Chen X, Draper JS, Antosiewicz JE, Chon CH, Jones SB, et al. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(23):13350-5.
72. Shyamsundar R, Kim YH, Higgins JP, Montgomery K, Jorden M, Sethuraman A, et al. A DNA microarray survey of gene expression in normal human tissues. *Genome biology*. 2005;6(3):R22.
73. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America*. 2001;98(9):5116-21.
74. Parry RM, Jones W, Stokes TH, Phan JH, Moffitt RA, Fang H, et al. k-Nearest neighbor models for microarray gene expression analysis and clinical outcome prediction. *The pharmacogenomics journal*. 2010;10(4):292-309.