U.PORTO

MESTRADO ONCOLOGIA

ONCOLOGIA LABORATORIAL

PROSTYPING: Unraveling Novel Biomarkers for Prostate Cancer Risk Stratification

Ângela Aguiar Albuquerque Castro

M

2023

U.PORTO

Ângela Aguiar Albuquerque Castro. PROSTYPING: Unraveling Novel Biomarkers for Prostate Cancer Risk Stratification

M.ICBAS 2023

PROSTYPING: Unraveling Novel Biomarkers for Prostate Cancer Risk Stratification

Ângela Aguiar Albuquerque Castro

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

Ângela Aguiar Albuquerque Castro

PROSTYPING: Unraveling Novel Biomarkers for Prostate Cancer Risk Stratification

Dissertação de Candidatura ao grau de **Mestre em Oncologia** – Especialização em Oncologia Laboratorial submetida ao Instituto de Ciências Biomédicas de Abel Salazar da Universidade do Porto

Orientadora: **Professora Doutora Carmen de Lurdes Fonseca Jerónimo**
Professora Catedrática Convidada do
Departamento de Patologia e Imunologia Molecular
Instituto de Ciências Biomédicas Abel Salazar - Universidade do Porto
Diretora do Centro de Investigação, Coordenadora e Investigadora Auxiliar do
Grupo de Epigenética e Biologia do Cancro do
Centro de Investigação
Instituto Português de Oncologia do Porto Francisco Gentil, E.P.E

Coorientador: **Professor Doutor Rui Manuel Ferreira Henrique**
Professor Catedrático Convidado do
Departamento de Patologia e Imunologia Molecular
Instituto de Ciências Biomédicas Abel Salazar - Universidade do Porto
Diretor e Assistente Graduado Sénior do Serviço de Anatomia Patológica e
Investigador Sénior do Grupo de Epigenética e Biologia do Cancro do
Centro de Investigação
Instituto Português de Oncologia do Porto Francisco Gentil, E.P.E

“Try and leave this world a little better than you found it, and when your turn comes to die, you can die happy in feeling that at any rate, you have not wasted your time but have done your best.”

Robert Baden-Powell

This study was funded by a grant of the Research Centre of Portuguese Oncology Institute of Porto (PI27-FB-GEBC_CI-IPOP-2016)

AGRADECIMENTOS

E chegou ao fim. Dois anos que passaram a correr. Em primeiro lugar quero agradecer à Professora Carmen Jerónimo, minha orientadora, por logo no primeiro ano desta aventura me ter acolhido e me ter dado a oportunidade de crescer enquanto estudante na investigação. Esteve sempre disponível para me receber e ajudar, discutir ideias novas tendo-me sempre em consideração. Ao meu co-orientador, Professor Rui Henrique, pelo espírito crítico que adicionou a este projeto, pela ajuda em trazer a investigação para a clínica. Ao João Lobo tenho a agradecer toda a disponibilidade, mesmo com mil e uma coisas para fazer, conseguiu ter um papel imprescindível no desenvolvimento deste projeto.

Aos membros do GEBC, obrigada por me terem recebido tão bem. Obrigada pelos lanches, pelas horas de almoço, pelos "Vamos lá fora", por todas as sessões de zumba e karaoke. Tornaram este ano mais fácil. Agradeço com especial carinho à (Ana Catarina) Macedo. Aturas-me desde o início desta jornada, ensinaste-me e continuas a ensinar tanto. Contigo tornei-me mais autónoma, desafiaste-me a ser melhor e sem ti não era possível ter chegado até aqui. Obrigada!

Um obrigada gigante ao Portoing. A Mariana, a pessoa que mais ouviu os meus desabafos, as minhas preocupações. Pelo verão em que levantámos mais de 1000 lâminas e blocos (eternamente grata), pelos workshops the Excel e Word, por todos os lanches, passeios, conversas…Viemos de Aveiro juntas, mas foi no Porto que crescemos juntas. Obrigada por tudo!! À Bea que se juntou a nós este ano, aos nossos almoços, à nossa lista gigante de rooftops para visitar, de museus onde ir, de cafés para experimentar.

À Raquel, uma surpresa inesperada que veio no primeiro ano de mestrado. O primeiro e emocionalmente o mais desafiante, ajudaste imenso, nem sabes o quando. Fizeste do Porto casa, tiveste lá sempre, apoiaste-me sempre.

Às S.O.S., o grupo que surgiu da necessidade e se tornou um porto seguro. À Inês, Carlota, Catarina, Filó, Joana e Rita. A vocês que ouvem podcasts de devaneios meus enquanto cozinho, que dão opinião sobre moda, ciência e política. Que ouvem os meus dilemas e crises existenciais, obrigada por estarem sempre lá. Que nunca faltem *brinners* e encontros anuais (creio que ainda temos uma visita a Braga para agendar este ano). Sou uma sortuda por vos ter comigo. ADORO-VOS!!

Sarita, a ti que estás comigo há 17 anos, que fizeste este caminho comigo desde o início, um gigante obrigada. Fico genuinamente feliz por tudo o que já conquistaste, mereces tudo

de bom. És a pessoa mais genuína e altruísta que conheço. Sou muito grata por te ter na minha vida. Que comecem os nossos mil planos de viagens…

Para a família:

Tio Carlos e Teresa, vocês merecem também um agradecimento especial. Por todas as conversas, conselhos e pela força que me deram ao longo deste percurso.

À minha Mariazinha, sobrinha e afilhada, que me faz rir em todas as horas…A luz que que precisávamos e não sabíamos. A minha "cara chapada". Que gosto é ver-te crescer…

Aos meus pilares, avós, pais e irmã, que me viram crescer e acompanharam todo este percurso atribulado. Que celebraram cada conquista como deles, que superaram cada adversidade comigo. Ao ano mais desafiante, sem vocês nada disto era possível. Não tenho como agradecer o esforço que fizeram nos últimos meses para hoje estarmos aqui a celebrar. Esta dissertação é tanto minha como vossa!!

RESUMO

Introdução: O cancro da próstata é a segunda neoplasia mais comum nos homens em todo o mundo e, devido ao intenso rastreio com PSA, cerca de 90% dos tumores diagnosticados estão confinados ao órgão. Na prática clínica, a decisão do tratamento primário é baseada na avaliação do risco clínico ou no CAPRA *score*. No entanto, a previsão do prognóstico dos doentes continua a ser desafiante, especialmente em doentes de baixo risco. Assim, torna-se necessário a implementação de novos sistemas de classificação molecular e a identificação de novos biomarcadores de prognóstico, no diagnóstico deste cancro. Especificamente, os biomarcadores moleculares poderão ser relevantes para aumentar a acuidade das decisões terapêuticas, distinguindo cancro da próstata indolente das formas clinicamente relevantes, dado a sua agressividade.

Materiais e métodos: Neste estudo, a imunoexpressão dos marcadores luminais, NKX3.1 e CK8/28, dos marcadores basais, CK5/6, CK14 e CD49f, bem como do marcador de proliferação celular, KI-67, foram avaliadas por imunohistoquímica numa série de 103 amostras de prostatectomias radicais, sem qualquer tratamento prévio. Os níveis de metilação do gene *KLF8* foram também avaliados por PCR quantitativo específico de metilação, na mesma coorte. Os *endepoints* do estudo incluíram a sobrevivência global e livre de recorrência bioquímica. Por fim, foram desenvolvidos nomogramas representativos de calculadoras de risco, incorporando os parâmetros clinico-patológicos relevantes juntamente com a imunoexpressão do Ki-67 e os níveis de metilação do *KLF8*.

Resultados: Na coorte de doentes em estudo, não foram encontradas diferenças na sobrevivência global ou na sobrevivência livre de recorrência bioquímica entre tumores *luminal-like* e tumores do *basal-like*. Notavelmente, o índice Ki-67 mais elevado foi associado a uma menor sobrevivência livre de recorrência, tendo sido verificada a mesma associação nos doentes com tumores de baixo grau (1 e 2). Por outro lado, níveis mais elevados de metilação do *KLF8* associaram-se significativamente a uma melhor sobrevivência global em doentes com tumores de baixo grau. O desempenho destes marcadores foi avaliado no contexto de uma calculadora de risco, considerando outras variáveis clínicas convencionais, designadamente o PSA, o estadio clínico e a idade dos doentes ao diagnóstico. Estes modelos demonstraram um bom desempenho, com uma área sob a curva (AUC) superior a 0,75 em todas as simulações de análise realizadas. Adicionalmente, foram obtidos resultados similares quando a mesma calculadora de risco de recorrência foi aplicada a uma série independente de 153 biópsias de diagnóstico de cancro da próstata.

Discussão e conclusão: Este estudo permitiu verificar que a subclassificação molecular dos tumores da próstata em *luminal-like* e *basal-like* descrita na literatura não tem aplicabilidade na prática clínica quando avaliados os níveis de proteína dos marcadores característicos de células luminais e basais. No entanto, o índice Ki-67 e os níveis de metilação do *KLF8* emergem como biomarcadores independentes com capacidade para estratificação de doentes, especialmente em doentes de baixo risco, os quais necessitam de uma estratégia mais precisa para a tomada de decisão terapêutica. Desta forma, a implementação da calculadora de risco de recorrência testada seria de extrema utilidade na prática clínica, dado permitir distinguir cancro da próstata indolente de formas mais agressivas, possibilitando uma decisão terapêutica mais ajustada a cada doente.

ABSTRACT

Background: Prostate cancer is the second most common malignancy in men worldwide, and due to the PSA screening nearly 90% of the diagnosed tumors are organ-confined. In clinical practice, the primary treatment decision is based on the clinical risk assessment or CAPRA score. Nonetheless, it is still challenging to predict patients' prognosis, specifically in low-risk tumors. Therefore, the implementation of new molecular classifications and the identification of new prognostic biomarkers in the diagnosis of this cancer is needed. Particularly, the molecular biomarkers could be valuable to improve the accuracy of the therapeutic decision-making, distinguishing indolent PCa from the clinic relevant forms, due to their aggressiveness.

Material and methods: In this study, the immunoexpression of the luminal markers, NKX3.1 and CK8/18, the basal markers, CK5/6, CK14, and CD49f, as well as the proliferation marker, Ki-67, were accessed by immunohistochemistry in a series of 103 specimens from radical prostatectomies, without any prior treatment. The methylation levels of the *KLF8* gene were also evaluated, by quantitative methylation-specific PCR, in the same cohort. Study endpoints included overall and BCR-free survival. Subsequently, we assembled nomograms that serve as risk calculators, incorporating relevant clinico-pathological parameters alongside Ki-67 immunoexpression and *KLF8* methylation levels.

Results: In the study cohort, no differences were found in the overall survival or biochemical-recurrence free survival between luminal-like and basal-like tumors. Remarkably, higher Ki-67 index was associated with lower BCR-free survival, and the same association was found in lower-grade tumors (1 and 2). Conversely, higher *KLF8* methylation levels were significantly associated with improved overall survival in patients with lower-grade tumors. The performance of these markers was evaluated in the context of a risk calculator, considering other conventional clinical variables, namely PSA, T stage, and patients age at diagnosis. Markedly, these models demonstrated good performance with an area under the curve (AUC) exceeding 0.75 across all analysis simulations carried out. Additionally, similar encouraging results were obtained when the BCR calculator was applied to an independent series of 153 PCa diagnostic biopsies.

Discussion and conclusion: This study allows us to verify that the molecular subclassification of prostate tumors into luminal-like as basal-like, described in the literature, does not have applicability in clinical practice when assessing the protein levels of the luminal and basal characteristic markers. Nonetheless, Ki-67 index and *KLF8* methylation levels emerge as independent prognostic biomarkers with the ability for patient stratification,

especially in lower-risk patients, which require a better tool for therapeutic decision-making. Henceforth, the implementation of the risk of recurrence calculator tested would be extremely useful in the clinical practice to distinguish indolent PCa from more aggressive forms, allowing a more adjusted treatment decision for each patient.

INDEX OF CONTENT

INDEX OF FIGURES

INDEX OF TABLES

ABBREVIATIONS

PCa – Prostate Cancer

DNA – Deoxyribonucleic Acid

DRE – Digital Rectal Examination

PSA – Prostate Specific Antigen

TRUS – Transrectal Ultrasound

mpMRI – Multiparametric Magnetic Resonance Imaging

BCR – Biochemical Recurrence

TNM – Tumor Node Metastasis

CT – Computed Tomography

MRI – Magnetic Resonance Imaging

GS – Gleason Score

ISUP – International Society of Urological Pathology

GG – Grade Groups

PIN – Prostatic Intraepithelial Neoplasia

mCRPC – Metastatic Castration-Resistant Prostate Cancer

RP – Radical Prostatectomy

RT – Radical Radiotherapy

ADT – Androgen Deprivation Therapy

CRPC – Castration-Resistant Prostate Cancer

NEPC – Neuroendocrine Prostate Cancer

WHO – World Health Organization

AR – Androgen Receptor

AR-V – Splice Variants of AR

LBD – Ligand-Binding Domain

CAPRA – Cancer of the Prostate Risk Assessment

RT-PCR – Reverse Transcription Polymerase Chain Reaction

RNA – Ribonucleic Acid

DNMT – DNA Methyltransferase

5mC – 5-Methylcytosine

SAM – S-Adenyl Methionine

CpG – Cytosine phosphatase Guanine

TET – Ten-Eleven Translocation

FFPE – Formalin-Fixed Paraffin-Embedded

TMA – Tissue Microarray

IHC – Immunohistochemistry

DAB – 3,3'-Diaminobenzidine

TCGA – The Cancer Genome Atlas

SMART – Shiny Methylation Analysis Resource Tool

qMSP – Quantitative Methylation Specific PCR

NTC – No Template Controls

ROC – Receiver Operator Characteristic

AUC – Area Under the Curve

KLF8^{me} – *KLF8* methylation

EMT – Epithelial-Mesenchymal Transition

GENERAL INTRODUCTION

PROSTATE CANCER EPIDEMIOLOGY AND ETIOLOGY

Prostate cancer (PCa) is the second most frequent cancer and the fifth leading cause of cancer-related death in men worldwide, with an estimation of 1.4 million new cases and 375,000 deaths in 2020 (Figure 1). Unlike mortality rates, the incidence rates greatly varies among different continents, with the highest rates found in Northern and Western Europe and the lowest rates found in Asia and Northern Africa. These variations are mainly due to differences in the diagnostic practices and men life span. Despite PCa being more prevalent in high-income countries, little is known about the etiology of this disease. Advanced age, African origin, family history, and certain genetic mutations remain the only established risk factors associated with PCa (1). Compared with Caucasian men, PCa is particularly aggressive and fatal for Africans and African Americans (2). Men with a family history of PCa show a greater risk of developing this disease during life (3). Indeed, certain gene alterations are associated with heritable PCa, being the majority related with DNA (Deoxyribonucleic Acid) damage repair machinery. These alterations include germline *BRCA1* and *BRCA2* mutations which are associated with a significant increased risk of the PCa development. Men with germline *BRCA2* mutations are usually diagnosed at a younger age and have more aggressive tumors, with a poorer median survival than men with a sporadic type of the disease. Genes involved in mismatch repair, like *MLH1*, *MSH2*, *MSH6*, and *PMS2*, can be also altered, being drivers for the Lynch Syndrome. Although this syndrome is strongly associated with colorectal, gynecologic, and stomach cancers, men with Lynch Syndrome have also an increased risk for PCa (4). On the other hand, a few modifier factors related with lifestyle, such as smoking and excess body weight, may increase the risk of developing advanced disease (1).

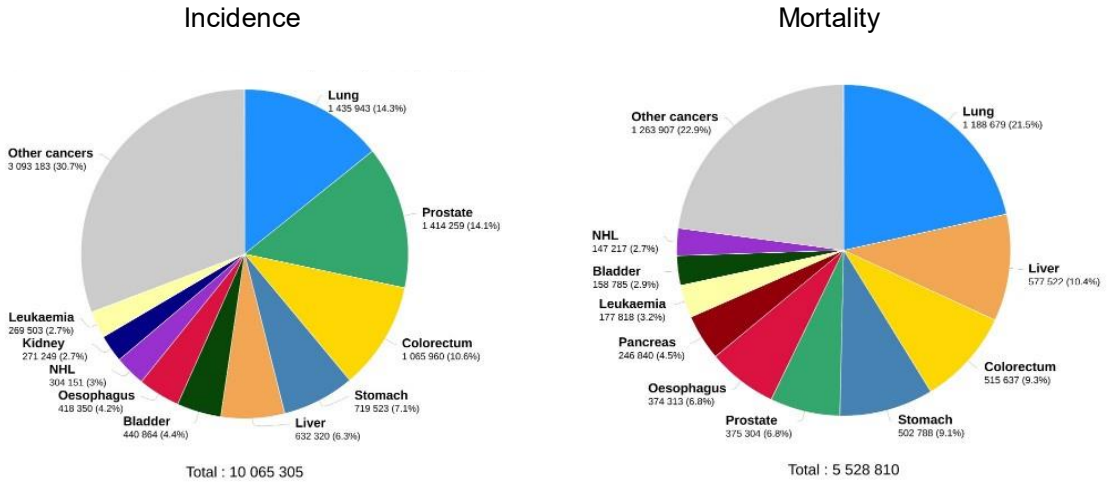


Figure 1- Graphic representation of the 10 most common cancers among men worldwide, regarding their incidence and mortality rates, in 2020. Adapted from (Sung, 2021). Prostate cancer was the second most common malignancy in men and the fifth leading cause of cancer-related death.

PROSTATE CANCER DIAGNOSTIC

Screening and early detection

PCa screening consists in a systematic examination of asymptomatic aging men using diagnostic tools such as digital rectal examination (DRE), prostate-specific antigen (PSA) blood test and transrectal ultrasound (TRUS) guided biopsy, in men over 50 years old. An organized screening program for PCa is currently not recommended in most countries, remaining a controversial topic within the urologic community (3). In 2017, the US Preventive Services Task Force recommendation for PSA-based screening for PCa was updated. Men aged as 55-69 years old must be informed of the benefits and harms of PSA-based screening before any decision, which might reduce overdiagnosis and overtreatment. Nonetheless, for men over 70 years old, PSA-based screening is discouraged. (5).

Digital Rectal Examination

DRE is often used as a diagnostic tool to evaluate the presence of prostate gland malignancy. Although this method is easy to perform and minimally invasive, it is subjective, relying on the experience and skills of the examiner in determining the size, consistency, nodularity, and asymmetry of the prostate (6, 7).

Moreover, DRE cannot be used as a single and definitive diagnostic tool, since only the posterior and lateral surfaces of the prostate can be reached during the examination. Besides that, this method cannot exclude the presence of malignancy due to its low sensitivity, below 60% (8, 9).

PSA-based screening

PSA is a serine protease regulated by androgens. This protein is produced by secretory epithelial cells from the prostate glandular tissue and then, released in seminal fluid with a normal circulating concentration of 0.5 to 2.0 ng/mL. (10).

In prostate carcinogenesis, the basal cell layer and basement membrane are disrupted allowing PSA releasing. Suspicious of PCa is considered when serum PSA values are greater than 4 ng/mL. Although serum PSA can be used to identify PCa patients, its use as a screening tool is limited due to the lack of specificity, since PSA can also be increased in benign lesions as prostatic hyperplasia and some inflammatory conditions (10, 11). Moreover, the use of PSA as screening tools can lead to unnecessary biopsies, resulting in overdiagnosis and overtreatment of indolent diseases. Conversely, men with normal PSA can be missed diagnoses (12).

Prostate Biopsy

The histopathological verification based on microscopic evaluation of the prostate tissue obtained by needle biopsy is necessary for a definitive diagnosis of PCa. A biopsy is recommended when other pre-screening tools detect some suspicious lesion (3). Prostate biopsy can be performed by either a transrectal or transperineal approach (13). Most of the diagnosed tumors are localized in the peripheral region of the prostate. Therefore, the systematic biopsies, regardless of the approach used, should be directed to the peripheral gland bilaterally from apex to base as far posterior and lateral as possible (14). By convention, 10 to 12 systematic biopsies are recommended depending on the prostate size (15). Although systematic prostate biopsy remains the standard of approach for definitive PCa diagnosis, it misses 21 to 28% of all PCa, and about 14% of the detected tumors could be under-graded (15, 16). To overcome these issues, imaging techniques have been adapted to improve diagnostic performance, being multiparametric magnetic resonance imaging (mpMRI) the most used one (15).

CLASSIFICATION AND STAGING

To better decide the primary PCa treatment, the risk stratification for biochemical recurrence (BCR) is assessed, taking into account the PSA serum values, as well as the clinical staging (TNM classification) and the biopsy grading into low-, intermediate-, and high-risk based on the Gleason Grading System (17).

In the TNM classification, the local stage of PCa is based on DRE findings, being essential to define the extraprostatic extension of the disease, as well as circumvent the limits of the tumor, therefore guiding the surgical decisions and additional therapeutic management (18, 19). Further, imaging techniques, such as computed tomography (CT) or magnetic resonance imaging (MRI), are needed to predict lymph node involvement in intermediate or high-risk patients. Bone scintigraphy can be also performed to confirm bone metastasis in high-risk or symptomatic patients (18).

Due to the heterogeneity of PCa, the Gleason Score (GS) is used to graduate the tumor, being the Gleason grading system the most relevant predictor of PCa prognosis. This system is based on histopathological features of PCa tissues. After pathologic evaluation, the two most abundant patterns of the tumor are graded in 1 through 5 and added to form the Gleason Score, ranging from 2 to 10 (20). Since tumors with a GS lower than 6 are not clinically significant, in 2014, the International Society of Urological Pathology (ISUP) conference proposed a new grading system aiming to decrease the overtreatment of low-grade PCa (17, 20). This 2014 ISUP grading system limits the number of PCa classification grades, resulting in 5 prognostically distinct ISUP grades, also called grade groups (GG) (Table 1) (17).

Table 1- The International Society of Urological Pathology (ISUP) 2014 grade system

ISUP Grade	Gleason Score	Gleason Pattern
1	≤6	≤3+3
2	7	3+4
3	7	4+3
4	8	4+4; 3+5; 5+3
5	9 or 10	4+5; 5+4; 5+5

This new grade system limits the PCa classification grades, based on Gleason Score. Thus, PCa is classified into one of the five prognostically distinct grade groups. Adapted from (Epstein, 2016).

BIOLOGY AND MANAGEMENT OF PROSTATE TUMORS

The carcinogenic process of the prostate gland occurs during a long period and commonly initiate as prostatic intraepithelial neoplasia (PIN), followed by localized PCa and locally advanced prostate adenocarcinomas. The locally advanced prostate adenocarcinomas can progress into a castration-resistant phenotype with or without neuroendocrine features, ultimately leading to the development of metastasis. Similarly, hormone-naïve metastatic tumors eventually progress to metastatic castration-resistant prostate cancer (mCRPC) (Figure 2) (21).

The treatment guidelines for PCa are based on PSA serum concentration, clinical stage of the disease, and histological grade. Regarding primary PCa, these parameters are determined to decide between active surveillance, for the more indolent forms of the disease, or active treatment, in a single or multimodality approach (22). Radical prostatectomy (RP), radical radiotherapy (RT), or brachytherapy can be potentially curable for localized PCa. Conversely for high-risk and locally advanced diseases, neoadjuvant and concomitant androgen-deprivation therapy (ADT) should be considered (22-24). Metastatic disease is not curable, however, can be managed to slow down the progression and downstage the tumor burden. Indeed, ADT is effective for hormone-naïve metastatic PCa (22, 23, 25). However, the disease may evolve to a more aggressive behavior, arising castration-resistant PCa (CRPC) or even neuroendocrine PCa (NEPC). Unfortunately, no curative-intent treatment options are available so far for those aggressive forms, and chemotherapy is often used as a palliative treatment (23-25).

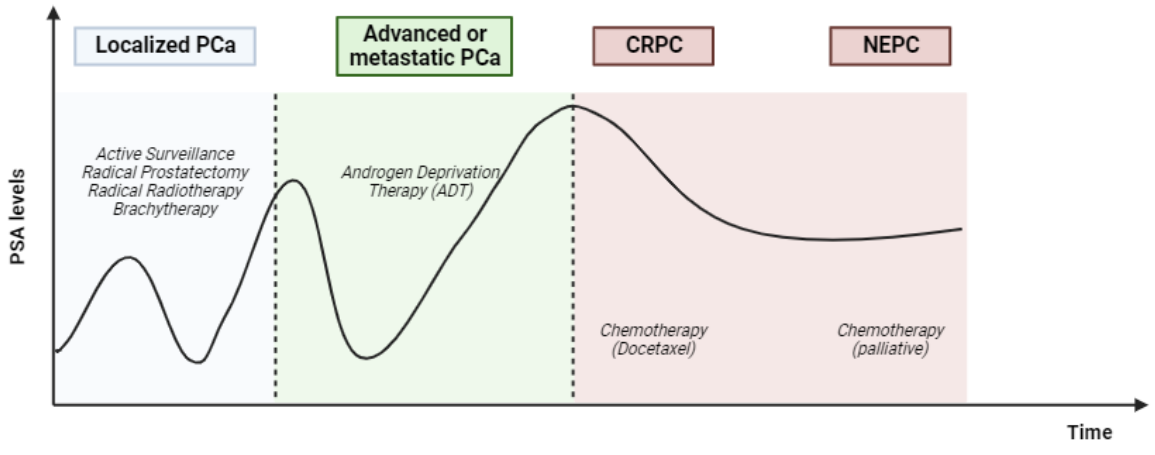


Figure 2- Therapeutical approaches according to PCa different stages. PSA levels vary with the course of the disease. For localized PCa, RP is the standard of care, nonetheless, patients with low-risk tumors may benefit from active surveillance. As for advanced or hormone-naïve metastatic disease, ADT is recommended. Unfortunately, after some time on ADT, the disease progress, arising CRPC or even NEPC. At this point, no curative-intent treatments are available, being chemotherapy used as a palliative treatment. **Abbreviations:** PCa – prostate cancer; RP – radical prostatectomy; ADT – androgen deprivation therapy; CRPC – castration resistant prostate cancer; NEPC – neuroendocrine prostate cancer. Created with Biorender.

Localized prostate cancer

Due to the intense PSA screening, nearly 90% of all PCa tumors are locally-confined at the time of diagnosis (26). Localized PCa arises from the uncontrolled proliferation of luminal cells with loss of the basal layer, and it is restricted to the prostate gland (27). The majority of PCa arise from the peripheral region of the prostate. Nonetheless, PCa is a very heterogeneous disease with a multifocal nature, with 80% of primary PCa displaying divergent morphology and topography (19, 28).

Histologically, the majority of PCa are acinar adenocarcinomas (19, 29, 30). However, in 2016, the World Health Organization (WHO) described a new histological type classified as intraductal carcinoma of the prostate. These cases are characterized by the proliferation of neoplastic cells within the ducts and acini, with the formation of a solid or dense cribriform pattern, involving at least 50% of the lumen (29, 31). Intraductal carcinoma is associated with more aggressive PCa, and Gleason Grade is not applicable since this type of tumor is linked to a high-grade and high-volume carcinoma (31).

Genomic landscape

PCa harbors high genomic complexity, with multiple recurrent genomic alterations, that were reported to be useful for molecular subtyping. These genomic alterations include mutations, DNA copy-number alterations, rearrangements, and gene fusions (26, 32). The most frequent genomic alteration in PCa is the fusion of the androgen-regulated promoters with members of the ETS family. In particular, nearly 50% of all prostate tumors display the *TMPRSS2-ERG* fusion (26). This fusion is an early event in PCa carcinogenesis but is not sufficient to initiate PCa, requiring other genomic alterations, like *PTEN* deletions, *TP53* alterations, and *RB1* deletions (33, 34).

Regarding somatic point mutations, *SPOP*, *TP53*, *FOXA1*, and *PTEN* are the most frequent mutated genes in primary PCa (26). *SPOP* mutations are present in 10-15% of localized prostate tumors and are mutually exclusive of tumors with ETS rearrangements. Androgen receptor (AR) and its' co-factors are SPOP substrates in PCa. Likewise, dysregulation of *SPOP* is associated with increased *AR* transcriptional activity and AR target gene expression (26, 33).

Locally advanced and metastatic prostate cancer

Locally advanced PCa is characterized by an extension of the tumor outside the prostate gland with disruption of the prostatic capsule, with or without invasion of the seminal vesicles, facilitating the invasion of adjacent lymph nodes and distant organs (35).

Metastatic PCa is a leading cause of PCa-related deaths, and about 80% of metastatic patients experience bone metastasis, in which cancer cells directly interact with the bone microenvironment. Besides bone, metastasis can also be formed in other distant organs like liver and lungs (21, 36).

The genetic landscape is even more complex in metastatic tumors than in the primary tumors, due to the high selective pressure in this stage of the disease. Indeed, some alterations promote tumor progression and invasion (33). For instance, *PTEN* loss is associated with more aggressive tumors and increased risk for metastization. Moreover, *MYC* activation with *PTEN* loss led to increased genomic instability in metastatic PCa (21).

Castration-resistant prostate cancer

CRPC occurs with the progression of the disease, mostly due to ADT resistance. Nearly all patients on ADT, eventually progress to a castration-resistant phenotype (37). The CRPC diagnosis is considered when the PSA serum levels are greater than 2 ng/mL, in two consecutive measurements, in a castration environment, while testosterone levels are maintained below 50 ng/dL (38).

Some genetic alterations are implicated in the transformation of locally advanced PCa to a CRPC phenotype. *AR* is very often amplified and mutated in these samples, leading to consequent alterations in AR signaling pathways (26, 39). Additionally, about 50% of metastatic castration-resistant prostate cancer (mCRPC) cases have *TP53* loss and 15-35% have alterations in DNA repair machinery (21, 33, 40). Moreover, PI3K signaling alterations as well as deletions or mutation of *RB1*, *KMT2C*, and *KMT2D* are also frequently found in mCRPC (26).

Notably, most mechanisms that lead to a CRPC phenotype are related to *AR* signaling pathway. Despite the androgen blockade by the ADT, low levels of androgens persist. Thus, a subset of cells can adapt through *AR* amplification or *AR* mutations, becoming more sensitive to the low levels of androgens in circulation (41). Nonetheless, some resistant mechanisms are independent of the AR ligand. Particularly, some growth factors, cytokines, and kinases are capable of increasing AR signaling, promoting disease progression. Moreover, the persistent castration by ADT can lead to the development of splice variants of AR (AR-V). These variants have lost the C-terminal of the ligand-binding domain (LBD), being constitutively activated (41, 42).

Thus, for these patients, androgen-receptor pathway target therapies, such as enzalutamide and abiraterone, are considered first-line treatment in asymptomatic patients. Docetaxel should be considered as well, for both symptomatic and asymptomatic patients.

For patients who acquired resistance to first-line treatment with docetaxel, enzalutamide and abiraterone should be used as second-line treatments. Additionally, taxanes are also a second-line treatment option for men who had primary resistance to first-line treatment with enzalutamide or abiraterone (24). These therapeutic approaches can slow down the progression of the disease, however, all CRPC patients eventually gain resistance to therapy (41).

Neuroendocrine prostate cancer

Neuroendocrine PCa is a lethal form characterized by the low or negative expression of prostate gland characteristic markers, such as AR and PSA. Less than 1% of the prostate tumors are pure localized small cell carcinoma. Nonetheless, about 5 to 10% of the prostate cancer tumors show focal neuroendocrine differentiation admixed with adenocarcinoma (43). Nearly 20% of CRPC tumors lose their dependence on AR signaling, with a conversion into an AR-negative tumors, poorly differentiated small cell neuroendocrine carcinoma, without an increase in PSA levels (43, 44). NEPC tumors start to express typical neuroendocrine markers, such as enolase 2, chromogranin A, synaptophysin, and CD56. The evaluation of the immunoexpression of these markers is crucial to the NEPC diagnosis (44, 45).

Loss of *TP53* and *RB1* in mCRPC induces cell plasticity to a neuroendocrine phenotype. This process is driven by genetic and epigenetic reprogramming. About 40% of NEPCs display overexpression or amplification of *NMYC*, being a driver for NEPC initiation (21). *EZH2* is also upregulated, especially in tumors with distant metastasis, promoting lineage plasticity (21, 44, 45). The synergic interaction between the *EZH2*, *NMYC*, and *AR* leads to an inhibition of AR-signaling by the formation of the N-MYC/AR/EZH2-PRC2 complex, resulting in neuroendocrine transformation (45). Similarly with adenocarcinomas, *TMPRSS2-ERG* gene fusion is observed in nearly 50% of NEPC (43, 44).

RISK ASSESSMENT OF LOCALIZED PROSTATE CANCER

In the last decade, new strategies for the molecular classification of PCa have been investigated to better predict patients' prognosis (46). Based on RNA expression and transcriptome data, PCa may be classified into three different subtypes based on the characteristics of the basal and luminal epithelial cells. Therefore, like in breast cancer, it was possible to categorize the tumors into a luminal A-like PCa, a luminal B-like PCa, and a basal-like PCa. Luminal B-like PCa displayed the worse outcome (47-49). Despite the promising results of this classification system based on RNAseq data, further research is needed to confirm its practicability in the clinical practice.

The Clinical Risk Assessment remains the better predictor of patients' prognosis and is crucial for the therapeutic guidance. At diagnosis, the patients with PCa are classified for the risk grouped based on similar clinico-pathological characteristics (50, 51). A more accurate way to individually assess the risk of PCa progression is using the Cancer of the Prostate Risk Assessment (CAPRA) score. CAPRA score is a 0 to 10 scale based on PSA levels, GS, clinical tumor stage, age at diagnosis, and the percentage of positive biopsy cores. This tool is currently the most validated and used risk assessment method in localized PCa (50, 52, 53).

High-risk patients are defined as having one of the following characteristics: metastatic spread, GS greater than 8 and very high PSA levels (Table 2) (50, 54, 55). Those parameters are indicative of a more aggressive disease with high susceptibility for disease progression, requiring alternative therapeutic interventions. As for PCa patients with low- (GS 3+3, GG1) and favorable intermediate-risk (GS 3+4, GG2), the choice of the best therapeutic approach remains challenging. The decision between active surveillance and curative treatment is complex, and currently it is difficult to predict which patients may benefit from an interventive approach (50). To overcome this issue, new strategies based on genomic tests, such as the Decipher score®, Prolaris® cell cycle progression test, and Oncotype DX® Prostate Score, have been investigated.

The Decipher score® is a tissue-based microarray classifier that evaluates the expression of 22 genes involved in the AR signaling pathway, cell proliferation, differentiation and motility, and immune modulation. Both biopsy and RP specimens can be used to perform the test, predicting the risk for distant metastasis, and the risk for advanced disease at RP. Although this test can be used to select the patients that may benefit from active surveillance, the Decipher score® is specially indicated to guide the decision between post-prostatectomy adjuvant treatment or salvage radiotherapy (12, 56). The Prolaris® test is a biopsy-based genomic test that measures the expression of 31 cell cycle progression

genes and 15 housekeeping genes, by reverse transcription polymerase chain reaction (RT-PCR), generating a score that provides information about the tumor cells' proliferation. It is validated to predict PCa-specific mortality, as well as to predict the risk of BCR and metastatic spread (50, 57, 58). The Oncotype DX® score is also biopsy-based but measures the expression of 17 genes related to prostate tumorigenesis, involving the AR pathway, cellular organization, proliferation, and stromal response, using quantitative RT-PCR. This test predicts adverse PCa pathology as the probability of advanced disease and higher stage and grade (58-60).

Although these tests provide information that may help in the risk assessment of low- and intermediate-risk PCa, the technology used is very expensive, with the price ranging from 3.900 to 5.150 dollars per patient. Hence, finding new strategies that may use less expensive tools, like immunohistochemistry, seems to be imperative to introduce these risk calculators into the clinical practice (56). Moreover, the patient's preferences along with the clinician's advice are, currently, the more relevant factors in the therapeutic decision, making it necessary a more evidence-based approach to better decide between active surveillance and curative treatment (56, 60).

Table 2- Risk group stratification for biochemical recurrence

Very Low-Risk	Low-Risk	Intermediate-Risk		High-Risk	Very High-Risk
		Favorable	**Unfavorable**		
T1c AND GG1 AND PSA<10ng/mL	T1-T2a AND GG1 AND PSA<10ng/mL	T2b-T2c OR GG2 OR PSA 10-20 ng/mL	T2b-T2c OR GG3 OR PSA 10-20 ng/mL	T3a OR GG4 or GG5 OR PSA>20ng/mL	T3b-T4 OR N+ Any GG Any PSA

According to the National Comprehensive Cancer Network (NCCN) it is possible to find 5 groups with different prognosis, based on clinical tumor stage, GG and PSA serum levels. Specifically, the NCCN guidelines subdivide the intermediate-risk group into favorable intermediate-risk and unfavorable intermediate-risk groups, being the GG the differential parameter. Adapted from (Mohler, 2019).

Role of the proliferation markers in risk stratification and prognosis

Tumor cell proliferation is currently an important independent prognostic factor in a variety of solid tumors, like breast, lung, and colorectal cancers, predicting patients' outcomes (61). Ki-67 is a protein directly involved in cell proliferation, and its immunoexpression evaluation can be used to assess the proliferative status of the tumor cells. This marker can recognize cells in all phases of the cell cycle, except in the G0 phase, therefore, Ki-67 can be used to identify almost every population of cells in proliferation (62-64). Moreover, the Ki-67 evaluation by immunohistochemistry is promising, since is highly

reproducible, easy to perform, and widely used in pathological laboratories (65). Remarkably, several studies reported that Ki-67 expression can predict worse outcomes, in localized prostate tumors. Specifically, high levels of Ki-67 have been associated with poor prognosis, regarding biochemical recurrence, development of distant metastasis, and overall survival (65-67). Thus, the Ki-67 evaluation may help in therapeutic decision-making, particularly for patients considered for active surveillance (67). Nonetheless, for the implementation of this marker in the clinical practice more studies are needed.

EPIGENETICS

Epigenetics comprises multiple mechanisms that can regulate gene expression without changing the DNA sequence. This non-mutational reprogramming is flexible and reversible, with high relevance in embryogenesis and tumorigenesis (68, 69). DNA methylation, histone modifications (acetylation, methylation, ubiquitylation, phosphorylation, and ribosylation), chromatin remodeling complexes, and non-coding ribonucleic acid (RNA) regulation are the main epigenetic mechanisms involved in gene expression regulation (68, 69). These events are capable of reprogramming neoplastic cells, contributing to the acquisition of malignant characteristics during tumor development, allowing its' progression. Thus, the epigenetic landscape in the cell is still considered an important hallmark of cancer (70).

DNA Methylation

DNA methylation is the most studied epigenetic mechanism. This process is catalyzed by DNA methyltransferases (DNMTs) and culminate in the formation of a 5-Methylcytosine (5mC), by the transference of a methyl group from S-Adenyl Methionine (SAM) to the fifth carbon of the cytosine residue (68, 71). The majority of DNA methylation occurs in cytosines that precede a guanine nucleotide, a Cytosine phosphatase Guanine (CpG), or the so-called CpG sites (71). These CpG sites are spread across the whole genome and are often methylated. The CpG islands are defined as a cluster of CpG nucleotides with more than 60% CG content and are found in 70% of the promotor regions of human genes, usually being hypomethylated in active genes (71, 72). In neoplastic cells transformation, a switch in methylation pattern is seen, and the CpG islands became methylated mostly in genes with tumor-suppressive functions, repressing their transcription (Figure 3) (69).

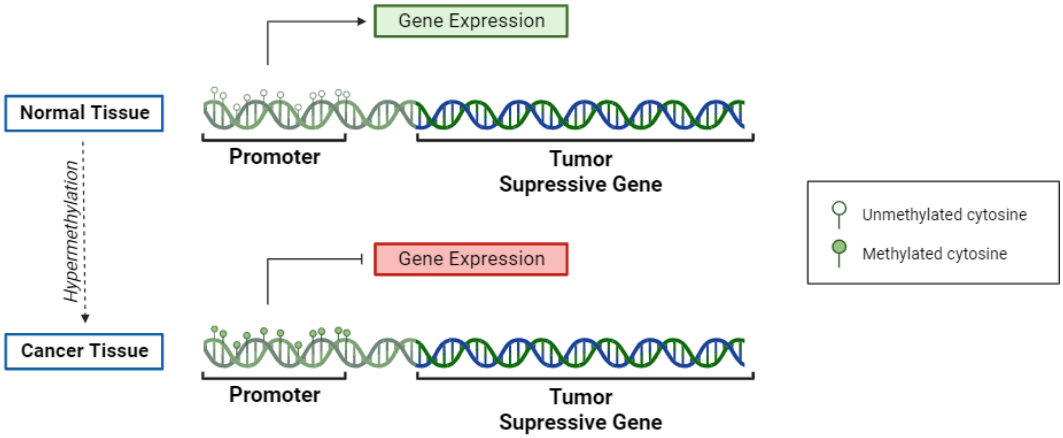


Figure 3- Schematic representation of DNA methylation. DNA methylation is a process catalyzed by DNA methyltransferases. They transfer a methyl group from the S-Adenyl Methionine (SAM) to the fifth carbon of the cytosine residue, forming a 5-methylcytosine. The promoter region of the tumor suppressive genes is usually hypomethylated, allowing the expression of the genes. However, in tumorigenesis, a switch in the methylation pattern is observed, with the hypermethylation of the promoter region of tumor suppressive genes, repressing their transcription. Created with Biorender.

The DNMTs family comprises two main types of enzymes, DNMT1 responsible for the maintenance of the DNA methylation after replication, and DNMT3 responsible for *de novo* methylation. DNMT2 is also part of the DNMT family, however, its function as DNA methyltransferase is limited, having a more significant role in RNA processing (68, 73). DNA methylation can be reverted by the reduction of the DNMTs activity or by a reaction actively catalyzed by a family of enzymes named Ten-Eleven Translocation (TET) (73). In a tumoral context, DNMTs are overexpressed in several solid tumors, leading to a hypermethylated status and oncogenic activation. Specifically, DNMT1 overexpression is associated with lymph node metastasis and consequently poor patients' prognosis (74).

DNA methylation in prostate cancer

In PCa, DNMTs can also be overexpressed, being associated with cancer progression and metastasis. DNMT1 has been associated with therapy resistance by the repression of the androgen receptor gene, promoting an androgen-independent phenotype, like NEPC. As for DNMT3A and DNMT3B, they are found overexpressed in more aggressive tumors (75).

DNA hypermethylation is an early event in PCa carcinogenesis, and genes like *GSTP1*, *APC*, and *RASSF1A* are frequently silenced due to promoter hypermethylation (76). Genomic subtypes can also harbor a specific DNA hypermethylation pattern. For instance, *STAT6* is predominantly epigenetic silenced in ETS-positive tumors, while *HEXA* is aberrantly methylated in *SPOP*-mutant tumors. Conversely, some genes, like *GSTP1* and *KLF8*, are silenced in almost all prostate tumors, being excellent candidates for PCa detection (26). Moreover, *GSTP1* promoter hypermethylation is found in more than 90% of prostate tumors and is associated with disease progression and aggressiveness (77). Despite *KLF8* being hypermethylated in many prostate tumors, its role in PCa carcinogenesis is still poorly understood. Nonetheless, *KLF8* expression has been associated with worse overall survival and disease progression into a hormone-refractory phenotype. Therefore, differences in methylation levels can be helpful in stratifying patients (78).

Additionally, prostate tumor cells can be directly released in the urethra trough the prostatic ducts, therefore, circulating tumor DNA can be found in the urine (79). Thus, the DNA methylation assessment trough liquid biopsies, like urine, is a minimal-invasive technique that can be used as an early diagnostic and monitoring tool (80-82). Taking this into account, DNA methylation assessment seems to be a promising approach to predict patients' prognosis, in prostate cancer.

AIMS

The main goal of this master's dissertation was to determine the practicability of subtyping prostate cancers and identifying prognostic biomarkers. These biomarkers could be valuable to improve therapeutic decision-making, particularly for patients diagnosed with low- or intermediate-risk PCa that currently are spared from an active treatment intervention, but without any well-stablished criteria guidelines, yet. Thus, with this research we aim to contribute with valued insights into refining risk assessment strategies and ultimately improving PCa patients' prognosis and life-quality outcomes.

Overall, this dissertation is organized into two major sections, each reflecting distinct approaches to the applicability of prognostic biomarkers in the clinical practice (Figure 4).

Section 1- Exploratory data

1.1- The practicability of luminal and basal molecular subtyping in PCa

i. Evaluation of the expression of luminal and basal biomarkers as well as the expression of a proliferation marker in a series of localized Prostate Cancer, by immunohistochemistry.

1.2- Specific gene promoter hypermethylation associated with well-known PCa genomic signatures.

ii. *In silico* analysis to unveil the most relevant CpG islands of the *GSTP1*, *KLF8*, *HEXA*, and *STAT6* gene promoter and corroborate their dysregulation in PCa.

Section 2: Ki-67 immunoexpression and $KLF8^{me}$ levels as independent prognostic biomarkers for PCa *(Manuscript under submission)*

1. Evaluation of Ki-67 immunoexpression in a series of localized PCa;
2. Assessment of the methylation levels of *GSTP1* and *KLF8* genes in the same series of localized PCa;
3. Establishment of a risk calculator model based on the clinico-pathological characteristics of the localized PCa patients and the markers with prognostic value;
4. Validation of the risk calculator model in an independent series of localized PCa biopsy samples.

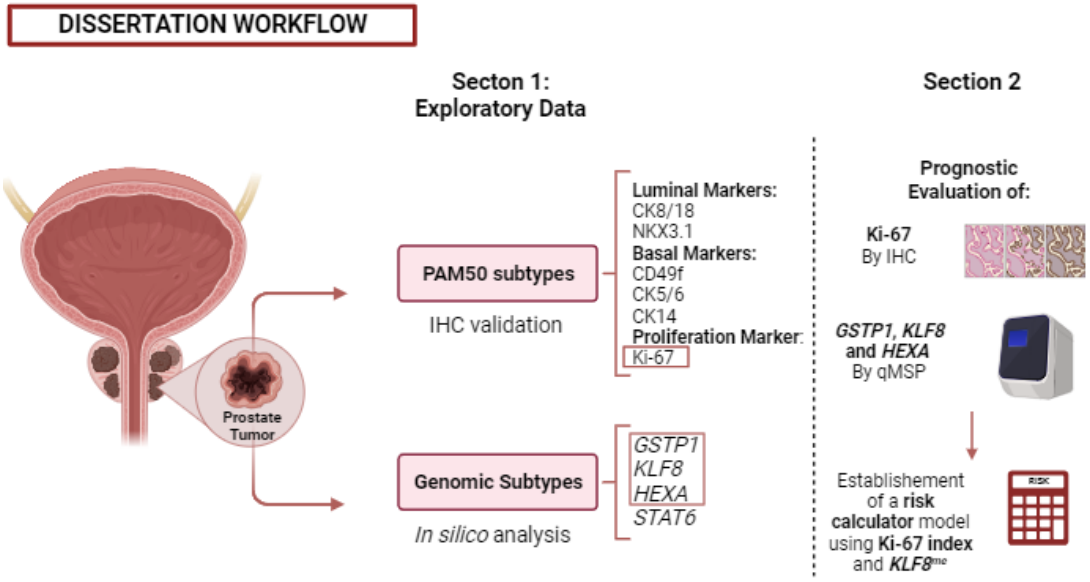


Figure 4- Schematic representation of the workflow of this dissertation. Created with Biorender.

MATERIAL AND METHODS

PATIENTS AND SAMPLES

One hundred and three PCa specimens (Cohort 1) were obtained from the archives of IPO-Porto Biobank. These formalin-fixed paraffin-embedded (FFPE) tissue samples were collected from patients with localized prostate adenocarcinoma submitted to radical prostatectomy at IPO-Porto between 2004 and 2008, without any previous treatment. Moreover, Cohort 1 was stratified to form the Cohort 1.1, which includes exclusively patients with Grade Group 1 and 2, according to ISUP classification system. The clinical data of the patients is summarized in Table 3. Additionally, a validation series of PCa biopsies (Supplementary Table 1), previously reported by other colleagues was used to validate the risk predictor calculators discovered in cohort 1 and cohort 1.1 (83). This study was approved by the Institutional Review Board of the Portuguese Oncology Institute of Porto (CES86/2022).

Table 3- Detailed clinico-pathological data of PCa patients

	Cohort 1	**Cohort 1.1 (Subgroup of GG1 and 2 patients)**
Clinico-Pathological Characteristics	**Mean ± standard deviation, total n=103**	**Mean ± standard deviation, total n=66**
Age (Years)	63.3 ± 6.03	63.4 ± 5.67
Clinico-Pathological Characteristics	**Patient Numbers n (%), total n=103**	**Patient Numbers n (%), total n=66**
Age (Categorized-years)		
≤55	12 (11.65%)	7 (10.61%)
55-65	53 (51.46%)	33 (50.00%)
>65	38 (36.89%)	26 (39.39%)
Serum PSA concentration (ng/mL)		
≤10	66 (64.08%)	48 (72.73%)
10-20	36 (34.95%)	18 (27.27%)
>20	1 (0.97%)	N/A
ISUP Grade Group		
1	20 (19.42%)	20 (30.30%)
2	46 (44.66%)	46 (69.70%)
3	9 (8.74%)	N/A
4	8 (7.77%)	N/A
5	20 (19.42%)	N/A
Clinical Stage		
pT2a	1 (0.97%)	1 (1.52%)
pT2b	2 (1.94%)	2 (3.03%)
pT2c	42 (40.78%)	39 (59.09%)
pT3a	43 (41.75%)	20 (30.30%)
pT3b	14 (13.59%)	4 (6.06%)
pT4	1 (0.97%)	N/A
Biochemical Recurrence	35 (33.98%)	18 (27.27%)
Clinical Recurrence	18 (17.48%)	3 (4.54%)
Deaths	46 (44.66%)	27 (40.91%)

Abbreviations: N/A- Non applicable

TISSUE MICROARRAY (TMA) CONSTRUCTION

Firstly, using a rubber cast of 60 cores (T-SueTM Microarray Mold 60 cores W/plunger 2mm Red, Simport Scientific, Canada), the TMA recipient blocks were made in a paraffin inclusion device (Modular Tissue Embedding Center EC 500). Here the liquid paraffin (at 61ºC) was placed in the mold and then it rested for about 30 minutes in the cold part of the device to solidify.

After the TMA recipient blocks were done, a puncher (T-SueTM punch needle plunger Simport Scientific, Canada) was used to remove 2mm diameter of tumor tissue (core) from donor blocks that were transferred to the recipient block. For each case, 3 cores were needed to ensure tumors' representativeness. A liver tissue core was used as a technical TMA guide. Furthermore, breast, colon, stomach, and tumor-adjacent prostate samples were used as controls.

Finally, after TMAs constructions, 3µm of thick sections were cut using a Rotary Microtome to further immunohistochemistry (IHC) evaluation.

IMMUNOHISTOCHEMISTRY (IHC)

Immunostaining was performed using the NovoLink TM Max Polymer Detection System (Leica Biosystems, Germany). The antibodies against CK8/18, NKX3.1, CK5/6, CK14, CD49f (section 1) and Ki-67 (section 2) were used at properly optimized conditions (Supplementary Table 2).

Briefly, after 15 minutes in the oven at 60°C, the slides were deparaffinized in xylol and hydrated in a decreasing series of alcohols. Then, the slides were washed in current water followed by distilled water for 3 minutes each. The antigen retrieval was realized using the retrieval buffer indicated for each antibody (Supplementary Table 2). Then, the endogenous peroxidases were neutralized using 3% hydrogen peroxide, for 10 minutes, in a humidified chamber, at room temperature. The slides were washed again with water, followed by two washes in TBST for 3 minutes. Subsequently, the horse serum was incubated at a dilution of 1:50 for 20 minutes, at room temperature. After this, the primary antibody was incubated with the appropriate time and dilution (Supplementary Table 2).

Afterward, the slides were incubated with post-primary and polymer solutions for 30 minutes each at room temperature. Then, 3,3'-Diaminobenzidine (DAB) was incubated for 10 minutes at a dilution of 1:10 at room temperature. Hematoxylin was used as a counterstained solution.

Finally, the slides were dehydrated with an increased series of alcohols followed by diaphanization with xylol. The assembly of the slides was made using an automatic slide assembler (Leica CV5030, Germany).

The slides were analyzed by an experienced pathologist according to a semi-quantitative method. The IHC analysis considers the intensity, the localization, and the percentage of positive staining cells. Specifically, for Ki-67, cases were categorized according the following score: 0: <5% of positive cells, 1: 5-10% of positive cells, and 2: >10% of positive cells, as previously described by us (83).

NUCLEIC ACID EXTRACTION AND QUANTIFICATION

The total DNA isolation was performed using the FFPE RNA/DNA Purification Plus Kit (Norgen Biotek, Thorold, Canada) for tissue samples, according to manufacturing instructions. The RNA was also extracted since the kit used is optimized for the extraction of both nucleic acids. However, only the isolated DNA was used in this study for methylation analysis. The DNAs were eluted in a final volume of 10µL of elution buffer and then the DNA was quantified using the High Sensitivity dsDNA Qubit Assay, according to manufacturing instructions. Finally, the samples were stored at -20°C.

BISULFITE MODIFICATION

After the quantification, 50 ng of DNA were used for bisulfite modification using the EZ DNA Methylation-Gold Kit (REF: D5006), with a final concentration of 2ng/µL, according to manufacturing instructions.

IN SILICO ANALYSIS

To unveil the most relevant CpG islands within the promotor region of *GSTP1*, *KLF8*, *HEXA*, and *STAT6* genes, The Cancer Genome Atlas (TCGA) Human methylation 450k array data were retrieved from Shiny Methylation Analysis Resource Tool (SMART) App website **http://www.bioinfo-zs.com/smartapp/** (84). SMART App website was used to perform a pair-wise correlation analysis of differential methylation using the following criteria: Dataset- PRAD-TCGA dataset; Methylation value- Beta-value; Gene expression- Log2-scaled (TMP+1) values; Correlation coefficient- Spearman; and Aggregation Method- mean. The CpG sites within the gene promoter that most likely influence gene expression are the ones inversely correlated with gene expression levels. Therefore, the most significant and highly anti-correlated CpG sites were selected for further quantitative methylation specific PCR (qMSP) analysis. Nonetheless, no data was available for *STAT6* gene promoter and no further analysis was conducted for this gene.

PRIMER AND PROBE DESIGN

The primers and probes were specifically designed based on the precise location of the selected CpGs, except for *GSTP1* gene, since the primers designed based on this criterion were not validated in our cohort. So, the primers used for *GSTP1* methylation assessment were the same as previously described in our research group (85). The final sequences of primers and probes are listed in Table 4.

Primers and probes were designed 1500bp upstream of promoter region using the template of the genomic DNA sequence, available at UCSC Genome Browser on Human Dec.2013 (GRCh38/hg38) Assembly. The bisulfite-modified DNA sequence was obtained through Methyl Primer Express v1.0. For the primers design the following criteria were considered: the melting temperature should be between 58 and 60°C, the optimal primer length between 18 and 20 bases, and the product size from 80 to 150bp. The probes should not hybridize or overlap the primers and the melting temperature should be about 10°C higher than that of the primers. The probe length should be between 13 to 30 bases and in the 5' end it should be avoided a G residue. Additionally, both primers and probe sequences should have an ideal CpG content between 30 and 80%.

For quality control, primers and probes sequences were analyzed using the Primer Express 3.0-Primer Probe Test Tool, as well as and the Beacon Designer, Premier Biosoft (Palo Alto, CA, USA) and Bisearch Primer Design and Search Tool to assure PCR product specificity.

Table 4- Primers and Probes sequences with respective fluorochrome and quencher

Gene	Sequence	
β-Actin	**Primers**	F – 5' TGGTGATGGAGGAGGTTTAGTAAGT 3'
		R – 5' AACCAATAAAACCTACTCCTCCCTTAA 3'
	Probe	5' Cy5 – ACCACCACCCAACACACAATAACAAACACA – BHQ 3'
GSTP1	**Primers**	F – 5' AGTTGCGCGGCGATTTC 3'
		R – 5' GCCCCAATACTAAATCACGACG 3'
	Probe	5' FAM – CGGTCGACGTTCGGGGTGTAGCG – TAMRA 3'
KLF8	**Primers**	F – 5' GGCGCGAGGTATTTCGTG 3'
		R – 5' CCCCAAAACAAAAACTTCAACG 3'
	Probe	5' FAM – TTCGTTTTAGAGGATTCGTACGAGTTGTTG – BHQ1 3'
HEXA	**Primers**	F – 5' GGGTCGTAGGCGTAGAGTCG 3'
		R – 5' TCCTACGAACGCTACCGCC 3'
	Probe	5' HEX – CGTCGATAAGTTACGGGGGCGTCG 3'

QUANTITATIVE METHYLATION-SPECIFIC REAL-TIME PCR (qMSP)

The qMSP reactions were run in 384-well plates in QuantStudio 12K Flex Real-Time PCR System (Thermo Fisher, Forter, CA, USA). In brief, 1.5µL of bisulfite-modified DNA,

5µL of MasterMix Xpert Fast Probe (GRISP) with ROX, an optimized volume of primers and probe at 10µM and sterile bi-distilled water in a total volume of 10µL were added to each well (Table 5). A multiplex qMSP reaction was run, consisting of 'panel #1' (*KLF8* and *β-Actin*) and 'panel #2' (*HEXA*, *GSTP1* and *β-Actin*). *β-Actin* was used as a housekeeping gene in both panels to ensure the reproducibility of the PCR conditions. Primer/probe annealing temperature was optimized for all genes at 60°C.

Table 5- Conditions of qMSP for each gene

Gene	Primers, volume (µL), F+R	Probe, volume (µL)	Annealing Temperature (°C)
β-Actin	0.5	0.05	60°C
GSTP1	0.3	0.05	60°C
KLF8	0.3	0.05	60°C
HEXA	0.5	0.05	60°C

No template controls (NTC) were used in every plate assuring the absence of contamination. Negative controls (Human HCT116 DKO Non-Methylated DNA, D5014-1; Zymo) were used in primers optimization to ensure the specificity for the detection of the methylated DNA template. Serial dilutions (five, at 1:5) of positive control (Human HCT116 DKO Methylated DNA, D5014-2; Zymo) were included in each plate to compute a standard curve and assure run efficiency to a final interpolation of the PCR results. The run efficiency was considered valid if the values were between 90 and 110%. All samples were run in triplicate. Results were plotted as relative methylation levels (ratio between gene and housekeeping mean quantities) multiplied by 1000, for easier tabulation.

STATISTICAL ANALYSIS

Non-parametric tests were used to determine de statistical power of the differences among comparable groups, through GraphPad Prism version 8.0 (GraphPad Software, La Jolla California USA). Specifically, the Mann-Whitney test and the Kruskal-Wallis test were used to compare between two groups and among multiple groups, respectively. Tumor samples were classified into low or high methylation levels based on the frequency distribution of the data, using the 50th percentile as the cut-off value. Moreover, the Log-rank test was applied to assess differences in overall survival and BCR free survival, for each marker.

For risk calculator computation, nomograms were constructed based on multivariable logistic regression analysis, using Rstudio from R software (version 4.3.1). Afterward, receiver operator characteristic (ROC) curves were used to evaluate the performance of our models, based on the calculation of the area under the curve (AUC). The rms, caTools, epicalc, and pROC packages were installed for these analyses. Youden's J index (value combining highest sensitivity and specificity) was used for categorization of results and subsequent assessment of the performance of our model in predicting BCR risk, dividing the cohort in two groups (high-risk and low-risk) according to the cut-off established for the Linear predictor value at the nomogram.

For each analysis, the *p-values* lower than 0.05 were considered statistically significant. Specifically, *$p<0.05$; **$p<0.01$; ***$p<0.001$; ****$p<0.0001$, ns: non-significant.

SECTION 1-
Exploratory Data

1.1- The practicability of Luminal and Basal molecular subtyping in PCa

INTRODUCTION

The glandular prostate is the most relevant part of the prostate's carcinogenic process. The prostatic gland is composed mostly for two main cell types, luminal and basal cells, in pseudostratified epithelium, in which some neuroendocrine cells can be seen (86). These prostate cell subtypes can be distinguished based on the expression of different cell markers (87-89). It is also possible to find some rare cells that co-express luminal and basal cell markers, the intermediate cell population, also called transit-amplifying cells (90). Although prostate tumors are typically recognized by the proliferation of luminal cells and the absence of basal cells, it should be noted that PCa has been associated with not only the involvement of luminal cells but also basal cells and transit-amplifying cells as possible cells of origin. Therefore, PCa patients might have different treatment outcomes, based on those tumor cell heterogeneity (91-93).

The distinction between luminal and basal-like cells extends beyond PCa model. The luminal and basal cell-like characteristics are important for defining distinct molecular subtypes also in other cancer types, including bladder cancer and breast cancer, the later much more stablished in daily care (47, 94). Nonetheless, the clinical practicability and impact of this diagnostic classification in PCa is not yet known. Previously, in another study, the PAM50 classifier was used to subtype a large retrospective and an additional prospective cohort of PCa patients, by assessing the prognostic value of several defined markers (47). Herein, PCa tumors were segregated into three distinct subtypes – luminal A, luminal B and basal - based on luminal-like and basal-like lineage characteristic markers. Notably, the luminal B-like subtype was found associated with a worse outcome, although a better response to ADT was attained by the patients harboring these tumors (47).

This classification could be of valuable interest to potentially redirect ADT treatment to those patients, understanding who will really benefit from it, based on tumor aggressiveness. However, the clinical applicability and practicability of this approach remains controversial since those results were obtained using microarray transcriptomic analysis.

Thus, we aim to study whether these observations were reflected in the same way at the protein level, unveiling the clinical practicability of subtyping PCa according to tumor cell type.

RESULTS

Immunoexpression of NKX3.1 and CK8/18 (luminal markers), CK5/6, CK14, and CD49f (basal markers), and Ki-67, a proliferative marker

To evaluate if the previous findings are reflected at protein level, immunohistochemistry was performed in a series of 103 localized PCa samples to attempt to identify luminal and basal-like markers enriched in specific cell subpopulations. The clinico-pathological data of these patients are described in Table 3. In our study, NKX3.1 and CK8/18 were highly expressed across all the cases. Conversely, CK14 was almost absent in our PCa tissue series. Hence, these markers may not be helpful for distinguishing or categorizing PCa tumors. Nonetheless, CD49f and CK5/6, which were previously recognized as characteristic basal cell markers, displayed a rather variable protein expression in our samples' set. Moreover, Ki-67, a proliferation marker often associated with tumor aggressiveness, previously used to distinguish between luminal A and luminal B-like tumors, also displayed a wide range of expression across all the studied samples (Figure 5).

Based on this, the cases showing concurrent CK5/6 and CD49f expression were classified as basal-like, while the remaining cases were categorized as luminal-like and associations with clinico-pathological data were investigated, as outlined in the following subsections.

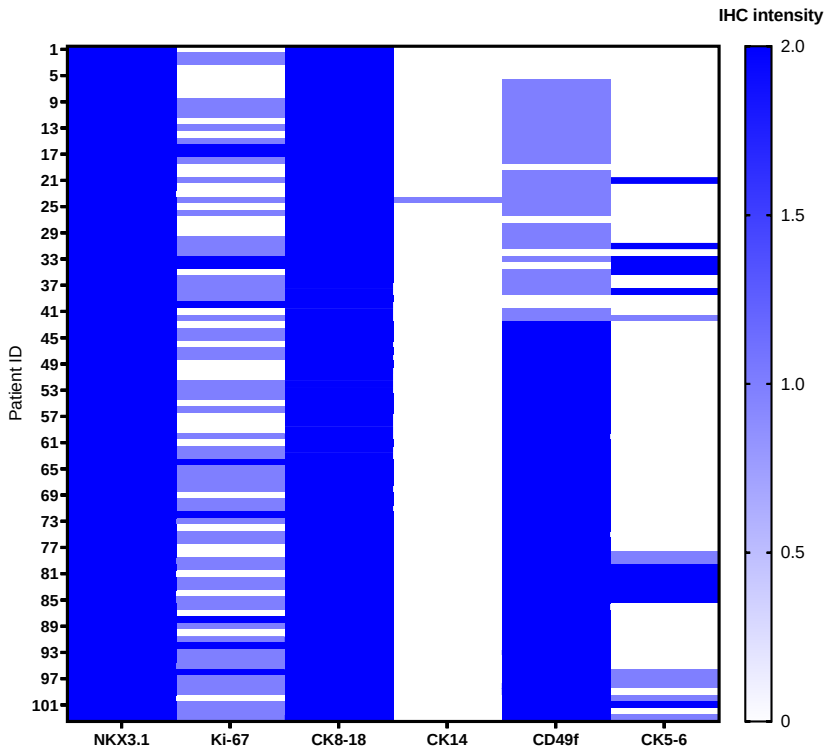


Figure 5- Array representative of the luminal (NKX3.1 and CK8/18) and basal (CK14, CK5/6 and CD49f) markers, as well as Ki-67 immunoexpression across the cohort.

Association of CK5/6 and CD49f (basal markers) immunoexpression with clinico-pathological variables

Overall, stage 3 tumors displayed significantly higher percentage of positive staining for CK5/6 immunoexpression than stage 2 tumors (Figure 6A). Conversely, no differences were found for CD49f immunoexpression between the two groups (Figure 6B). Moreover, high-grade tumors presented higher percentage of positive CK5/6 staining comparing with low-grade lesions (Figure 6C). Meanwhile, there was no apparent pattern observed regarding the association between CD49f immunoexpression and the GG. Nonetheless, statistically significant differences were found in certain comparisons between groups (GG1 vs GG2, GG1 vs GG5, GG2 vs GG5, and GG4 vs GG5) (Figure 6D). Remarkably, neither CK5/6 or CD49f expression impacted patients' prognosis (overall survival and BCR-free survival) (Supplementary Figure 1).

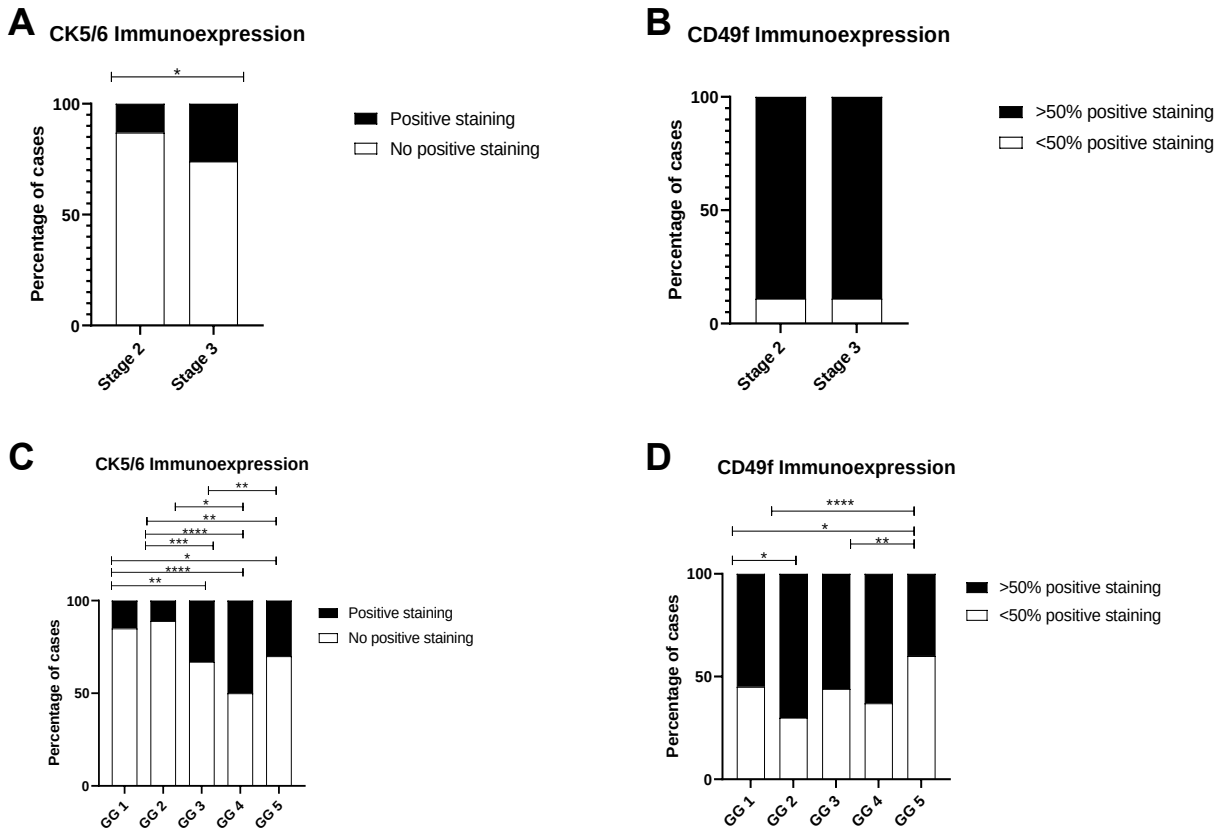


Figure 6- Association of CK5/6 and CD49f immunoexpression with clinico-pathological features. Comparisons between stage 2 and stage 3 tumors samples were made based on the immunoexpression of the CK5/6 **(A)** and the CD49f **(B)**. **C-** CK5/6 and **D-** CD49f immunoexpression amoung Gleason grade groups. Median ranks between two groups were compared using the Mann-Whitney test and the comparisons between grade groups were performed using Kruskal-Wallis test. *, **, ***, **** represent a p-value lower than 0.05, 0.01, 0.001, and 0.0001, respectively.

CK5/6 and CD49f immunoexpression defined according to the luminal-like and basal-like subclassification and its impact for overall-survival rates

Overall, considering the selected markers immunoexpression, only seven cases were classified as basal-like, and no differences were obtained regarding overall and BCR-free survivals, between luminal-like and basal-like tumors (Figure 7).

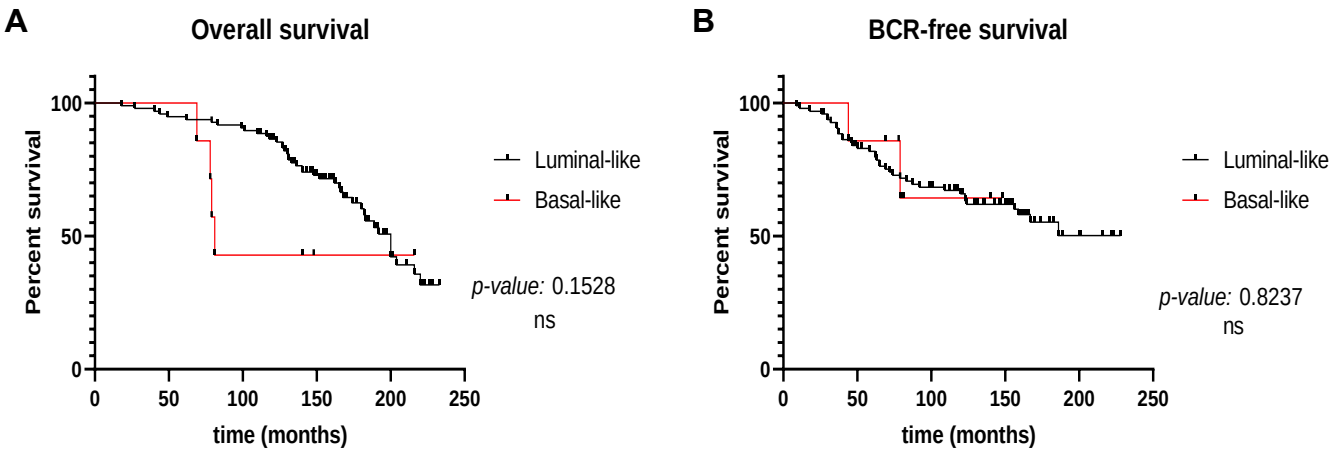


Figure 7- Kaplan-Meier curves for the overall survival **(A)** and the biochemical-recurrence free survival **(B)**, based on the subclassification into luminal-like (n=96) and basal-like (n=7). The Log-Rank test was used to evaluate the differences in overall survival and BCR-free survival between groups, where ns represent *p-values* with no statistical significance.

High Ki-67 immunoexpression predicts worse patient prognosis

Notably, in our study, Ki-67 index emerged as a valuable biomarker for PCa recurrence prediction. Consequently, this protein-based biomarker was selected for further in-depth analysis in the second section of this dissertation.

1.2- Specific gene promoter hypermethylation associated with well-known PCa genomic signatures

STATE OF ART

As previously mentioned, neoplastic cells experience a switch in methylation pattern, commonly with aberrant promoter hypermethylation, leading to gene silencing (69). Curiously, different hypermethylated genes were associated with different PCa genomic signatures (26). Specifically, in a previous study, *STAT6* was epigenetically silenced in ERG-positive tumors, while *HEXA* were preferentially hypermethylated in *SPOP*-mutant tumors. Moreover, *GSTP1* and *KLF8* promoter genes hypermethylation is considered an early event in PCa carcinogenesis, being present in almost all PCa tumors (26). Henceforth, we went to confirm whether these genes were altered in a series of PCa samples from TCGA. Remarkably, all mentioned genes were hypermethylated in PCa tumor samples compared to normal adjacent tissues (Supplementary Figure 2). This may be suggestive of the relevance of specific gene promoter hypermethylation in the process of PCa development and/or progression.

Thus, next we ascertain the most relevant CpG islands on the promoter of *STAT6*, *HEXA*, *GSTP1* and *KLF8* genes.

IN SILICO ANALYSIS FOR THE MOST RELEVANT CpG SITES

We started by using TCGA Human methylation 450k array *in silico* data retrieved from SMART App website http://www.bioinfo-zs.com/smartapp/ to address the most relevant specific CpG islands within the promoter region of the selected genes. By using this algorithm, we pinpointed specific *GSTP1*, *KLF8*, and *HEXA* CpGs aggregation methylation values with significant negative correlation with gene expression, emphasizing that the main mechanism of regulation of these genes is promoter methylation (Figure 8). No data was available for *STAT6* promoter in the aforementioned publicly available database, and then, we did not include this gene for further analysis. The number of significant hypermethylated CpG islands of *GSTP1*, *KLF8*, and *HEXA* are listed in Table 6. Overall, these CpG islands were further used to design the qMSP primers (as previous mentioned in the materials and methods section).

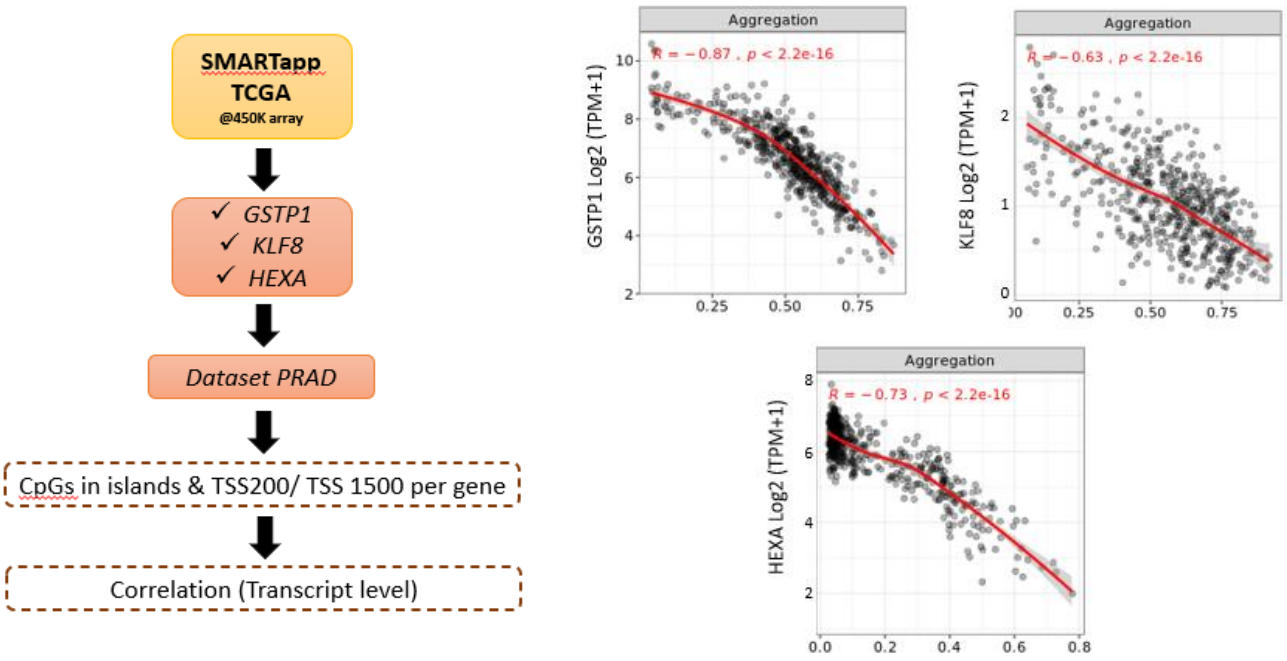


Figure 8- *In silico* analysis for the most relevant methylated CpGs in PCa. A- Schematic representation of the *in silico* analysis performed for differential methylation patterns in PCa. Most relevant specific CpG islands within the promoter region of *GSTP1*, *KLF8*, and *HEXA* genes were assessed using TCGA Human methylation 450k array data retrieved from SMART App website http://www.bioinfo-zs.com/smartapp/. **B-** Graphic representation of aggregation analysis representative of the inverse correlation between gene expression and methylation levels for each gene. **Abbreviations:** PCa- Prostate cancer; TCGA- The Cancer Genome Atlas.

Table 6- Relative Methylation levels of *GSTP1*, *KLF8,* and *HEXA* promoter regions

Gene	Number of Hypermethylated CpGs
GSTP1	6
KLF8	3
HEXA	8

RATIONAL

Currently, the primary therapeutic-decision is based on the Clinical Risk Assessment, that categorize PCa patients in risk groups based on similar clinico-pathological characteristics. Nonetheless, new strategies have been investigated to increase the accuracy of the risk stratification and, consequently, improve the therapeutic decision making. Indeed, a previous study, using the PAM50 classifier, was able to segregate prostate tumors into luminal A, luminal B, and basal subtypes, entailing significant differences in the prognostic outcomes (47). However, the results obtained were based on transcriptomic data, which not always reflect the levels of protein expression. Thus, in our study, we sought to test this molecular classification using immunohistochemistry and assess its clinical applicability in PCa diagnosis. Additionally, due to the clinical relevance of the genomic subtypes we also aimed to perform an *in silico* analysis to unveil the methylation landscape of such subtypes and to try to find biomarkers with significant prognostic value for this tumor model.

Since all the analyzed tumors expressed luminal markers, it was not informative to use them for the classification system. Contrarily, CK5/6 and CD49f expression allowed to define the basal-like origin subtype. However, in our series of localized prostate tumors, no differences in the patients' prognosis were found between luminal-like and basal-like PCa patients, according to this sub-categorization. Therefore, the molecular classification seemed to fail for the clinical applicability or prognostic value, so far, by immunohistochemistry, an approach closer to the clinical practice, faster and more cost-effective than the transcriptomic analysis.

Nonetheless, in line with transcriptomic data, Ki-67 index, used to distinguish luminal A-like from luminal B-like tumors, was associated with patients' prognosis (Detailed results in section 2). Thus, we hypothesize whether Ki-67 index can be used as an independent PCa prognostic biomarker and be helpful for therapeutic decision-making. Along with Ki-67, we resolve to access the prognostic value of other potential biomarker candidates. So, based in our *in silico* analysis, we went to explore whether the methylation status of *GSTP1*, *KLF8* and *HEX*A has impact on PCa patient's prognosis (Detailed results in section 2).

SECTION 2

Manuscript in preparation

Ki-67 immunoexpression and *KLF8^{me}* levels as independent prognostic biomarkers for localized prostate cancer

INTRODUCTION

Prostate cancer (PCa) remains a highly prevalent disease in men worldwide (1, 21). Characteristically, this disease develops gradually and is frequently diagnosed in older individuals (1, 21). The high incidence of this disease is mostly attributed to the widespread PSA screening, which results in nearly 90% of PCa diagnosed as organ-confined disease (26). When initially diagnosed, defining the optimal primary treatment for PCa is articulated with the risk of biochemical recurrence (BCR). This risk assessment considers factors like PSA serum levels, clinical stage, and tumor grade, with the International Society of Urological Pathology (ISUP) grade system playing a central role (17, 50, 51). For high-risk patients (PSA>20ng/mL and ISUP grade greater than 4 or cT2b), therapeutic procedures commonly involved neoadjuvant or concurrent androgen deprivation therapy (ADT) plus radiotherapy (50, 54). Nonetheless, the approach to low- and intermediate-risk PCa patients remains challenging. Currently, it is difficult to predict which patients may benefit most from an active intervention-based approach or, on the other hand, may be better placed under active surveillance. Indeed, a subset of patients undergoing active surveillance eventually progress, with 11 to 32% of men ultimately receiving active treatment (50, 95). However, it remains difficult to accurately predict which patients will have poor outcome and require treatment. The lack of reliable biomarkers of PCa aggressiveness with prognostic value is a major drawback. Nevertheless, efforts to address this issue have been accomplished with the emergency of new strategies based on genomic tests, such as the Decipher® score, Prolaris® cell cycle progression test, and Oncotype DX® Prostate Score. However, the results are not always clinically informative (12, 50, 56-60).

These tests, which rely in whole transcriptomic data, offer valuable prognostic insights, especially by predicting the risk of metastasis, the probability of tumor regrowth and the risk of death (12, 50, 56-60). Despite the advantages of these newly “cancer prognostic tests”, they are very expensive, with a price ranging from 3.900 to 5.150 dollars per patient, being difficult to implement in the clinic practice. Consequently, there is an urgent need to explore alternative strategies that can leverage more cost-effective and convenient tools, such as immunohistochemistry (IHC) or PCR, to make this risk assessment simpler in clinical practice (56). For instance, developing a more precise and reliable method may be pivotal to define the most appropriate therapeutic choice, particularly when dealing with a subset of patients with low- and intermediate-risk tumors, since they can be candidates for active surveillance.

As such, in the present study, our main goal was to construct a risk stratification calculator. This tool is based on relevant clinico-pathological characteristics and new biomarkers, employing techniques that can be easily integrated into routine clinical practice. To accomplish this aim, we evaluated Ki-67 immunoexpression, known to be associated with breast, lung and colorectal cancers aggressiveness and increased recurrence (61, 64). As for PCa, Ki-67 immunoexpression has been tested as an independent disease-specific survival biomarker. Specifically, Ki-67 index prognostic value was particularly informative in grade group 1 and 2 tumors, enlightening its potential to discriminate indolent tumors from more aggressive ones (65).

Moreover, we employed quantitative methylation specific PCR (qMSP) to assess the methylation levels of *GSTP1* and *KLF8* genes. Given their hypermethylation as an early event in prostate carcinogenesis, these genes hold great potential as biomarkers for PCa (26, 96). Although *GSTP1* has strong value for PCa diagnosis, as previously reported by us, $KLF8^{me}$ had not yet been explored as a biomarker in this model. Notably, KLF8 expression was associated with invasion and progression in breast, bladder, and colorectal cancers (97-99).

Therefore, we sought to determine whether relative methylation levels of these selected candidate genes were significantly associated with a higher risk of recurrence or worst overall PCa survival. Remarkably, our findings underscore the significance of both *KLF8* methylation ($KLF8^{me}$) and Ki-67 immunoexpression as valuable prognostic indicators, making them strong candidates for assessing the risk of PCa progression and/or recurrence, offering a straightforward and useful approach for clinical implementation, particularly in the context of low- and intermediate-risk PCa.

RESULTS

Association of Ki-67 immunoexpression and *GSTP1* and *KLF8* methylationlevels with clinico-pathological variables

Considering all the cases, only residual methylation levels were found for *HEXA.* Further analyses were performed only for *GSTP1* and *KLF8* (Supplementary Figure 3). The clinico-pathological data of the 103 PCa patients used in this study is described in Table 3. PCa patients diagnosed with stage 3 disease displayed significantly higher proportion of cases with high Ki-67 immunoexpression (>10%), (*p-value*<0.01, Figure 9A and Supplementary Figure 4). Significant differences among GG groups were found for Ki-67 immunoexpression (Figure 9B). Specifically, high-grade tumors (GG 3, 4, and 5) displayed higher Ki-67 immuno-score compared with low-grade tumors (GG 1 and 2). Moreover, *GSTP1* and *KLF8* methylation levels were significantly higher in high-stage tumors (stage 3) than in low-stage tumors (stage 2) (*p-value*=0.0089 and *p-value*=0.0166 respectively) (Figure 9C and D). However, no associations were found among GG groups (Supplementary Figure 5).

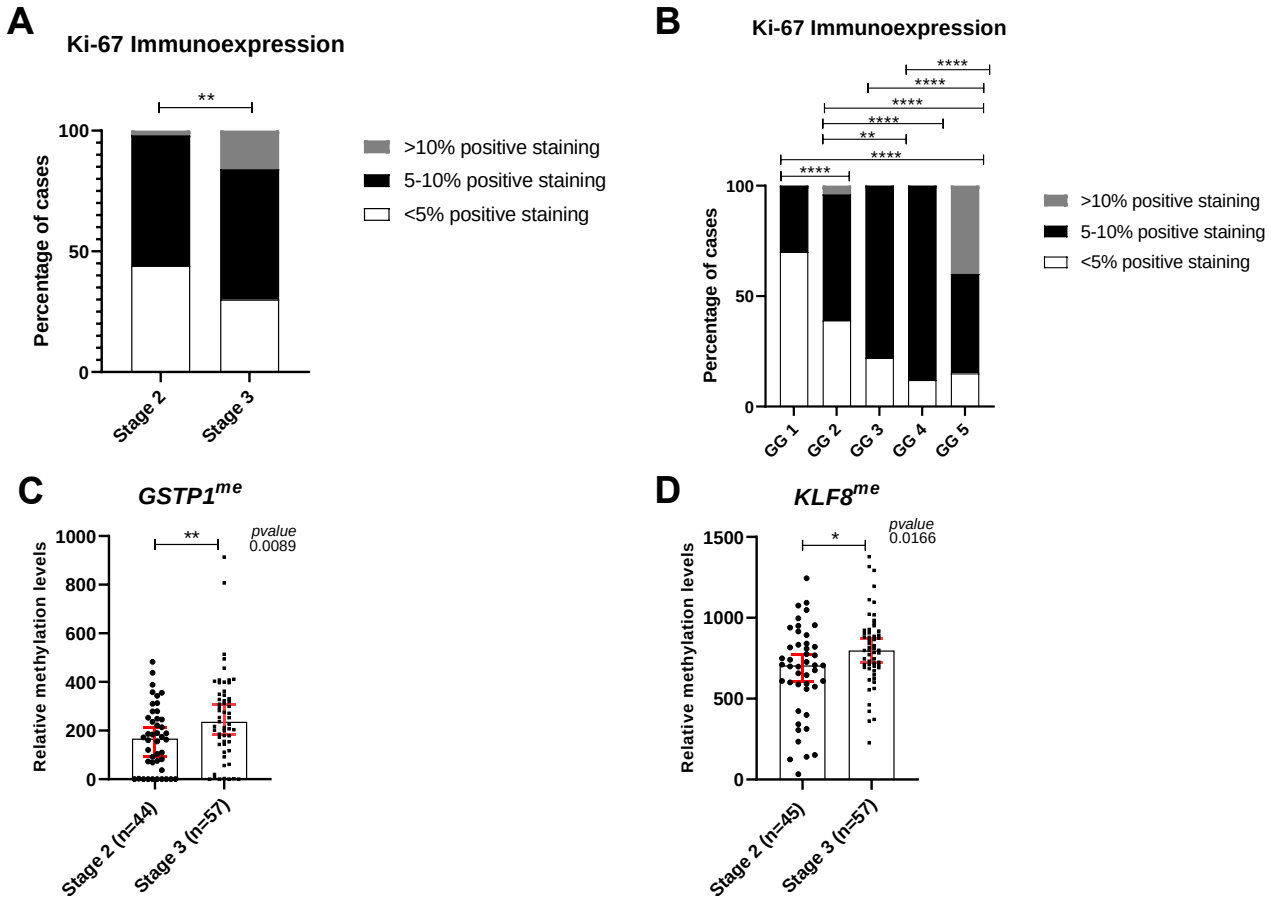


Figure 9- Ki-67 immunoexpression and *GSTP1* and *KLF8* relative methylation levels according with clinico-pathological features. Ki-67 immunoexpression in stage 2 and stage 3 tumors **(A)** and Gleason grade groups **(B)**. **C-** *GSTP1*^{me} and **D-** *KLF8*^{me} levels in stage 2 and stage 3 prostate tumor samples. Median ranks between two groups were compared using the Mann-Whitney test and the comparisons between grade groups were performed using the Kruskal-Wallis test. *, **, and **** represent a *p-value* lower than 0.05, 0.01, and 0.0001, respectively.

Ki-67 positive staining and KLF8 methylation levels independently predict PCa patients' prognosis

GSTP1 and *KLF8* methylation levels did not significantly associate with overall survival or the risk of biochemical recurrence considering all the cases (n=103, identified as cohort 1) (Supplementary Figure 6). Nonetheless, when we limited the analysis to lower-grade PCa patients (GG 1 and 2 in cohort 1.1), lower *KLF8* methylation levels were associated with shorter overall survival (*p-value*=0.0263) (Figure 10A). Additionally, patients displaying moderate (5-10% of positive staining) and strong (>10% of positive staining) Ki-67 immunoexpression faced significantly reduced BCR-free survival, in both cohort 1 and 1.1 (*p-value*=0.0002 and *p-value*=0.0121, respectively) (Figure 10B and C), although no association was found concerning overall survival (Supplementary Figure 7). Nonetheless, in lower-grade tumors (cohort 1.1), Ki-67 immunoexpression and *KLF8* methylation levels showed prognostic value, although no associations were observed when they were combined. (Figure 10D).

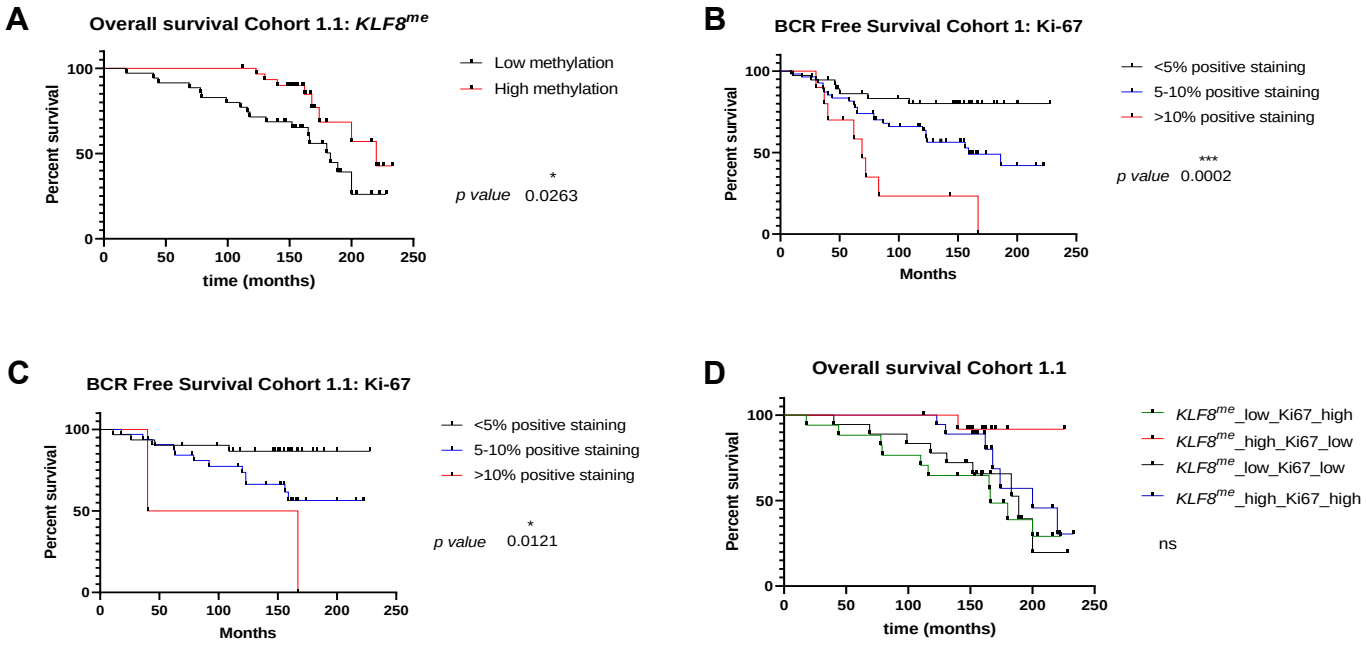


Figure 10- A- Kaplan-Meier curve for the overall survival based on the methylation levels of *KLF8*, in cohort 1.1. Kaplan-Meier curves for the biochemical-recurrence free survival based on the expression of Ki-67, in cohort 1 **(B)** and in cohort 1.1 **(C)**. **D-** Kaplan-Meier for the overall survival for the combination of both markers, in cohort 1.1. The Log-Rank test was used to evaluate the differences in overall survival and BCR-free survival, where * and *** represent p-values lower than 0.05 and 0.001, respectively.

Nomogram models accurately predict patients' outcomes

Several pipeline models were developed and tested to find the one that better predict patients' prognosis. All models developed for Cohort 1 included the clinico-pathological variables: age, clinical stage, GG and PSA levels due to their recognized clinical relevance. For Cohort 1.1, GG variable was omitted, given that this group setting was already focused only for low-grade tumors. Given that Ki-67 high immunoexpression was significantly associated with the risk of BCR (*p-value*=0.0057, 95% CI: 0.4933-2.8908), this marker was selectively included in models aiming at predicting the risk of BCR. Conversely, *KLF8* methylation levels were only considered in models for predicting death (*p-value*=0.0064, 95% CI: 0.4932-3.0238). Consequently, we formulated nomogram models for overall survival and BCR-free survival, representative of the respective equations – risk of death = -4.0594+1.0339*__AGEcat__+1.7585*__KLF8__+0.8255*__Tstage__-0.5032*__PSA__ and risk of BCR = -4.5509-0.1447*__AGEcat__+1.6921*__Ki67__+0.5982*__Tstage__+0.9164*__PSA__, resulting in the generation of risk calculator models for both risk of death and risk of BCR, which we designated **ProstARK** calculator (Figure 11A and B). For all models, described in detail in Table 7 and Supplementary Table 3, the area under the curve was greater than 0.75 (Figure 11C and Supplementary Figure 8). Indeed, considering only low-grade tumors, the area under the curve for Model 1 and Model 2 defined in Table 7 were 0.764 (95% CI: 0.644-0.885) and 0.752 (95% CI: 0.615-0.888), respectively.

Table 7- Summary of the logistic models representative of the Risk of Death and the Risk of BCR calculators, in cohort 1.1

					95% Confidence Interval (CI)	
	Coefficient	**S.E.**	**Wald Z**	***p-value***	**Lower Bound**	**Upper Bound**
Model 1: Risk of Death Cohort 1.1						
Intercept	-4.0594	1.9136	-2.12	0.0339	-7.8101	-0.3088
AGEcat	1.0339	0.4913	2.10	0.0353	0.0710	1.9968
KLF8	1.7585	0.6456	2.72	0.0064	0.4932	3.0238
Tstage	0.8255	0.6406	1.29	0.1976	-0.4302	2.0812
PSA	-0.5032	0.6781	-0.74	0.4581	-1.8322	0.8259
Model 2: Risk of BCR Cohort 1.1						
Intercept	-4.5509	1.9821	-2.30	0.0217	-8.4357	-0.6661
AGEcat	-0.1447	0.4912	-0.29	0.7683	-1.1075	0.8180
Ki-67	1.6921	0.6116	2.77	0.0057	0.4933	2.8908
Tstage	0.5982	0.6454	0.93	0.3540	-0.6666	1.8631
PSA	0.9164	0.6630	1.38	0.1669	-0.3831	2.2160

Abbreviations: S.E.- Standard Error; CI- Confidence Interval; BCR- Biochemical-recurrence.

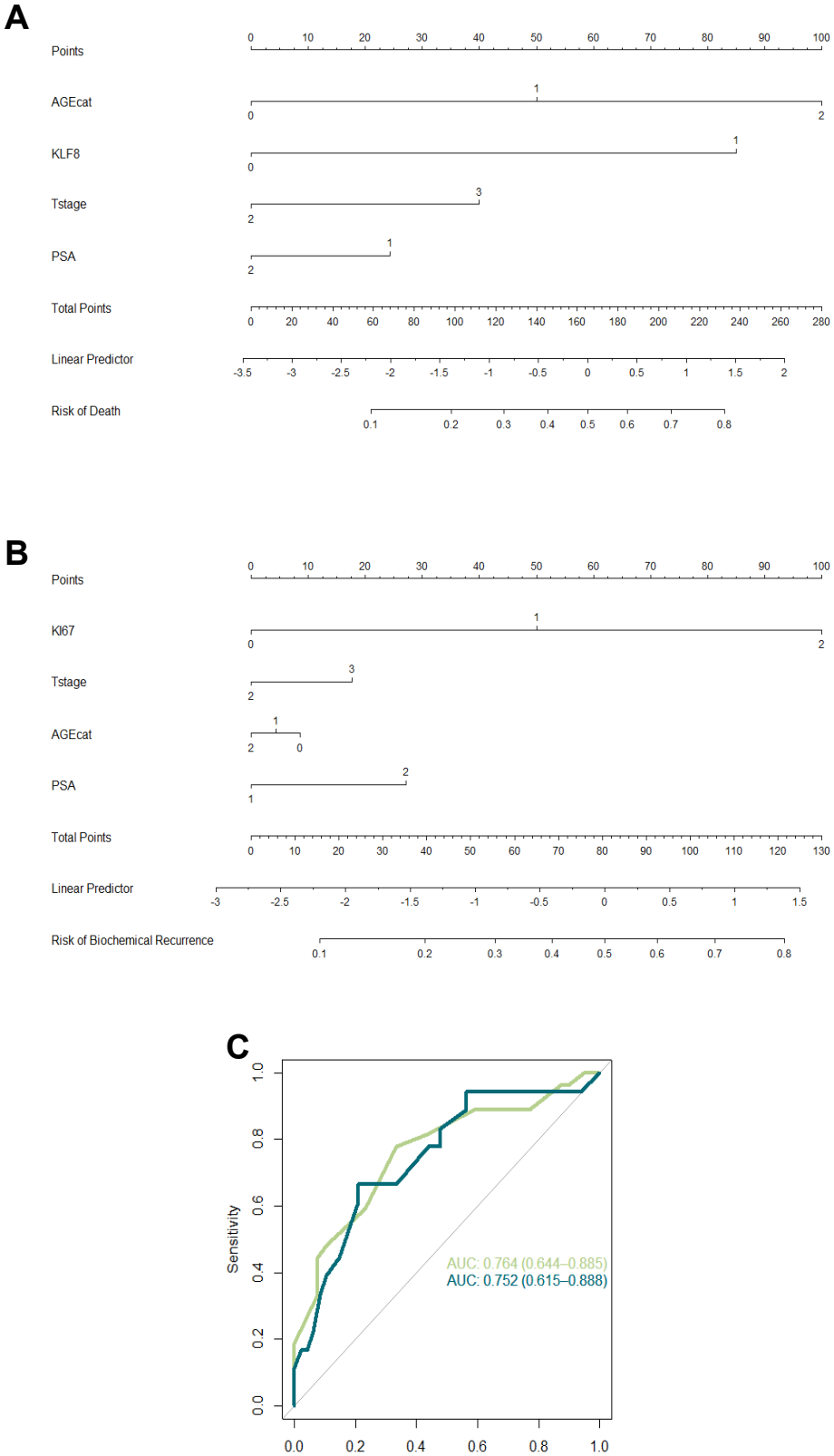


Figure 11- Nomogram representative of the Risk of Death **(A)** and the Risk of Biochemical-recurrence **(B)** calculators, for cohort 1.1, with the relevant clinico-pathological variables used in the risk stratification: age at diagnosis (0: ≤55, 1: 55-65, 2: >65 years), clinical stage and PSA serum levels (1: <10ng/mL, 2: >10ng/mL). **C-** Roc curves to evaluate the performance of the nomogram models. ***KLF8^{me}*** **categorization-** 0: high methylation levels, 1: low methylation levels. **Ki-67 immunoexpression categorization-** 0: <5%, 1: 5-10%, 2: >10% of positive staining. **Green and Blue Curves:** ROC curve for the Risk of Death calculator and for the Risk of Biochemical-recurrence calculator, respectively.

ProstARK predicts the risk of biochemical recurrence in an independent series of prostate biopsies

Ultimately, the BCR prediction models were assessed in an independent series of PCa biopsies. When considering the entire series, the area under the curve obtained was 0.813 (95% CI: 0.745-0.882) (Figure 12A), while if only considered the lower-grade tumors (GG 1 and 2), the area under the curve value achieved 0.723 (95% CI: 0.571-0.875) (Figure 12B), meaning that there is a 72% chance of our model correctly predicting the patients' outcome.

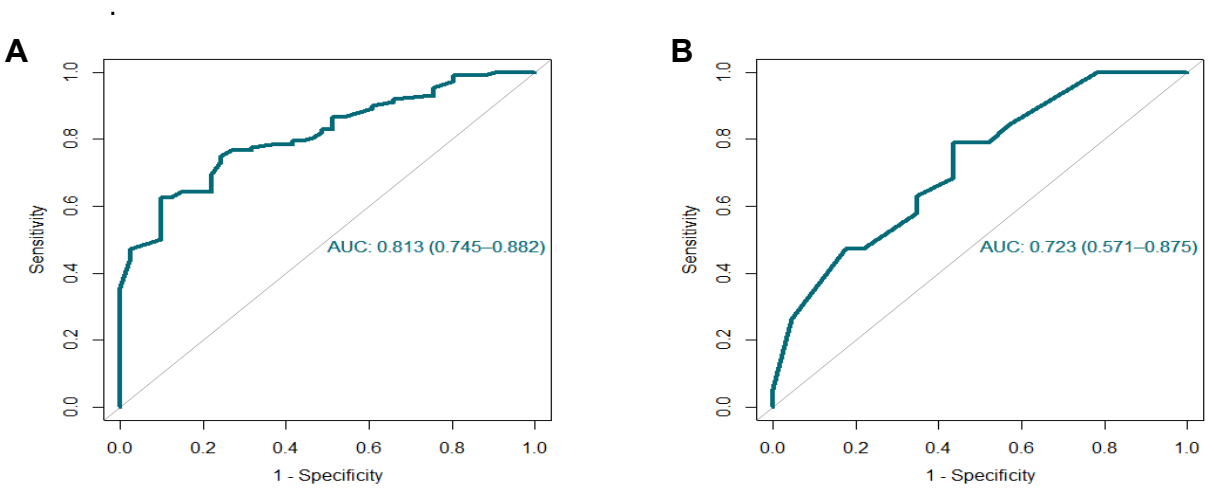


Figure 12- ROC curve to evaluate the performance of the Risk of Biochemical-recurrence calculator in an independent cohort of localized prostate tumor biopsies.

Clinical applicability of death- and BCR-risk assessment calculators for PCa prognostication

The integration of ProstARK in the clinics may be key for accurate risk assessment and improvement of patient management. Specifically, concerning the risk of death calculator, although based in overall survival data, a 0.41 cut-off of Linear predictor constructed on the Optimal Youden Index, that is equivalent to a 0.60 risk of death, stratified the low GG PCa patients into low- and high-risk of death (Figure 13A). In relation to this last calculator, ProstARK, a cut-off of 0.31 for the Linear predictor, translated into recurrence/progression rate of 0.58, entailing that patients diagnosed with low GG PCa but displaying a risk of progression exceeding 0.58 are classified as high-risk patients (Figure 13B). These high-risk patients might benefit from an active treatment instead of active surveillance. Moreover, the other low GG PCa patients that do not meet this criterion, will have their unnecessary interventions as they have favourable prognosis.

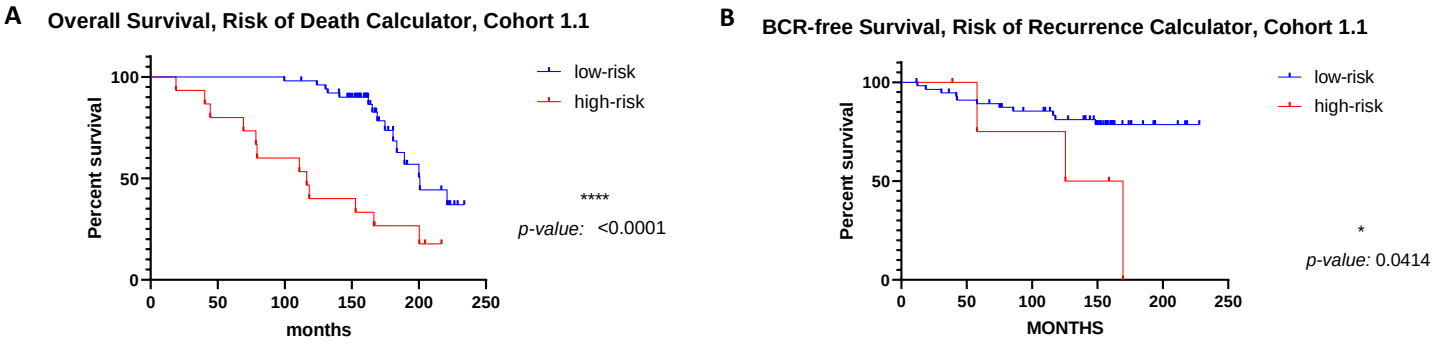


Figure 13 - Kaplan-Meier curve for the overall survival based on the Risk of death calculator, in cohort 1.1. Kaplan-Meier curves for the biochemical-recurrence free survival based on the ProstARK calculator, in cohort 1.1.

Hence, this calculator enables a more precise risk stratification among low-grade PCa patients, optimizing clinical decisions on the most appropriate therapeutic strategy. Furthermore, assessment of prognostic variables may further refine follow-up strategies, allowing for timely adjustments in treatment plans and providing a more attentive monitoring of patients at higher risk for recurrence/progression. Indeed, both calculators significantly discriminated low *versus* high-risk patients in cohort 1, which comprises all PCa grades (Supplementary Figure 9).

DISCUSSION AND CONCLUSION

PCa ranks as the second most frequent malignancy in men worldwide, and due to the intensive PSA screening, nearly 90% of the tumors are diagnosed as organ-confined (1, 26). In clinical practice, therapeutic strategies often depend on tools such as Clinical Risk Assessment or the CAPRA score. Specifically, for low- and intermediate-risk tumors, characterized by GG1 or GG2, active surveillance is frequently considered (50). Nonetheless, the choice between active surveillance or curative-intent treatment is still controversial, since a minor group of these patients ultimately experience disease progression or long-term recurrences. Consequently, there is a pressing need to identify new markers with prognostic value, as these markers can play a pivotal role in helping clinicians in selecting the most suitable treatment approach for individual patients.

GSTP1 promoter hypermethylation is an early event in prostate carcinogenesis. Previous studies have established a correlation between *GSTP1* hypermethylation and certain clinico-pathological features (77, 100). Specifically, *GSTP1* hypermethylation has been associated with higher clinical stage and higher grade, being a marker for PCa aggressiveness (77, 100). In our study, we observed analogous results concerning clinical stage, with stage 3 tumors harboring higher levels of *GSTP1* hypermethylation compared to stage 2 tumors. Conversely, no association was found for tumor grade. While *GSTP1* methylation holds promise as a diagnostic biomarker, its performance in predicting prognosis was rather limited. KLF8 expression has been studied in several tumor models, due to its role in epithelial-mesenchymal transition (EMT) (97-99). However, only a few studies described the role of *KLF8* in PCa, and its regulation by promoter methylation remains poorly understood (101, 102). Consequently, *KLF8* methylation levels' value as a biomarker has yet to be explored. Notably, in our study, we found that low *KLF8* methylation levels were significantly associated with worse overall survival for patients with low- and intermediate-risk PCa. In the line with our results, a previous study showed that higher KLF8 immunoexpression was associated with aggressive features, resulting in patients' worse overall survival (78). Additionally, our data reinforces the value of Ki-67 as the most promising biomarker for predicting patients' prognosis. Herein, we found that elevated Ki-67 immuno-score correlates with higher GG, suggesting that tumors exhibiting elevated Ki-67 expression tend to be more aggressive. Likewise, high Ki-67 immunoexpression was significantly associated with shorter BCR-free survival. These findings are in agreement with data gathered for other tumor models (103-105), although no consistent results are available for PCa, mostly due to the different cutoff values used to define low and high Ki-67 immunoexpression (66, 83, 106). Therefore, in our study, we selected the cutoff values

based on a previous study, where several cutoff values were tested to define low and high Ki-67 indexes (83).

Ultimately, considering low- and intermediate-risk PCa patients, we successfully formulated two distinct risk calculators. The “risk of death” calculator incorporated relevant clinico-pathological data as well as *KLF8* methylation levels, herein unveiled as independent biomarker of overall survival in PCa patients. Indeed, since we used overall survival instead of disease-specific survival, our model was widely influenced by the age at diagnosis variable, outlined as “AGEcat”. PCa is predominantly diagnosed in elderly men and, consequently, the risk of death tends to be elevated with age regardless of other cancer-related characteristics. Additionally, we developed a “BCR-risk” calculator, designated ProstARK, specifically tailored for GG1 and GG2 PCa patients. This graphic tool was constructed based on Ki-67 immunoexpression along with relevant clinico-pathological factors routinely considered by clinicians, including age at diagnosis, PSA serum levels, and clinical stage. Remarkably, in our model, greater significance was assigned to Ki-67 immunoexpression compared to PSA serum levels, widely used in the clinic routine for PCa patient monitoring. Indeed, PSA alone is not an ideal biomarker, due to its limited specificity that restricts its use as a screening and monitoring tool for PCa. However, the combination of PSA values with other biomarkers, such as Ki-67, seems to improve the prognostic accuracy for those same patients.

After the evaluation of the performance of the risk of BCR calculator model using ROC curves, validation was performed in an independent cohort of biopsies of localized prostate tumors. The use of diagnostic biopsies to validate our model was imperative to assess the practicability of this risk of recurrence calculator in clinical practice. Importantly, the use of this graphic calculator in patients with low- and intermediate-risk tumors would improve the accuracy of the therapeutic-decision making. Thus, the group of patients that progress on active surveillance may be reduced.

Currently, other risk calculators based on genome sequencing analysis are proved to assist in the clinical decision-making process between curative-intent treatment and active surveillance. Regarding Decipher® score, in a retrospective study including only very-low, low, and favorable intermediate-risk PCa patients (GG1 and GG2), the risk of adverse pathology was evaluated and the Decipher® score had an AUC of 0.65 (107, 108). Therefore, the use of this tool to select the patients that may benefit from active surveillance is rather limited, being the Decipher® score specifically indicated to guide the post-prostatectomy treatment (12). In our study, we performed a ROC analysis to evaluate the performance of ProstARK in predicting BCR. Specifically, considering all patients, we

obtained an AUC of 0.766, that slightly decreased to 0.752 when we stratify our analysis for low- and favorable intermediate-risk patients (GG1 and GG2). Thus, our risk calculator obtained better results when comparing to Oncotype DX®, that had an AUC of 0.68 in a retrospective study with very-low, low and intermediate-risk PCa patients (109). In line with our risk calculator, Prolaris® test is a genomic test that also provides information about tumor cells' proliferation. Indeed, the results obtained for predicting BCR recurrence was better than ours, with an AUC of 0.825 (110). However, the high cost entails a significant obstacle to its widespread implementation in clinical practice. Hence, here we propose a "risk of recurrence" calculator model more cost-effective, using a more user-friendly and less expensive approach, such as IHC.

The combination of each patient's clinico-pathological features and Ki-67 immunoexpression levels provide a risk calculator easy to implement for PCa prognostication. Our future research endeavors will involve the validation of ProstARK in a larger prospective cohort of PCa patients, at diagnosis. Furthermore, the flexibility of our model allows for the integration of other prognostic biomarkers for predicting patient's outcomes and improve therapeutic decision-making.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians; 71(3):209-49.
2. Tan DS, Mok TS, Rebbeck TR. 2016. Cancer Genomics: Diversity and Disparity Across Ethnicity and Geography. J Clin Oncol; 34(1):91-101.
3. Mottet N, van den Bergh RCN, Briers E, Van den Broeck T, Cumberbatch MG, De Santis M, et al. 2021. EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer-2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. Eur Urol; 79(2):243-62.
4. Das S, Salami SS, Spratt DE, Kaffenberger SD, Jacobs MF, Morgan TM. 2019. Bringing Prostate Cancer Germline Genetics into Clinical Practice. J Urol; 202(2):223-30.
5. Grossman DC, Curry SJ, Owens DK, Bibbins-Domingo K, Caughey AB, Davidson KW, et al. 2018. Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement. Jama; 319(18):1901-13.
6. Cupp MR, Oesterling JE. 1993. Prostate-specific antigen, digital rectal examination, and transrectal ultrasonography: their roles in diagnosing early prostate cancer. Mayo Clin Proc; 68(3):297-306.
7. Ragsdale JW, 3rd, Halstater B, Martinez-Bianchi V. 2014. Prostate cancer screening. Prim Care; 41(2):355-70.
8. Steggall MJ. 2008. Digital rectal examination. Nurs Stand; 22(47):46-8.
9. Naji L, Randhawa H, Sohani Z, Dennis B, Lautenbach D, Kavanagh O, et al. 2018. Digital Rectal Examination for Prostate Cancer Screening in Primary Care: A Systematic Review and Meta-Analysis. Ann Fam Med; 16(2):149-54.
10. Balk SP, Ko YJ, Bubley GJ. 2003. Biology of prostate-specific antigen. J Clin Oncol; 21(2):383-91.
11. Catalona WJ, Smith DS, Ratliff TL, Dodds KM, Coplen DE, Yuan JJ, et al. 1991. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. N Engl J Med; 324(17):1156-61.
12. Farha MW, Salami SS. 2022. Biomarkers for prostate cancer detection and risk stratification. Ther Adv Urol; 14:17562872221103988.
13. Pilatz A, Veeratterapillay R, Köves B, Cai T, Bartoletti R, Wagenlehner F, et al. 2019. Update on Strategies to Reduce Infectious Complications After Prostate Biopsy. Eur Urol Focus; 5(1):20-8.
14. Shariat SF, Roehrborn CG. 2008. Using biopsy to detect prostate cancer. Rev Urol; 10(4):262-80.
15. Litwin MS, Tan HJ. 2017. The Diagnosis and Treatment of Prostate Cancer: A Review. Jama; 317(24):2532-42.
16. Bjurlin MA, Carter HB, Schellhammer P, Cookson MS, Gomella LG, Troyer D, et al. 2013. Optimization of initial prostate biopsy in clinical practice: sampling, labeling and specimen processing. J Urol; 189(6):2039-46.
17. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. 2016. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. Am J Surg Pathol; 40(2):244-52.
18. Descotes JL. 2019. Diagnosis of prostate cancer. Asian J Urol; 6(2):129-36.
19. Edge SB, Edge SB. 2017. AJCC cancer staging manual 8th ed: Springer.
20. Epstein JI, Zelefsky MJ, Sjoberg DD, Nelson JB, Egevad L, Magi-Galluzzi C, et al. 2016. A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score. Eur Urol; 69(3):428-35.
21. Wang G, Zhao D, Spring DJ, DePinho RA. 2018. Genetics and biology of prostate cancer. Genes Dev; 32(17-18):1105-40.
22. Sandhu S, Moore CM, Chiong E, Beltran H, Bristow RG, Williams SG. 2021. Prostate cancer. Lancet; 398(10305):1075-90.

23. Parker C, Castro E, Fizazi K, Heidenreich A, Ost P, Procopio G, et al. 2020. Prostate cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol; 31(9):1119-34.
24. Gillessen S, Attard G, Beer TM, Beltran H, Bossi A, Bristow R, et al. 2018. Management of Patients with Advanced Prostate Cancer: The Report of the Advanced Prostate Cancer Consensus Conference APCCC 2017. Eur Urol; 73(2):178-211.
25. Cornford P, van den Bergh RCN, Briers E, Van den Broeck T, Cumberbatch MG, De Santis M, et al. 2021. EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer. Part II-2020 Update: Treatment of Relapsing and Metastatic Prostate Cancer. Eur Urol; 79(2):263-82.
26. 2015. The Molecular Taxonomy of Primary Prostate Cancer. Cell; 163(4):1011-25.
27. Lee CH, Akin-Olugbade O, Kirschenbaum A. 2011. Overview of prostate anatomy, histology, and pathology. Endocrinol Metab Clin North Am; 40(3):565-75, viii-ix.
28. Haffner MC, Zwart W, Roudier MP, True LD, Nelson WG, Epstein JI, et al. 2021. Genomic and phenotypic heterogeneity in prostate cancer. Nat Rev Urol; 18(2):79-92.
29. Mohanty SK, Lobo A, Cheng L. 2023. The 2022 revision of the World Health Organization classification of tumors of the urinary system and male genital organs: advances and challenges. Hum Pathol; 136:123-43.
30. Humphrey PA. 2012. Histological variants of prostatic carcinoma and their significance. Histopathology; 60(1):59-74.
31. Humphrey PA. 2017. Histopathology of Prostate Cancer. Cold Spring Harb Perspect Med; 7(10).
32. Schoenborn JR, Nelson P, Fang M. 2013. Genomic profiling defines subtypes of prostate cancer with the potential for therapeutic stratification. Clin Cancer Res; 19(15):4058-66.
33. Jillson LK, Yette GA, Laajala TD, Tilley WD, Costello JC, Cramer SD. 2021. Androgen Receptor Signaling in Prostate Cancer Genomic Subtypes. Cancers (Basel); 13(13).
34. Qian C, Li D, Chen Y. 2022. ETS factors in prostate cancer. Cancer Lett; 530:181-9.
35. van den Ouden D, Schröder FH. 2000. Management of locally advanced prostate cancer. 1. Staging, natural history, and results of radical surgery. World J Urol; 18(3):194-203.
36. Gandaglia G, Abdollah F, Schiffmann J, Trudeau V, Shariat SF, Kim SP, et al. 2014. Distribution of metastatic sites in patients with prostate cancer: A population-based analysis. Prostate; 74(2):210-6.
37. Davies A, Conteduca V, Zoubeidi A, Beltran H. 2019. Biological Evolution of Castration-resistant Prostate Cancer. Eur Urol Focus; 5(2):147-54.
38. Morote J, Aguilar A, Planas J, Trilla E. 2022. Definition of Castrate Resistant Prostate Cancer: New Insights. Biomedicines; 10(3).
39. Labbé DP, Brown M. 2018. Transcriptional Regulation in Prostate Cancer. Cold Spring Harb Perspect Med; 8(11).
40. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, et al. 2015. Integrative clinical genomics of advanced prostate cancer. Cell; 161(5):1215-28.
41. Chandrasekar T, Yang JC, Gao AC, Evans CP. 2015. Mechanisms of resistance in castration-resistant prostate cancer (CRPC). Transl Androl Urol; 4(3):365-80.
42. Jamroze A, Chatta G, Tang DG. 2021. Androgen receptor (AR) heterogeneity in prostate cancer and therapy resistance. Cancer Lett; 518:1-9.
43. Aggarwal R, Zhang T, Small EJ, Armstrong AJ. 2014. Neuroendocrine prostate cancer: subtypes, biology, and clinical outcomes. J Natl Compr Canc Netw; 12(5):719-26.
44. Yamada Y, Beltran H. 2021. Clinical and Biological Features of Neuroendocrine Prostate Cancer. Curr Oncol Rep; 23(2):15.
45. Xie Y, Ning S, Hu J. 2022. Molecular mechanisms of neuroendocrine differentiation in prostate cancer progression. J Cancer Res Clin Oncol; 148(7):1813-23.
46. Logothetis CJ, Gallick GE, Maity SN, Kim J, Aparicio A, Efstathiou E, et al. 2013. Molecular classification of prostate cancer progression: foundation for marker-driven treatment of prostate cancer. Cancer Discov; 3(8):849-61.

47. Zhao SG, Chang SL, Erho N, Yu M, Lehrer J, Alshalalfa M, et al. 2017. Associations of Luminal and Basal Subtyping of Prostate Cancer With Prognosis and Response to Androgen Deprivation Therapy. JAMA Oncol; 3(12):1663-72.
48. You S, Knudsen BS, Erho N, Alshalalfa M, Takhar M, Al-Deen Ashab H, et al. 2016. Integrated Classification of Prostate Cancer Reveals a Novel Luminal Subtype with Poor Outcome. Cancer Res; 76(17):4948-58.
49. Yoon J, Kim M, Posadas EM, Freedland SJ, Liu Y, Davicioni E, et al. 2021. A comparative study of PCS and PAM50 prostate cancer classification schemes. Prostate Cancer Prostatic Dis; 24(3):733-42.
50. 2017. Prolaris Cell Cycle Progression Test for Localized Prostate Cancer: A Health Technology Assessment. Ont Health Technol Assess Ser; 17(6):1-75.
51. Preisser F, Cooperberg MR, Crook J, Feng F, Graefen M, Karakiewicz PI, et al. 2020. Intermediate-risk Prostate Cancer: Stratification and Management. Eur Urol Oncol; 3(3):270-80.
52. Cooperberg MR, Broering JM, Carroll PR. 2010. Time trends and local variation in primary treatment of localized prostate cancer. J Clin Oncol; 28(7):1117-23.
53. Cooperberg MR, Broering JM, Carroll PR. 2009. Risk assessment for prostate cancer metastasis and mortality at the time of diagnosis. J Natl Cancer Inst; 101(12):878-87.
54. Chang AJ, Autio KA, Roach M, 3rd, Scher HI. 2014. High-risk prostate cancer-classification and therapy. Nat Rev Clin Oncol; 11(6):308-23.
55. Mohler JL, Antonarakis ES, Armstrong AJ, D'Amico AV, Davis BJ, Dorff T, et al. 2019. Prostate Cancer, Version 2.2019, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw; 17(5):479-505.
56. Eggener SE, Rumble RB, Armstrong AJ, Morgan TM, Crispino T, Cornford P, et al. 2020. Molecular Biomarkers in Localized Prostate Cancer: ASCO Guideline. J Clin Oncol; 38(13):1474-94.
57. Shore N, Concepcion R, Saltzstein D, Lucia MS, van Breda A, Welbourn W, et al. 2014. Clinical utility of a biopsy-based cell cycle gene expression assay in localized prostate cancer. Curr Med Res Opin; 30(4):547-53.
58. Reiter RE. 2016. Risk stratification of prostate cancer 2016. Scand J Clin Lab Invest Suppl; 245:S54-9.
59. Knezevic D, Goddard AD, Natraj N, Cherbavaz DB, Clark-Langone KM, Snable J, et al. 2013. Analytical validation of the Oncotype DX prostate cancer assay - a clinical RT-PCR assay optimized for prostate needle biopsies. BMC Genomics; 14:690.
60. Klein EA, Cooperberg MR, Magi-Galluzzi C, Simko JP, Falzarano SM, Maddala T, et al. 2014. A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. Eur Urol; 66(3):550-60.
61. Daidone MG, Costa A, Silvestrini R. 2001. Cell proliferation markers in human solid tumors: assessing their impact in clinical oncology. Methods Cell Biol; 64:359-84.
62. Sahin AA, Ro JY, el-Naggar AK, Wilson PL, Teague K, Blick M, et al. 1991. Tumor proliferative fraction in solid malignant neoplasms. A comparative study of Ki-67 immunostaining and flow cytometric determinations. Am J Clin Pathol; 96(4):512-9.
63. Deshmukh P, Ramsey L, Garewal HS. 1990. Ki-67 labeling index is a more reliable measure of solid tumor proliferative activity than tritiated thymidine labeling. Am J Clin Pathol; 94(2):192-5.
64. Elias JM. 1997. Cell proliferation indexes: a biomarker in solid tumors. Biotech Histochem; 72(2):78-85.
65. Kammerer-Jacquet SF, Ahmad A, Møller H, Sandu H, Scardino P, Soosay G, et al. 2019. Ki-67 is an independent predictor of prostate cancer death in routine needle biopsy samples: proving utility for routine assessments. Mod Pathol; 32(9):1303-9.
66. Berlin A, Castro-Mesta JF, Rodriguez-Romo L, Hernandez-Barajas D, González-Guerrero JF, Rodríguez-Fernández IA, et al. 2017. Prognostic role of Ki-67 score in localized prostate cancer: A systematic review and meta-analysis. Urol Oncol; 35(8):499-506.
67. Lotan TL, Tomlins SA, Bismar TA, Van der Kwast TH, Grignon D, Egevad L, et al. 2020. Report From the International Society of Urological Pathology (ISUP) Consultation Conference on

Molecular Pathology of Urogenital Cancers. I. Molecular Biomarkers in Prostate Cancer. Am J Surg Pathol; 44(7):e15-e29.
68. Cyr AR, Domann FE. 2011. The redox basis of epigenetic modifications: from mechanisms to functional consequences. Antioxid Redox Signal; 15(2):551-89.
69. Darwiche N. 2020. Epigenetic mechanisms and the hallmarks of cancer: an intimate affair. Am J Cancer Res; 10(7):1954-78.
70. Hanahan D. 2022. Hallmarks of Cancer: New Dimensions. Cancer Discov; 12(1):31-46.
71. Moore LD, Le T, Fan G. 2013. DNA methylation and its basic function. Neuropsychopharmacology; 38(1):23-38.
72. Jurkowska RZ, Jurkowski TP, Jeltsch A. 2011. Structure and function of mammalian DNA methyltransferases. Chembiochem; 12(2):206-22.
73. Hassler MR, Redl E, Hudson QJ, Miller WJ, Egger G. 2016. Chapter 1 - Basic Epigenetic Mechanisms and Phenomena. In: Egger G, Arimondo P, editors. Drug Discovery in Cancer Epigenetics. Boston: Academic Press; 2016. p. 3-40.
74. Zhang W, Xu J. 2017. DNA methyltransferases and their roles in tumorigenesis. Biomark Res; 5:1.
75. Tzelepi V, Logotheti S, Efstathiou E, Troncoso P, Aparicio A, Sakellakis M, et al. 2020. Epigenetics and prostate cancer: defining the timing of DNA methyltransferase deregulation during prostate cancer progression. Pathology; 52(2):218-27.
76. Yegnasubramanian S, De Marzo AM, Nelson WG. 2019. Prostate Cancer Epigenetics: From Basic Mechanisms to Clinical Implications. Cold Spring Harb Perspect Med; 9(4).
77. Jerónimo C, Henrique R, Hoque MO, Mambo E, Ribeiro FR, Varzim G, et al. 2004. A quantitative promoter methylation profile of prostate cancer. Clin Cancer Res; 10(24):8472-8.
78. He HJ, Gu XF, Xu WH, Yang DJ, Wang XM, Su Y. 2013. Krüppel-like factor 8 is a novel androgen receptor co-activator in human prostate cancer. Acta Pharmacol Sin; 34(2):282-8.
79. Ahmed H. 2010. Promoter methylation in prostate cancer and its application for the early detection of prostate cancer using serum and urine samples. Biomark Cancer; 2:17-33.
80. Di Meo A, Bartlett J, Cheng Y, Pasic MD, Yousef GM. 2017. Liquid biopsy: a step forward towards precision medicine in urologic malignancies. Mol Cancer; 16(1):80.
81. Han X, Wang J, Sun Y. 2017. Circulating Tumor DNA as Biomarkers for Cancer Detection. Genomics Proteomics Bioinformatics; 15(2):59-72.
82. Constâncio V, Nunes SP, Henrique R, Jerónimo C. 2020. DNA Methylation-Based Testing in Liquid Biopsies as Detection and Prognostic Biomarkers for the Four Major Cancer Types. Cells; 9(3).
83. Lobo J, Rodrigues Â, Antunes L, Graça I, Ramalho-Carvalho J, Vieira FQ, et al. 2018. High immunoexpression of Ki67, EZH2, and SMYD3 in diagnostic prostate biopsies independently predicts outcome in patients with prostate cancer. Urol Oncol; 36(4):161.e7-.e17.
84. Li Y, Ge D, Lu C. 2019. The SMART App: an interactive web application for comprehensive DNA methylation analysis and visualization. Epigenetics Chromatin; 12(1):71.
85. Jerónimo C, Usadel H, Henrique R, Oliveira J, Lopes C, Nelson WG, et al. 2001. Quantitation of GSTP1 methylation in non-neoplastic prostatic tissue and organ-confined prostate adenocarcinoma. J Natl Cancer Inst; 93(22):1747-52.
86. Baures M, Dariane C, Tika E, Puig Lombardi E, Barry Delongchamps N, Blanpain C, et al. 2022. Prostate luminal progenitor cells: from mouse to human, from health to disease. Nat Rev Urol.
87. Verhagen AP, Ramaekers FC, Aalders TW, Schaafsma HE, Debruyne FM, Schalken JA. 1992. Colocalization of basal and luminal cell-type cytokeratins in human prostate cancer. Cancer Res; 52(22):6182-7.
88. Nagle RB, Brawer MK, Kittelson J, Clark V. 1991. Phenotypic relationships of prostatic intraepithelial neoplasia to invasive prostatic carcinoma. Am J Pathol; 138(1):119-28.
89. Sherwood ER, Berg LA, Mitchell NJ, McNeal JE, Kozlowski JM, Lee C. 1990. Differential cytokeratin expression in normal, hyperplastic and malignant epithelial cells from human prostate. J Urol; 143(1):167-71.

90. Zhang D, Zhao S, Li X, Kirk JS, Tang DG. 2018. Prostate Luminal Progenitor Cells in Development and Cancer. Trends Cancer; 4(11):769-83.
91. Shibata M, Shen MM. 2013. The roots of cancer: stem cells and the basis for tumor heterogeneity. Bioessays; 35(3):253-60.
92. Visvader JE. 2011. Cells of origin in cancer. Nature; 469(7330):314-22.
93. Lawson DA, Witte ON. 2007. Stem cells in prostate cancer initiation and progression. J Clin Invest; 117(8):2044-50.
94. Prat A, Pineda E, Adamo B, Galván P, Fernández A, Gaba L, et al. 2015. Clinical implications of the intrinsic molecular subtypes of breast cancer. Breast; 24 Suppl 2:S26-35.
95. Dall'Era MA, Albertsen PC, Bangma C, Carroll PR, Carter HB, Cooperberg MR, et al. 2012. Active surveillance for prostate cancer: a systematic review of the literature. Eur Urol; 62(6):976-83.
96. Jerónimo C, Bastian PJ, Bjartell A, Carbone GM, Catto JW, Clark SJ, et al. 2011. Epigenetics in prostate cancer: biologic and clinical relevance. Eur Urol; 60(4):753-66.
97. Liang K, Liu T, Chu N, Kang J, Zhang R, Yu Y, et al. 2015. KLF8 is required for bladder cancer cell proliferation and migration. Biotechnol Appl Biochem; 62(5):628-33.
98. Wang X, Lu H, Urvalek AM, Li T, Yu L, Lamar J, et al. 2011. KLF8 promotes human breast cancer cell invasion and metastasis by transcriptional activation of MMP9. Oncogene; 30(16):1901-11.
99. Yan Q, Zhang W, Wu Y, Wu M, Zhang M, Shi X, et al. 2015. KLF8 promotes tumorigenesis, invasion and metastasis of colorectal cancer cells by transcriptional activation of FHL2. Oncotarget; 6(28):25402-17.
100. Richiardi L, Fiano V, Vizzini L, De Marco L, Delsedime L, Akre O, et al. 2009. Promoter methylation in APC, RUNX3, and GSTP1 and mortality in prostate cancer patients. J Clin Oncol; 27(19):3161-8.
101. Haldrup C, Mundbjerg K, Vestergaard EM, Lamy P, Wild P, Schulz WA, et al. 2013. DNA methylation signatures for prediction of biochemical recurrence after radical prostatectomy of clinically localized prostate cancer. J Clin Oncol; 31(26):3250-8.
102. Møller M, Strand SH, Mundbjerg K, Liang G, Gill I, Haldrup C, et al. 2017. Heterogeneous patterns of DNA methylation-based field effects in histologically normal prostate tissue from cancer patients. Sci Rep; 7:40636.
103. Choi SB, Park JM, Ahn JH, Go J, Kim J, Park HS, et al. 2022. Ki-67 and breast cancer prognosis: does it matter if Ki-67 level is examined using preoperative biopsy or postoperative specimen? Breast Cancer Res Treat; 192(2):343-52.
104. Jeon HG, Jeong IG, Bae J, Lee JW, Won JK, Paik JH, et al. 2010. Expression of Ki-67 and COX-2 in patients with upper urinary tract urothelial carcinoma. Urology; 76(2):513.e7-12.
105. Ziaran S, Harsanyi S, Bevizova K, Varchulova Novakova Z, Trebaticky B, Bujdak P, et al. 2020. Expression of E-cadherin, Ki-67, and p53 in urinary bladder cancer in relation to progression, survival, and recurrence. Eur J Histochem; 64(2).
106. Fisher G, Yang ZH, Kudahetti S, Møller H, Scardino P, Cuzick J, et al. 2013. Prognostic value of Ki-67 for prostate cancer death in a conservatively managed cohort. Br J Cancer; 108(2):271-7.
107. Kim HL, Li P, Huang HC, Deheshi S, Marti T, Knudsen B, et al. 2019. Validation of the Decipher Test for predicting adverse pathology in candidates for prostate cancer active surveillance. Prostate Cancer Prostatic Dis; 22(3):399-405.
108. Herlemann A, Huang HC, Alam R, Tosoian JJ, Kim HL, Klein EA, et al. 2020. Decipher identifies men with otherwise clinically favorable-intermediate risk disease who may not be good candidates for active surveillance. Prostate Cancer Prostatic Dis; 23(1):136-43.
109. Cullen J, Rosner IL, Brand TC, Zhang N, Tsiatis AC, Moncur J, et al. 2015. 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. Eur Urol; 68(1):123-31.

110. Cuzick J, Swanson GP, Fisher G, Brothman AR, Berney DM, Reid JE, et al. 2011. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. Lancet Oncol; 12(3):245-55.
111. Raghav K, Hwang H, Jácome AA, Bhang E, Willett A, Huey RW, et al. 2021. Development and Validation of a Novel Nomogram for Individualized Prediction of Survival in Cancer of Unknown Primary. Clin Cancer Res; 27(12):3414-21.
112. Zhang Z, Kattan MW. 2017. Drawing Nomograms with R: applications to categorical outcome and survival data. Ann Transl Med; 5(10):211.

Appendix

Supplementary Table 1

Supplementary Table 1- Detailed clinico-pathological data of a series of prostate cancer biopsies, previously described in our group (83)

	All series	Low- and favorable intermediate-risk patients (GG1 and 2)
Clinico-Pathological Characteristics	**Mean ± standard deviation, total n=153**	**Mean ± standard deviation, total n=42**
Age (Years)	71.7 ± 6.78	69.2 ± 5.80
Clinico-Pathological Characteristics	**Patient Numbers n (%), total n=153**	**Patient Numbers n (%), total n=42**
Age (Categorized- years)		
≤55	2 (1.31%)	1 (2.38%)
55-65	27 (17.65%)	7 (16.67%)
>65	124 (81.04%)	34 (80.95%)
Serum PSA concentration (ng/mL)		
≤10	18 (11.77%)	12 (28.58%)
10-20	29 (18.95%)	15 (35.71%)
>20	106 (69.28%)	15 (35.71%)
ISUP Grade Group		
1	17 (11.11%)	17 (40.48%)
2	25 (16.34%)	25 (59.52%)
3	36 (23.53%)	N/A
4	60 (39.22%)	N/A
5	15 (9.80%)	N/A
Clinical Stage		
cT1	2 (1.31%)	1 (2.38%)
cT2	47 (30.72%)	29 (69.05%)
cT3	59 (38.56%)	11 (26.19%)
cT4	45 (29.41%)	1 (2.38%)
Biochemical Recurrence	112 (73.20%)	19 (45.24%)
Deaths	122 (79.74%)	23 (54.76%)

Abbreviations: N/A- Non applicable

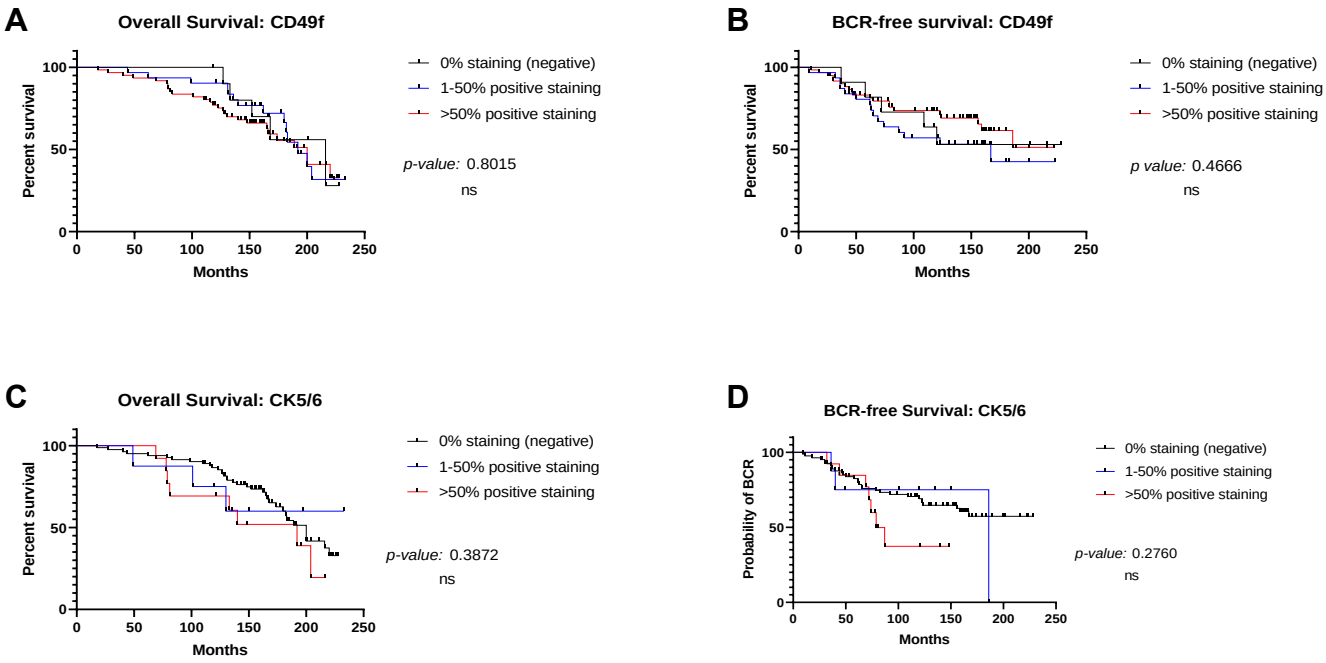
Supplementary Table 2

Supplementary Table 2- Optimized immunohistochemistry conditions for each antibody

Antibody	Dilution	Expression	Positive Control	Antigen Retrieval	Buffer	Incubation	Ab Reference	Company
Anti-NKX3.1	1:200	Nuclear	Prostate	20' MW + 10' RT	Citrate 1x	ON-RT	441R-14	Merck
Anti-Ki-67	1:150	Nuclear	Lymphoma	20' MW + 10' RT	Citrate 1x	ON-RT	M7240	Dako
Anti-CD49f	1:250	Cytoplasmic / Membrane	Colorectal Cancer	20' MW	Citrate 1x	ON-RT	HPA012696	Merck
Anti- CK5/6	1:200	Cytoplasmic	Squamous Cell Carcinoma	20' MW	EDTA 1x	1h-RT	Cytokeratin 5 & 6 (D5 & 16B4)	Cell Marque
Anti-CK14	1:300	Cytoplasmic	Prostate	20' MW	EDTA 1x	1h-RT	NCL-L-LL002	Leica Biosystems
Anti-CK8/18	1:150	Cytoplasmic	Thyroid	20' MW	EDTA 1x	1h-RT	NCL-L-5D3	Leica Biosystems

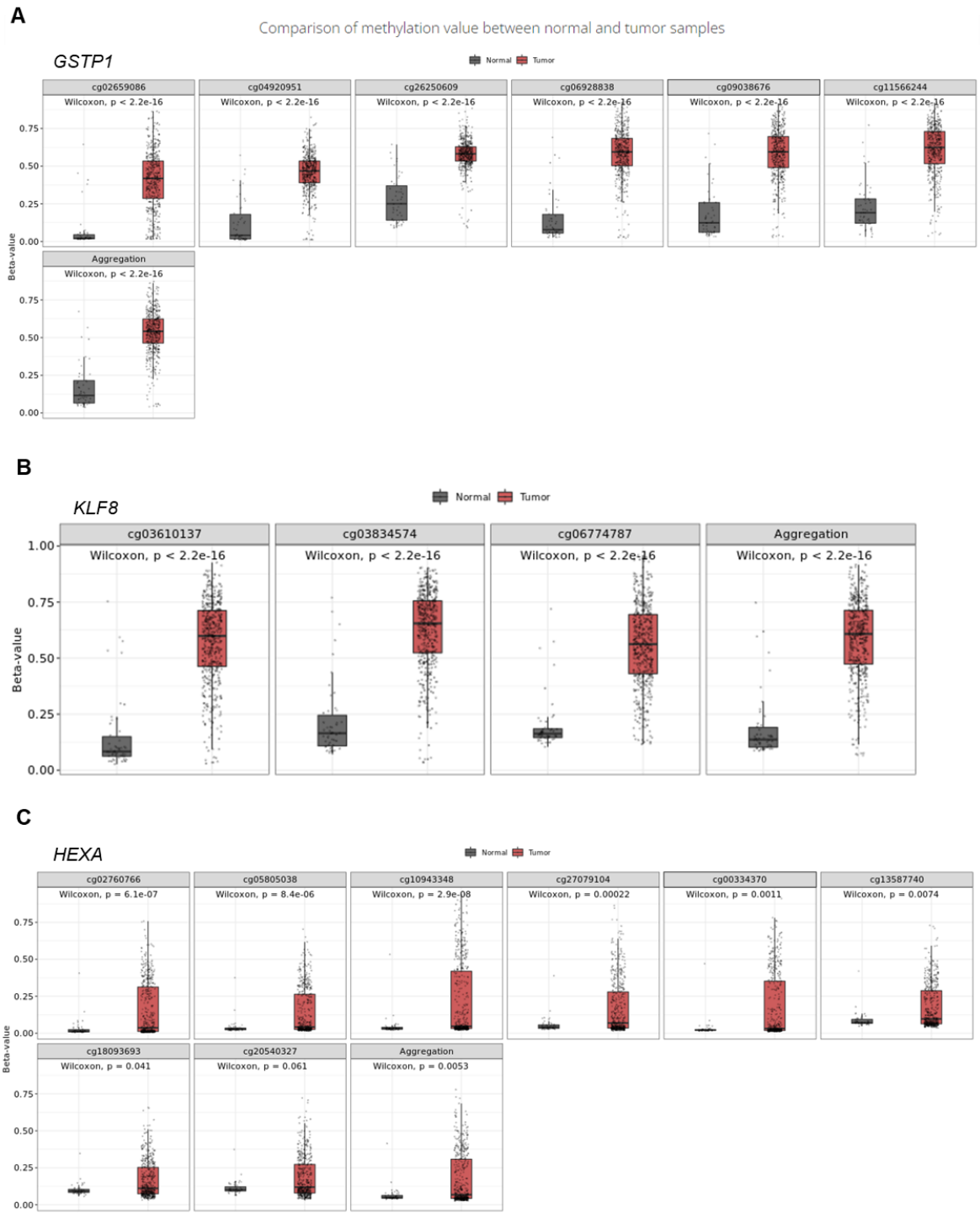
Abbreviations: Ab- antibody; MW- microwave; RT- room temperature; ON- overnight

Supplementary Figure 1



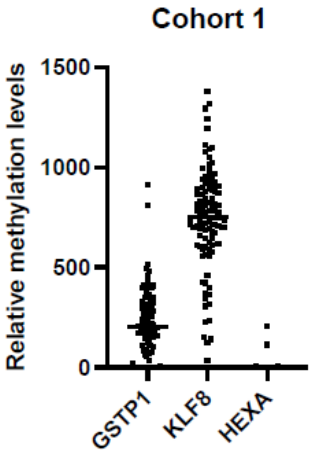
Supplementary Figure 1- Survival analysis based on the CK5/6 and CD49f immunoexpression. Kaplan-Meier curves for the overall **(A)** and BCR-free survival **(B)**, based on the CD49f immunoexpression. Kaplan-Meier curves, based on the CK5/6 immunoexpression, for the overall survival **(C)** and BCR-free survival **(D)**. The Log-Rank test was used to evaluate the differences in overall survival and BCR-free survival, where none of the analyses were statistically significant.

Supplementary Figure 2



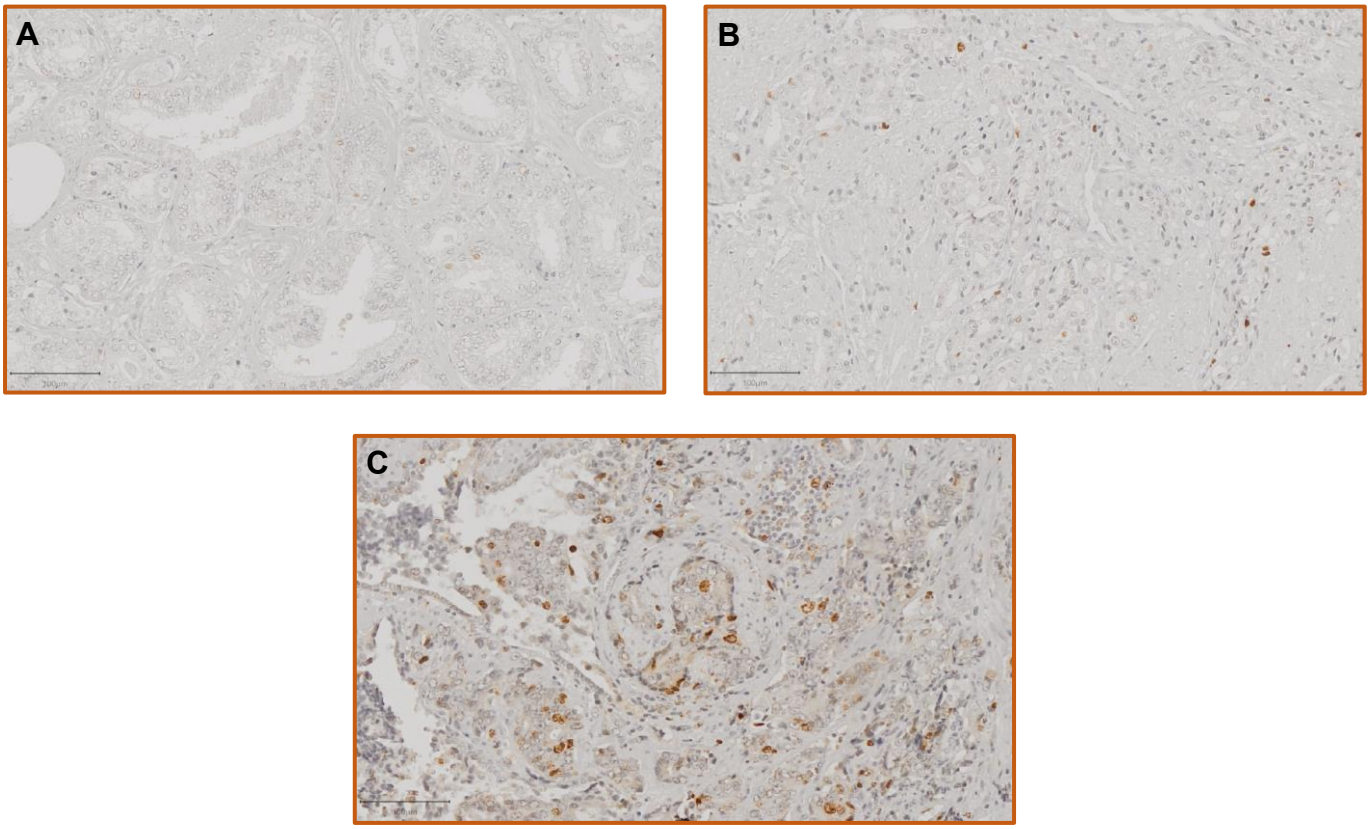
Supplementary Figure 2- Comparison of methylation levels between normal and tumor tissue of the specific CpG sites of *GSTP1* **(A)**, *KLF8* **(B)**, and *HEXA* **(C)** gene promoters, unveiled by the *in silico* analysis.

Supplementary Figure 3



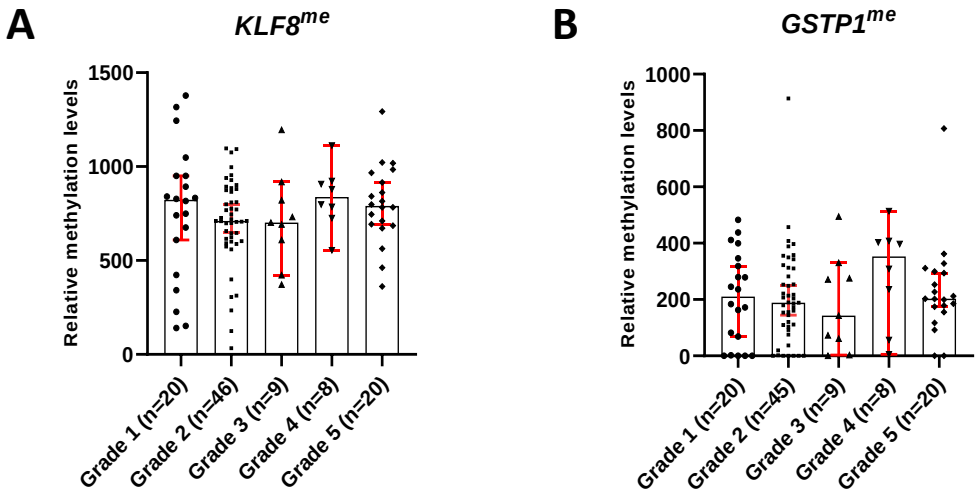
Supplementary Figure 3- Relative Methylation levels of *GSTP1*, *KLF8,* and *HEXA* promoter regions.

Supplementary Figure 4



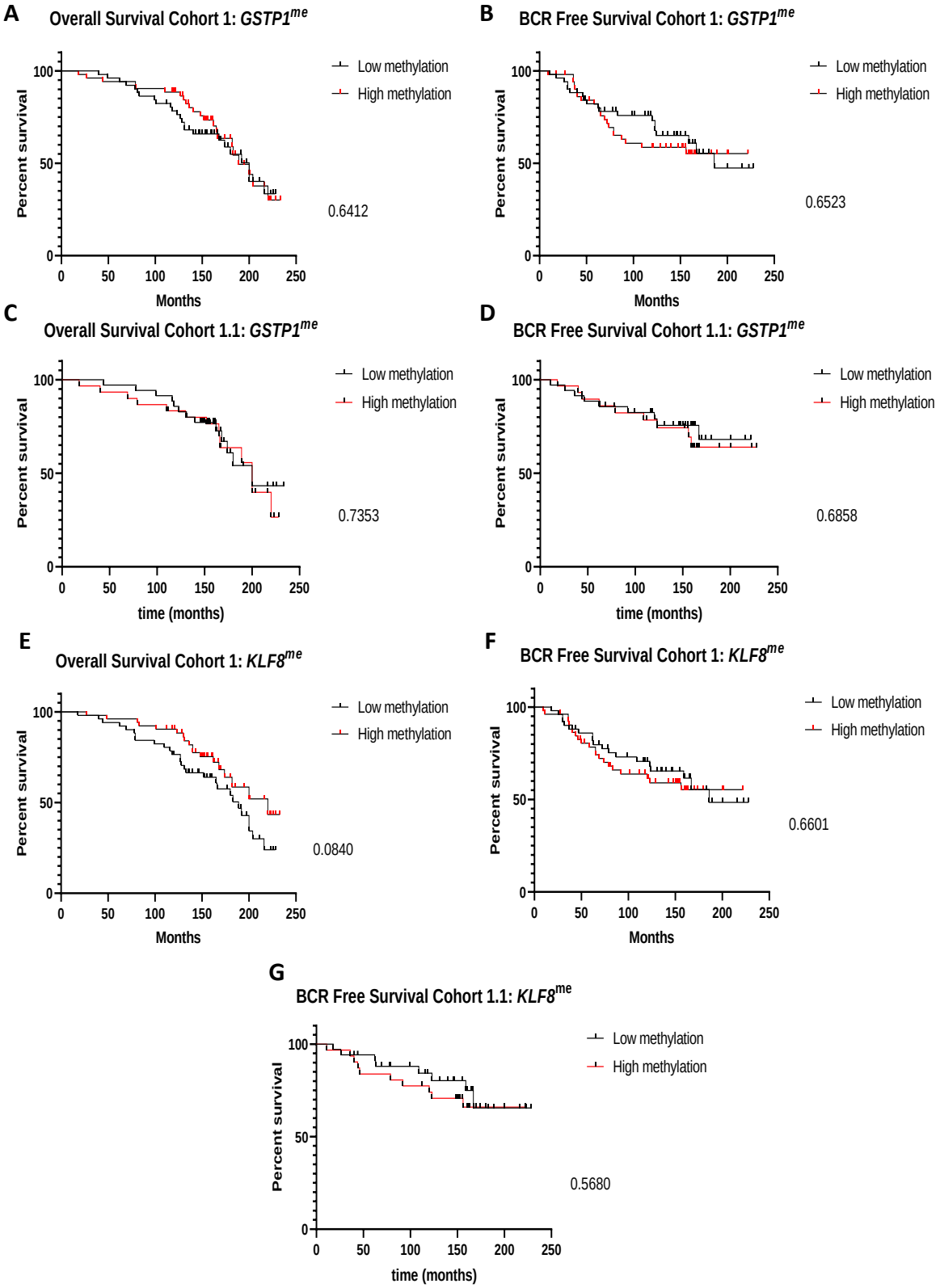
Supplementary Figure 4- Ki-67 expression in localized prostate tumors. Ki-67 nuclear staining **A-** less than 5% of positive staining cells; **B-** 5-10% of positive staining cells and **C-** more than 10% of positive staining cells. Pictures were taken at 8.36X200 magnification (scale bar 100µm).

Supplementary Figure 5



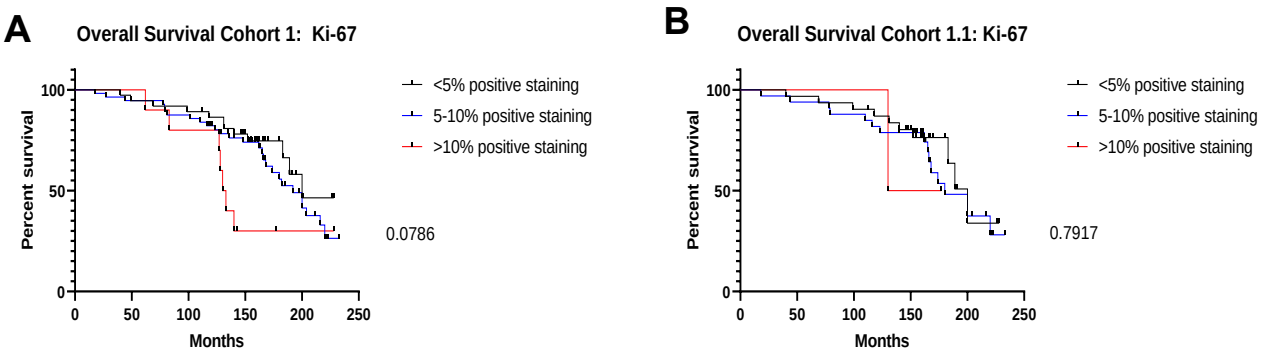
Supplementary Figure 5- Association of *GSTP1* and *KLF8* relative methylation levels with Gleason grade groups. A- *KLF8*^{me} and **B-** *GSTP1*^{me} levels were compared between Gleason grade groups. Median ranks between grade groups were compared using the Kruskal-Wallis test. No statistically significant differences were found.

Supplementary Figure 6



Supplementary Figure 6- Survival analysis based on the methylation levels of *GSTP1* and *KLF8*. Kaplan-Meier curves for the overall survival **(A)** and BCR-free survival **(B)**, in cohort 1. Kaplan-Meier for the overall survival **(C)** and BCR-free survival **(D)**, in cohort 1.1. Kaplan-Meier for the overall survival **(E)** and BCR-free survival **(F)** based on the *KLF8* methylation levels, in cohort 1. **G-** Kaplan-Meier for the BCR-free survival based on the $KLF8^{me}$ levels, in cohort 1.1. The Log-Rank test was used to evaluate the differences in overall survival and BCR-free survival, where none of the analyses were statistically significant.

Supplementary Figure 7



Supplementary Figure 7- Survival analysis based on the Ki-67 immunoexpression and *KLF8*^{me} levels. Kaplan-Meier for the overall survival based on the expression of the Ki-67, in cohort 1 **(A)** and cohort 1.1 **(B)**. The Log-Rank test was used to evaluate the differences in overall survival and BCR-free survival, where none of the analyses were statistically significant.

Supplementary Information 1

A nomogram is a graphical calculator device based on logistic regression models. This tool is widely used in clinical practice for patient outcome prediction, helping clinicians make treatment decisions (111, 112). The first step to creating a nomogram is creating a logistic regression model that fits better with our data. All models and consequently, all nomograms were created using the R software with the following codes. Afterward, ROC curves were used to evaluate the performance of our models. The rms, caTools, epicalc, and pROC packages were installed for these analyses.

Example 1: All variables are categorized. This example was used to perform a nomogram using a logistic regression model for overall survival with all variables.

```
>Model1<-lrm(VitalStatus~AGEcat+KLF8+TStage+GG+KI67+PSA,data=Geral)

>ddist<-datadist(Geral)

>options(datadist='ddist')

>nom1<-nomogram(Model1,fun=plogis,funlabel = "Risk of Death")

>plot(nom)
```

As for ROC curves, the following codes were applied:

```
>Model1<-glm(formula=VitalStatus~AGEcat+KLF8+TStage+GG+KI67+PSA,
family=binomial(link = "logit"),data=Geral)

>summary(Model1)

>logistic.display(Model1)

>par(pty="s")

>ROCcurve1<-roc(Geral$VitalStatus~Model1$fitted,
plot=TRUE,legacy.axes=TRUE,print.auc=TRUE,ci=TRUE,lwd=4,col="#b5d08d")
```

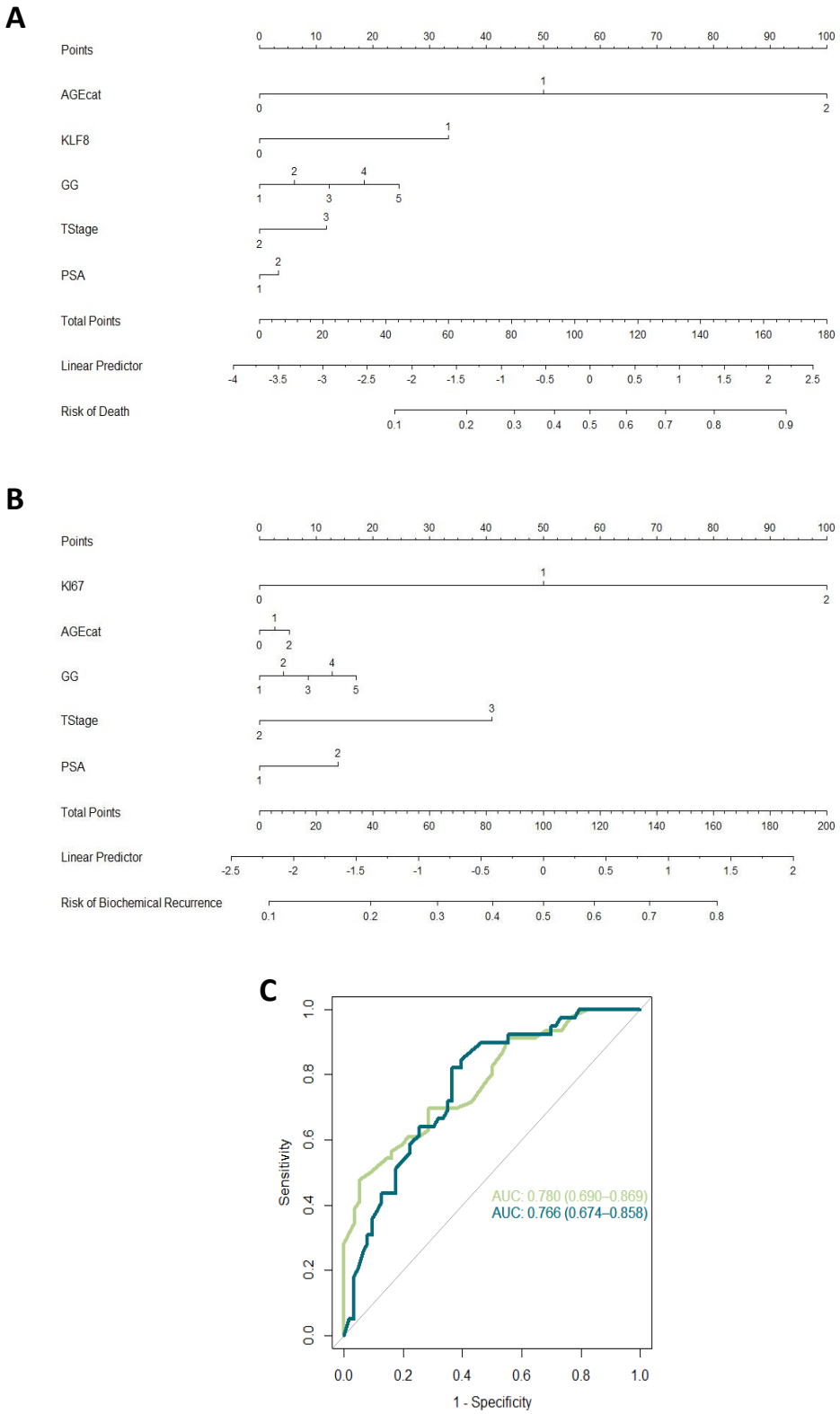
Supplementary Table 3

Supplementary Table 3- Summary of the logistic models representative of the Risk of Death and the Risk of BCR calculators, in cohort 1

					Confidence Interval	
	Coefficient	**S.E.**	**Wald Z**	***p-value***	**Lowe Bound**	**Upper Bound**
Model 3: Risk of Death Cohort 1						
Intercept	-4.8703	1.6838	-2.89	0.0038	-8.1704	-1.5702
AGEcat	1.7663	0.4357	4.05	<0.0001	0.9124	2.6202
KLF8	1.1751	0.4889	2.40	0.0162	0.2168	2.1334
GG	0.2169	0.2024	1.07	0.2839	-0.1798	0.6136
Tstage	0.4146	0.5522	0.75	0.4527	-0.6676	1.4968
PSA	0.1172	0.5055	0.23	0.8167	-0.8737	1.1080
Model 4: Risk of BCR Cohort 1						
Intercept	-4.5431	1.5207	-2.99	0.0028	-7.5237	-1.5626
AGEcat	0.0594	0.3511	0.17	0.8657	-0.6287	0.7475
Ki-67	1.1350	0.4332	2.62	0.0088	0.2859	1.9841
GG	0.0966	0.2077	0.47	0.6418	-0.3104	0.5036
Tstage	0.9295	0.5416	1.72	0.0861	-0.1320	1.9911
PSA	0.3139	0.4790	0.66	0.5123	-0.6250	1.2528

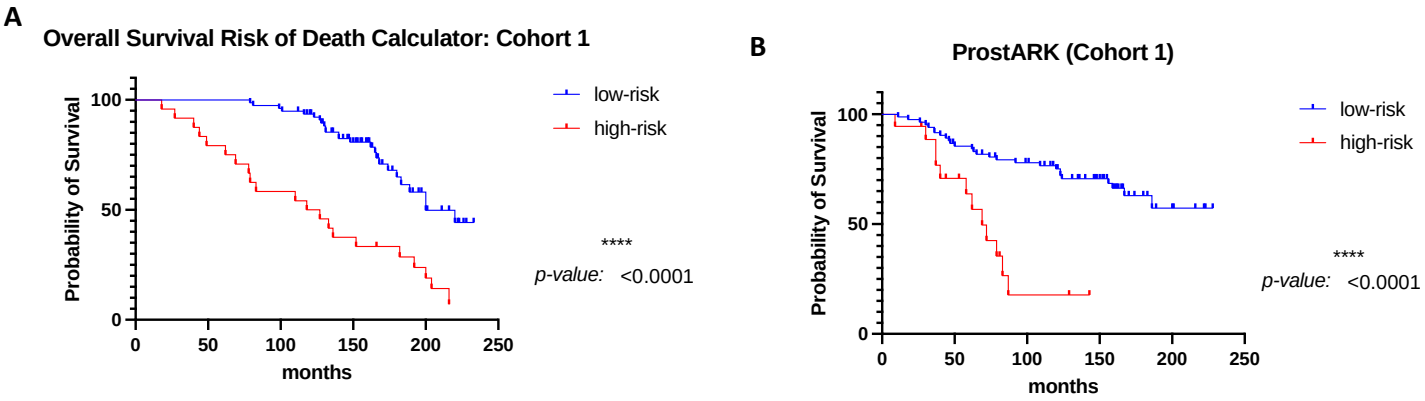
Abbreviations: S.E.- Standard Error; CI- Confidence Interval; BCR- Biochemical-recurrence; GG- Grade Group; PSA- Prostate-specific Antigen

Supplementary Figure 8



Supplementary Figure 8- Nomogram representative of of the Risk of Death **(A)** and the Risk of Biochemical-recurrence **(B)** calculators, for cohort 1, with the relevant clinico-pathological variables for the risk stratification: age at diagnosis (0: ≤55, 1: 55-65, 2: >65 years), clinical stage, grade group, and PSA serum levels (1: <10ng/mL, 2: >10ng/mL). **C-** Roc curves to evaluate the performance of the nomogram models. ***KLF8^{me}*** **categorization-** 0: high methylation levels, 1: low methylation levels. **Ki-67 immunoexpression categorization-** 0: <5%, 1: 5-10%, 2: >10% of positive staining. **Green Curve:** ROC curve for the Risk of Death calculator. **Blue Curve:** ROC curve for the Risk of Biochemical-recurrence calculator.

Supplementary Figure 9



Supplementary Figure 9- Kaplan-Meier curve for the overall survival based on the Risk of death calculator, in cohort 1. Kaplan-Meier curves for the biochemical-recurrence free survival based on the ProstARK calculator, in cohort 1.