

MESTRADO ONCOLOGIA
ONCOLOGIA LABORATORIAL

Aggressive prostate cancer GWAS SNPs and its association with clinical outcome.

Ana Rita Pinto

M

2019

Ana Rita Pinto. Aggressive prostate cancer GWAS SNPs and its association with clinical outcome.



Aggressive prostate cancer GWAS SNPs and its association with clinical outcome.
Ana Rita Mendes Pinto



Ana Rita Mendes Pinto

Aggressive prostate cancer GWAS SNPs and its association with clinical outcome.

Dissertação de Candidatura ao Grau de Mestre em Oncologia submetida ao Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto

Orientador – Professor Doutor Rui Manuel de Medeiros Melo Silva

Categoria – Professor Afiliado – ICBAS

Afiliação – Coordenador do Grupo de Oncologia Molecular & Patologia Viral do Centro de Investigação do Instituto Português de Oncologia do Porto

Co-orientadora – Doutora Jani Viviana Alves Vital da Silva

Categoria – Investigadora Pós-Doutoramento

Afiliação – Grupo de Oncologia Molecular e Patologia Viral do Centro de Investigação do Instituto Português de Oncologia do Porto

Improvise, Adapt, Overcome.

- Bear Grylls

Agradecimentos

Com a finalização desta etapa, não posso deixar de agradecer a todos aqueles que me fizeram crescer, tanto cientificamente como pessoalmente, e que de uma forma ou de outra estiveram presentes durante todo este percurso.

Deste modo, agradeço à Comissão Coordenadora do Mestrado em Oncologia, em especial à Professora Doutora Carmen Jerónimo, pela oportunidade de frequentar este mestrado e de aumentar o meu conhecimento científico.

Ao Professor Doutor Rui Medeiros, por me ter acolhido no seu grupo de investigação, por me aceitar como sua orientanda e me proporcionar esta nova experiência, e por puxar por mim nos momentos em que precisava.

À minha super co-orientadora, a Doutora Jani Silva, por todo o empenho e toda a ajuda (e por não me achincalhares), por me aturares nos momentos de pânico (e eu sei que não sou fácil de aturar) e por me animares quando precisava. Sem ti este trabalho não teria sido realizado, e por isso só te posso deixar o meu muito obrigado! Ps. E muito obrigada pela edição das imagens!

Ao Mestre Ricardo Pinto pelo incansável esforço, e que mesmo sendo um jovem investigador me auxiliou como muitos não conseguiriam.

A toda a equipa do Grupo de Oncologia Molecular e Patologia Viral, por me terem acolhido no laboratório, por toda ajuda e pelos momentos de descontração. Em especial, ao Mestre Augusto Nogueira, à Mestre Francisca Dias e à Mestre Mara Fernandes pelo apoio na parte laboratorial e na discussão de ideias, por clarearem a minha mente quando eu não via a luz ao fundo do túnel. E por fim às minhas meninas do laboratório: à Bruna, à Bela e à Joana, obrigada por todo o apoio nos momentos de desespero, pelas risadas quando só nos apetecia chorar e pelo ombro amigo para desabafar.

Aos meus amigos, porque sem vocês nada disto faria sentido. Aos meus amigos do Cerco, por me apoiarem e aturarem desde sempre e para sempre, e por acreditarem em mim mesmo quando eu não acreditava. Aos amigos que ganhei na faculdade, em especial ao meu trio fantástico de amigas que me dizem sempre a verdade, Cláudia, Andreia e Márcia, por nunca me deixarem sozinha mesmo quando estávamos separadas fisicamente. À OncoTeam, ao Augusto e à Mafalda, por embarcarem comigo nesta aventura, por todas as memórias que criamos nestes 2 anos, por me mostrarem que é

possível ultrapassar os obstáculos com a ajuda dos amigos, mas acima de tudo por sermos amebas juntos.

À Filipa, por ser a melhor amiga que é, pelos telefonemas diários com conversa da treta, pela constante preocupação e por me levantar quando teimo em cair, um milhão de obrigadas!

À Márcia, a pessoa que mais me ajudou este ano, que me aturou todos os dias quando já nem eu me conseguia aturar, que ouviu as risadas mas também os choros, que gozava comigo e de seguida me dava verdadeiros conselhos de amiga e que despendeu do seu tempo de trabalho para me acalmar quando eu stressava (que não foram poucas vezes), muito obrigada!

À minha família por tudo o que são e fazem por mim. Aos meus tios e primos, obrigada pela pergunta constante “Então como vai a tese?”, que apesar de induzir stress mostra a vossa preocupação pelo meu trabalho. Aos meus avós, obrigada por todo o apoio ao longo não só desta etapa, mas de toda a minha vida e por nunca me deixarem ir abaixo. E por fim aos meus padrinhos, por serem dois grandes pilares na minha vida, e por todas as conversas sem julgamentos.

Aos meus pais, por acreditarem sempre em mim, por me incentivarem a dar sempre mais e melhor, por me ensinarem a ser humilde no meu caminho e por nunca me deixarem fazer este percurso sozinha, o maior obrigado!

À minha irmã Carolina, por ser a minha irmã, e por me dar cabo do juízo mas por também saber sempre como me fazer para me animar.

À minha gata Nikita, por estar sempre ao meu lado durante as tardes de pesquisa e de escrita desta dissertação (literalmente), e por melhorar o meu dia instantaneamente.

E por fim, obrigada à minha ansiedade, por nunca me abandonar mesmo quando nada parecia estar a correr bem!

Abbreviations

A

A	Adenine
ADT	Androgen deprivation therapy
aLHRH	Luteinizing hormone-releasing hormone agonist
<i>APBB1IP</i>	Amyloid Beta Precursor Protein Binding Family B Member 1 Interacting Protein
AR	Androgen receptor
<i>ATM</i>	ATM Serine/Threonine Kinase
AUA	American Urology Association

B

<i>BARD1</i>	BRCA1 Associated RING Domain 1
BCR	Biochemical recurrence
<i>BCL2</i>	B -cell lymphoma 2
<i>BRCA1</i>	BRCA1 DNA Repair Associated
<i>BRCA2</i>	BRCA2 DNA Repair Associated
<i>BRIP1</i>	BRCA1 Interacting Protein C-Terminal Helicase 1

C

C	Cytosine
CAK	Cyclin-activated kinase complex
CAB	Complete androgen blockage
<i>CCND2</i>	Cyclin D2
<i>CCNH</i>	Cyclin H
CDK7	Cyclin Dependent Kinase 7
<i>CDK12</i>	Cyclin Dependent Kinase 12
<i>CEACAM21</i>	Carcinoembryonic Antigen Related Cell Adhesion Molecule 21
CI	Confidence interval
CHEK2	Checkpoint Kinase 2
CRPC	Castration-resistant prostate cancer
CZ	Central zone

D

<i>DAB2IP</i>	DAB2 Interacting Protein
DHT	Dihydrotestosterone
DNA	Deoxyribonucleic acid
DRE	Digital rectal exam

E

EAU	European Association of Urology
EBRT	External beam radiotherapy
EDTA	Ethylenediamine tetraacetic acid
ER	Estrogen receptor
ERG	ETS Transcription Factor ERG

F

<i>FAM83F</i>	Family With Sequence Similarity 83 Member F
<i>FAM84B</i>	Family With Sequence Similarity 84 Member B
<i>FGF10</i>	Fibroblast Growth Factor 10

<i>FGFR1</i>	Fibroblast Growth Factor Receptor 1
<i>FGFR2</i>	Fibroblast Growth Factor Receptor 2
<i>FOXO1</i>	Forkhead Box O1
<i>FSH</i>	Follicle-stimulating hormone
G	
<i>G</i>	Guanine
<i>GATM</i>	Glycine Amidinotransferase
<i>GnRH</i>	Gonadotropin-releasing hormone
<i>GWAS</i>	Genome-wide association study
H	
<i>HERC2</i>	HECT And RLD Domain Containing E3 Ubiquitin Protein Ligase 2
<i>HOXA2</i>	Homeobox A2
<i>HOXB13</i>	Homeobox B13
<i>HPC1</i>	Hereditary Prostate Cancer 1
<i>HR</i>	Hazard ratio
<i>HS3ST3A1</i>	Heparan Sulfate-Glucosamine 3-Sulfotransferase 3A1
J	
<i>JAZF1</i>	JAZF Zinc Finger 1
K	
<i>KLK3</i>	Kallikrein Related Peptidase 3
<i>KRT8</i>	Keratin 8
L	
<i>LD</i>	Linkage disequilibrium
<i>LH</i>	Luteinizing hormone
<i>LHRH</i>	Luteinizing hormone-releasing hormone
<i>LINC00598</i>	Long Intergenic Non-Protein Coding RNA 598
M	
<i>MAF</i>	Minor allele frequency
<i>MAT1</i>	RING Finger Protein MAT1
<i>MLH1</i>	MutL Homolog 1
<i>MLH3</i>	MutL Homolog 3
<i>MLPH</i>	Melanophilin
<i>MRI</i>	Magnetic resonance imaging
<i>MSH2</i>	MutS Homolog 2
<i>MSH6</i>	MutS Homolog 6
<i>MUTYH</i>	MutY DNA Glycosylase
<i>MYC</i>	MYC Proto-Oncogene, BHLH Transcription Factor
N	
<i>NAALADL2</i>	N-Acetylated Alpha-Linked Acidic Dipeptidase Like 2
<i>NER</i>	Nucleotide excision repair
<i>NR2C2</i>	Nuclear Receptor Subfamily 2 Group C Member 2
O	
<i>OR</i>	Odds ratio
<i>OS</i>	Overall survival

P

<i>P53</i>	Tumor Protein P53
<i>PALB2</i>	Partner And Localizer Of BRCA2
PCa	Prostate cancer
PCA3	Prostate Cancer Antigen 3
<i>PCAT19</i>	Prostate Cancer Associated Transcript 19
PCR	Polymerase chain reaction
PHI	Prostate health index
<i>PMS2</i>	PMS1 Homolog 2, Mismatch Repair System Component
PPC	Percentage of positive biopsy cores
PSA	Prostate specific antigen
PTEN	Phosphatase And Tensin Homolog
PZ	Peripheral zone

R

<i>RAP1</i>	Ras-related protein 1
<i>RASA1</i>	RAS P21 Protein Activator 1
RIAM	Rap1-GTP-interacting adapter molecule
RNA	Ribonucleic acid
RNF8	Ring Finger Protein 8
RP	Radical prostatectomy
RT	Radiotherapy
RUNX2	RUNX Family Transcription Factor 2

S

SNP	Single nucleotide polymorphism
-----	--------------------------------

T

T	Thymine
TF	Transcription factor
<i>TMPRSS2</i>	Transmembrane Serine Protease 2
<i>TNRC6B</i>	Trinucleotide Repeat Containing Adaptor 6B
TRUS	Transrectal ultrasound
TZ	Transition zone

U

UBC13	Ubiquitin-conjugating enzyme E2 13
-------	------------------------------------

W

<i>WDR11</i>	WD Repeat Domain 11
--------------	---------------------

Index

Abbreviations	- 1 -
Abstract	- 9 -
Resumo	- 13 -
1. Introduction.....	- 17 -
1.1. Prostate cancer	- 19 -
1.2. Advanced prostate cancer	- 22 -
1.3. Genetic variability in prostate cancer development	- 26 -
1.3.1. GWAS in Aggressive Prostate Cancer.....	- 30 -
1.3.2. The challenge of post-GWAS research.....	- 34 -
2. Objectives.....	- 37 -
2.1. Main Objective	- 39 -
2.2. Specific Objectives	- 39 -
3. Materials and Methods.....	- 41 -
3.1. Study population characterization	- 43 -
3.2. SNP selection	- 44 -
3.3. Laboratory procedures.....	- 46 -
3.4. Statistical analysis	- 47 -
4. Results	- 49 -
4.1. Descriptive statistics of the polymorphisms and the association with clinicopathologic characteristics	- 51 -
4.2. Overall survival analysis	- 52 -
4.3. Time to castration resistance analysis	- 55 -
5. Discussion	- 63 -
6. Conclusions and Future Perspectives.....	- 73 -
7. Appendix.....	- 77 -
Appendix 1	- 79 -
Appendix 2	- 85 -
8. References	- 87 -

Figure Index

Figure 1 - Estimated age-standardized prostate cancer global incidence rates in 2018.....	- 19 -
Figure 2 - Prostate cancer (PCa) natural history.	- 25 -
Figure 3 - Types of variants in cancer risk.....	- 28 -
Figure 4 - Clinical application of GWAS findings.	- 35 -
Figure 5 - Example of an allelic discrimination plot for rs2292884.....	- 47 -
Figure 6 - Overall survival by Kaplan-Meier and log-rank test according to <i>APBB1IP</i> rs1335540 polymorphism (dominant genetic model), stratifying by D'Amico classification. A) Subgroup of patients with non-high-risk D'Amico classification; B) Subgroup of patients with high-risk D'Amico classification.....	- 53 -
Figure 7 - Overall survival by Kaplan-Meier and log-rank test according to <i>MLPH</i> rs2292884 polymorphism (recessive genetic model), stratifying by D'Amico classification. A) Subgroup of patients with non-high-risk D'Amico classification; B) Subgroup of patients with high-risk D'Amico classification.....	- 54 -
Figure 8 - Overall survival by Kaplan-Meier and log-rank test according to <i>FGF10</i> rs2121875 polymorphism (dominant genetic model), stratifying by D'Amico classification. A) Subgroup of patients with non-high-risk D'Amico classification; B) Subgroup of patients with high-risk D'Amico classification.....	- 55 -
Figure 9 - Time to castration resistance by Kaplan-Meier and log-rank test according to <i>HERC2</i> rs6497287 polymorphism (dominant genetic model), stratified by PSA levels at diagnosis. A) Subgroup of patients with PSA <20ng/mL; B) Subgroup of patients with PSA ≥20ng/mL.....	- 56 -
Figure 10 - Time to castration resistance by Kaplan-Meier and log-rank test according to <i>FOXO1/LINC00598</i> rs7327461 polymorphism (recessive genetic model), stratifying for PSA levels at diagnosis; A) Subgroup of patients with PSA <20ng/mL; B) Subgroup of patients with PSA ≥20ng/mL.	- 57 -
Figure 11 - Time to castration resistance by Kaplan-Meier and log-rank test according to <i>RASA1/CCNH</i> rs35148638 polymorphism (recessive genetic model), stratified by PSA levels at diagnosis in a. A) Subgroup of patients with PSA <20ng/mL; B) Subgroup of patients with PSA ≥20ng/mL.	- 59 -
Figure 12 - Proposed model for the impact of the risk SNP panel in time to castration resistance.	- 71 -

Table Index

Table 1 - European Association of Urology risk groups for biochemical recurrence of localized and locally advanced prostate cancer [24].	- 23 -
Table 2 - Mechanisms of castration resistance prostate cancer (CRPC) development.....	- 26 -
Table 3 - Reported loci associated with prostate cancer aggressiveness through GWAS....	- 33 -
Table 4 - Clinicopathological characteristics of the study population.	- 44 -
Table 5 - Characteristics used in the selection of the study polymorphisms.	- 45 -
Table 6 - Context sequence of the study polymorphisms.	- 46 -
Table 7 - Descriptive statistics obtained for each polymorphism.	- 51 -
Table 8 - Multivariate Cox regression analysis for <i>HERC2</i> rs6497287 on the risk of earlier castration resistance in patients with PSA levels <20ng/mL at diagnosis, considering Gleason score and clinical stage.	- 57 -
Table 9 - Multivariate Cox regression analysis for <i>FOXO1/LINC00598</i> rs7327461 on the risk of earlier castration resistance in patients with PSA levels<20ng/mL at diagnosis, considering Gleason score and clinical stage.	- 58 -
Table 10 - Multivariate Cox regression analysis for <i>RASA1/CCNH</i> rs35148638 on the risk of earlier castration resistance in patients with PSA levels<20ng/mL at diagnosis, considering Gleason score and clinical stage.	- 59 -
Table 11 - Multivariate Cox regression analysis for combined <i>HERC2</i> rs6497287, <i>RASA1/CCNH</i> rs35148638 or <i>FOXO1/LINC00598</i> rs7327461 on the risk of earlier castration resistance in patients with PSA levels<20ng/mL at diagnosis, considering Gleason score and clinical stage.	- 60 -
Table 12 - Multivariate Cox regression analysis for combined <i>HERC2</i> rs6497287 and <i>FOXO1/LINC00598</i> rs7327461 or <i>HERC2</i> rs6497287 and <i>RASA1/CCNH</i> rs35148638 on the risk of earlier castration resistance in patients with PSA levels<20ng/mL at diagnosis, considering Gleason score and clinical stage.	- 61 -
Supplementary Table 1 - Overview of aggressive PCa susceptibility GWAS.....	- 79 -

Abstract

Prostate cancer (PCa) is the second most incident cancer in men worldwide. This high incidence is partially due to mainstream PSA screening, leading to an increase in the diagnosis of more initial and potentially indolent forms of PCa. Although the majority of cases are diagnosed at early stages, what allows monitoring the disease and delay treatment, the mortality associated to PCa is still elevated, mostly because of the poor prognosis of patients with advanced and castration resistance prostate cancer (CRPC). Genetic variants, such as single nucleotide polymorphisms (SNPs), may help elucidate the mechanisms leading to CRPC as well as predict those who are at higher risk to progress to CRPC.

Genome-wide approaches, like genome-wide association studies (GWAS), are a powerful tool that allows scanning the entire genome to search associations of SNPs with susceptibility or outcome of complex traits, such as PCa. To date, 8 GWAS on aggressive PCa have resulted in the report of 16 SNPs with ORs ranging from 0.94 to 2.90, and *P*-values ranging from 1.0×10^{-3} to 7.2×10^{-11} . However, the associations that result from GWAS need to be confirmed to meet the desired goal of clinical implementation of the variants. Thus, the purpose of this study was to select GWAS-discovered variants that have been associated with aggressive PCa and assess their impact on the outcome of a cohort of PCa patients from the North of Portugal.

All of the gathered variants were submitted to the GWAS4D software that elaborated a ranking according to a set of criteria, and variants *MLPH* rs2292884, *APBB1IP* rs1335540, *HERC2* rs6497287, *FOXO1/LINC00598* rs7327461, *RASA1/CCNH* rs35148638 and *FGF10* rs2121875 were selected. A retrospective hospital-based cohort study with 746 PCa patients was performed and the SNPs were genotyped by Real-Time Polymerase Chain Reaction (Real-Time PCR) using the TaqMan® Allelic discrimination Assay methodology. The two outcomes studied were overall survival (OS) and time to castration resistance.

Regarding OS, we observed that AA genotype carriers of *APBB1IP* rs1335540 and GG genotype carriers of *FOXO1/LINC00598* rs7327461 showed decreased survival when compared with the other genotypes in the subgroup of patients with PSA levels at diagnosis $<20\text{ng/mL}$ (log-rank test $P=0.002$ and $P=0.019$ respectively). A further stratification according to D'Amico classification revealed significant differences for three variants. *APBB1IP* rs1335540 AA homozygous and *MLPH* rs2292884 GG homozygous showed a decrease in OS in the non-high-risk group (log-rank test $P=0.020$ and $P=0.019$ respectively), while for *FGF10* rs2121875 G allele carriers an increase in survival was observed, also in the non-high-risk group (log-rank test $P=0.010$). All of these significant differences in survival were corroborated by Cox regression. We theorize that the possible

involvement of these variants in metastization and the AR pathway could be correlated with the OS in these patients.

The time to castration resistance analysis showed that *HERC2* rs6497287 allele C carriers, *FOXO1/LINC00598* rs7327461 GG and *RASA1/CCNH* rs35148638 CC homozygous led to a shorter time to castration resistance in the subgroup of patients with PSA levels <20ng/mL (log-rank test $P=0.009$, $P=0.001$ and $P=0.042$ respectively). A further multivariate Cox regression showed that when combined in pairs in a SNP risk panel, these variants helped predict individuals at higher risk for earlier castration resistance (HR=4.30, 95% CI=1.51–12.20, $P=0.006$) for individuals with PSA levels <20ng/mL, adjusting for Gleason score and clinical stage. We hypothesize that these variants are involved in DNA repair mechanisms and the AR pathway, which could contribute to progression to CRPC.

The complete understanding of the functionality of these variants and the biological mechanisms underlying the observed results would help implement them in the clinical practice, using them as biomarkers for prediction of outcome, evaluating response to androgen deprivation therapy (ADT). Moreover, they can potentially be of interest in the discovery of new therapeutical targets. To reach this goal, post-GWAS studies need to be performed, namely fine mapping and functional studies.

Keywords: Prostate cancer, aggressive prostate cancer, single nucleotide polymorphism, GWAS, clinical outcome, castration resistance, biomarker

Resumo

O cancro da próstata (CaP) é o segundo tipo de cancro mais incidente mundialmente nos homens. Esta elevada incidência é parcialmente devida ao uso convencional do rastreio através do antígeno específico da próstata (PSA), originando um aumento no diagnóstico de formas mais iniciais e potencialmente indolentes de CaP. Apesar de a maioria dos casos serem diagnosticados em estadios iniciais, permitindo a monitorização da doença e o adiamento de tratamento, a mortalidade associada ao CaP continua elevada, principalmente associada ao mau prognóstico dos doentes com CaP avançado e cancro da próstata resistente à castração (CRPC). Variantes genéticas, como os polimorfismos de nucleótido único (SNPs), podem ajudar a elucidar os mecanismos que levam ao CRPC e prever os indivíduos que têm maior risco de progredir para CRPC.

As abordagens *genome-wide*, como os *genome-wide association studies* (GWAS), são uma ferramenta que permite analisar todo o genoma na procura de associações de SNPs com a evolução clínica ou com a suscetibilidade para traços complexos, como o CaP. Até a data, foram realizados 8 GWAS em CaP agressivo que resultaram em 16 SNPs com *odds ratios* (OR) entre 0.94 e 2.90, e valores de P entre 1.0×10^{-3} e 7.2×10^{-11} . Contudo, as associações que advêm de GWAS têm de ser confirmadas de forma a poderem ser implementadas na prática clínica. Assim, o propósito deste estudo foi selecionar variantes reportadas por GWAS associadas a CaP agressivo e avaliar o seu impacto na evolução clínica de doentes com CaP do Norte de Portugal.

Todas as variantes recolhidas foram submetidas ao *software* GWAS4D que elaborou um *ranking* de acordo com um conjunto de critérios, e assim as variantes *MLPH* rs2292884, *APBB1IP* rs1335540, *HERC2* rs6497287, *FOXO1/LINC00598* rs7327461, *RASA1/CCNH* rs35148638 e *FGF10* rs2121875 foram selecionadas. Fizemos um estudo coorte retrospectivo de base hospitalar com 746 doentes com CaP, e os SNPs foram genotipados por reação de polimerase em cadeia (PCR) em tempo real usando a metodologia de discriminação alélica TaqMan®. Os dois parâmetros de avaliação da evolução clínica estudados foram a sobrevivência global e o tempo até ao desenvolvimento de resistência à castração.

Em relação à sobrevivência global, foi observado que os portadores do genótipo AA para *APBB1IP* rs1335540 e os portadores do genótipo GG para *FOXO1/LINC00598* rs7327461 mostraram uma menor sobrevivência global em comparação com os outros genótipos no subgrupo de doentes com níveis de PSA ao diagnóstico $<20\text{ng/mL}$ (teste *log-rank* $P=0.002$ e $P=0.019$ respetivamente). A estratificação de acordo com a classificação D'Amico revelou diferenças significativas em três variantes. Homozigóticos AA para *APBB1IP* rs1335540 e homozigóticos GG para *MLPH* rs2292884 exibiram uma menor sobrevivência global no subgrupo de não-alto-risco (teste *log-rank* $P=0.020$ e $P=0.019$ respetivamente), enquanto portadores do alelo G para *FGF10* rs2121875

tiveram um aumento na sobrevivência global, também no subgrupo de não-alto-risco ($P=0.010$). Todas as diferenças significativas foram corroboradas pela regressão de Cox. Foi colocada a hipótese de que o possível envolvimento destas variantes na metastização e na via do recetor de androgénio (RA) podem estar relacionados com a sobrevivência global nestes doentes.

A análise relativa ao tempo até ao desenvolvimento de resistência à castração mostrou que portadores do alelo C para *HERC2* rs6497287, portadores do genótipo GG para *FOXO1/LINC00598* rs7327461 e portadores do genótipo CC para *RASA1/CCNH* rs35148638 apresentaram um menor tempo até ao desenvolvimento de resistência à castração no subgrupo de doentes com níveis de PSA <20ng/mL (teste *log-rank* $P=0.009$, $P=0.001$ e $P=0.042$ respetivamente). A regressão multivariada de Cox mostrou que, num modelo combinado, estes SNPs ajudam a prever os indivíduos que apresentam maior risco de progressão precoce para CRPC (HR=4.30, 95% CI=1.51–12.20, $P=0.006$), também no subgrupo de doentes com níveis de PSA <20ng/mL, ajustando para *score* de Gleason e estadio clínico. Foi elaborada a hipótese de que estas variantes podem estar envolvidas em mecanismos de reparação de DNA e na via do RA, podendo assim contribuir para o desenvolvimento de CRPC.

O conhecimento da funcionalidade destas variantes e dos mecanismos biológicos subjacentes a estes resultados poderá contribuir para a sua implementação na prática clínica, como biomarcadores preditivos da evolução clínica, avaliando a resposta à terapia de bloqueio androgénico (TBA). Adicionalmente, as variantes podem ter interesse na descoberta de novos alvos terapêuticos. Desta forma, é necessária a realização de estudos pós-GWAS, nomeadamente *fine mapping* e estudos funcionais.

1. Introduction

1.1. Prostate cancer

Prostate cancer (PCa) is among the most common types of cancer, ranking as the third most incident worldwide, after lung and female breast cancer, and the second most diagnosed in men, having been estimated approximately 1.3 million new cases in 2018 [1]. Although there is a large variation in PCa incidence across the world, around 70% of PCa cases occur in developed countries [2]. The age-standardized incidence rates are higher in Australia, New Zealand, United States of America, and in Western and Northern Europe, but remain low in Asian populations (Figure 1) [1]. In Europe, PCa incidence varies 5-fold, ranging from 37 new cases per 100,000 inhabitants observed in Albania, to 189 per 100,000 inhabitants in Ireland. Regarding Portugal, the most recent data show an incidence of 87.7 cases per 100,000 inhabitants in 2018, making it the most incident solid cancer in males [3]. Mortality rates for PCa have been described to vary considerably worldwide, however, less accentuated than incidence rates [4]. In 2018, it has been estimated 359 thousand deaths worldwide, with the majority occurring in less developed countries [1].

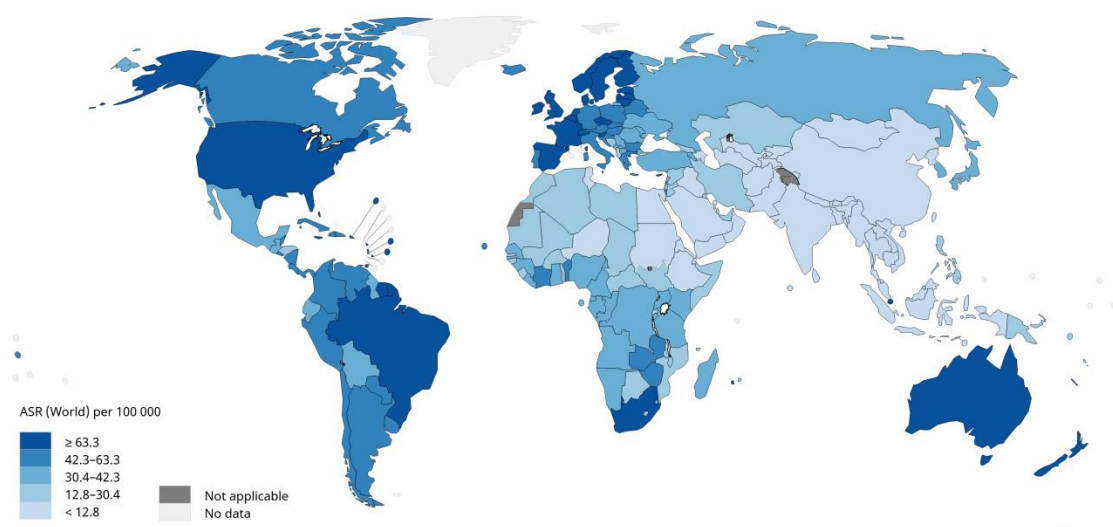


Figure 1 - Estimated age-standardized prostate cancer global incidence rates in 2018. Data source: GLOBOCAN 2018, Graph production: IARC (<http://gco.iarc.fr/today>) World Health Organization.

For a disease as common as PCa, little is known of its etiology, with age, ethnicity and family history being the better-established risk factors. In fact, individuals with first-degree relatives affected by the disease have a two-fold increased risk of developing PCa than those among the general population. Additionally, and considering ethnicity in the equation, a higher incidence of PCa is observed in people of African-American ancestry,

suggesting the role of genetic is determinant in the modulation of individual prostate tumorigenesis risk [5].

Lately, there has been an increase in the incidence of PCa, mostly attributed to population aging, which goes along with the presuppose that incidence is higher in more developed countries, but also to the improvement of the diagnostic techniques and the common use of prostate-specific antigen (PSA) measurements [6, 7]. PSA is a glycoprotein serine protease produced by prostatic epithelial cells whose job is to liquefy serum, help sperm motility and dissolve cervical mucus [8]. It was discovered that PSA was elevated in a percentage of PCa cases (serum PSA cutoff of 4 ng/mL) and so it began being widely used as a diagnostic tool which lead to an increase in the diagnosis of more early stages of PCa [9]. But not only did the use of PSA in screening lead to overdiagnosis and consequent overtreatment of indolent forms of PCa, that would, if left undiagnosed and untreated, not cause any harm to the patient, but also it showed lack of specificity for PCa, as higher levels of PSA can also be reported in cases of benign prostatic hyperplasia, prostatitis or in cases of recent manipulation of the prostate, and in individuals of older age [10]. Combining this to the fact that the PSA test does not discriminate between clinically aggressive and low-risk tumors, implies that this test alone is not sufficient to diagnose PCa and must be complementary to other diagnostic tools like digital rectal exam (DRE), transrectal ultrasound (TRUS) and the histopathological evaluation of biopsy or tumor resection specimens [11].

Currently, aside from serum PSA, there are two other FDA approved PCa biomarkers, Prostate Health Index (PHI) and PCA3 [12]. PHI combines three forms of PSA, total PSA, free PSA and p2PSA into a mathematical formula: $([-2]proPSA/free\ PSA) \times \sqrt{PSA}$ [13]. Unlike serum PSA, free PSA is unbound from proteins, circulating free in the bloodstream, and presents a better specificity to PCa than serum PSA [14]. Precursor forms of PSA are a component of free PSA. The [-2]proPSA has been proved to have better specificity for PCa than free PSA and total PSA, as this precursor form of PSA is enriched in PCa tissue when compared to benign prostate tissue, as well as being significantly related to the risk of aggressive PCa [15]. PCA3 is a long non-coding RNA that is highly overexpressed in PCa tissue, but with significantly lower levels in surrounding tissues and non-detectable levels in normal prostate tissue, making it an ideal biomarker for detection of PCa [16].

Anatomically the prostate can be divided into 3 sections, the central zone (CZ), the peripheral zone (PZ) and the transition zone (TZ) [17]. The most common zones for the development of PCa are the PZ, holding up to 70% of the cases, and the TZ close to the urethra, with around 20-25% of all PCa [18]. Histologically, there is not much variety in classification, as close to 95% of all cases are carcinomas [19]. The vast majority of PCa

cases are acinar adenocarcinomas, which are the biggest portion of the carcinomas, arising in prostate gland cells. Other than the acinar adenocarcinomas, other variants have been described, either arising in glands, such as ductal adenocarcinoma and signet ring adenocarcinoma, or not, like sarcomatoid carcinoma, urothelial carcinoma, basal cell carcinoma, etc. [19, 20]. The distinction between histological subtypes is important in a prognosis and treatment point of view.

Nevertheless, this neoplasia can present a multifocal occurrence, existing the possibility that there may be more than one primary tumor in different phases of the disease. So the Gleason grading system continues to be the standard approach to histologically grade prostate adenocarcinoma and a major predictor of outcome [20]. Resourcing to the multiple biopsy samples or tissue samples from radical prostatectomy (RP), the Gleason score is based on the extent of glandular differentiation and pattern of growth of the tumor in the prostatic stroma. A Gleason grade from 1 to 5, with 1 being the better differentiated and 5 the least differentiated, is attributed to the most prevalent histological pattern and the second most common pattern. These two grades are summed to provide the Gleason score that varies from 2 to 10 [11, 17]. Tumors bearing Gleason scores ≤ 6 are characterized as being differentiated to moderately differentiated, while tumors with Gleason scores ≥ 8 are less differentiated and present an aggressive nature with frequent metastization. Gleason score =7 encompass tumors with intermediate differentiation and aggressiveness, that can present a more favorable prognosis if the most extensive histological pattern is 3 (3+4) or a worst prognosis if the more common pattern is 4 (4+3) [21]. More recently, to overcome some inconsistencies associated with the Gleason score, it was implemented a new grading system based on it, dividing patients into 5 grade groups that provide a better prognostic stratification: Grade group 1 (Gleason score ≤ 6), Grade group 2 (Gleason score 3+4=7), Grade group 3 (Gleason score 4+3=7), Grade group 4 (Gleason score 8) and Grade group 5 (Gleason score 9 and 10) [22, 23].

Aside from the Gleason score, the clinical stage is another major indicator of prognosis and provides a better orientation for a treatment course. The most commonly used clinical staging system is the TNM system that classifies patients' tumors based on the size and extension of the primary tumor, the involvement of the lymph nodes and the presence of metastasis [24, 25].

In the last few decades, there has been continuing progress in PCa treatment in order to improve the outcome of patients. As so, the choice regarding treatment options entails a consideration of the patients' individual and clinical factors, such as age, clinical tumor stage, serum PSA levels, Gleason score or the number of positive prostate biopsies [26]. For localized, low-risk disease (T1-2aN0M0), the current challenge is to avoid the

overtreatment of clinically irrelevant tumors, which, usually, confers potentially harmful comorbidities to these patients [27]. More conservative strategies for tumors at this stage include watchful waiting and active surveillance [24]. Watchful waiting is limited to patients with low life expectancy and intends to minimize their treatment-related morbidity, having then a palliative purpose. Active surveillance, however, is more suitable for patients with early-stage PCa, which have high life expectancy, and consists in waiting for an evolution on the disease to initiate a curative treatment. Patients are monitored regarding their PSA levels, alongside with DRE, additional prostate biopsies and, eventually, magnetic resonance imaging (MRI) [24]. In addition to the more conservative treatment strategies, there are more active options for patients with localized disease. RP is the most effective method to completely eradicate prostate-confined PCa, by surgically removing the prostate *en bloc* with the seminal vesicles, and the bilateral pelvic lymph nodes dissection [12]. Also, external-beam radiotherapy (EBRT) and brachytherapy, which is a form of internal radiotherapy (RT) where the radiation is administered through seeds that are implanted temporarily or permanently in the patients' prostate, are a standard treatment for localized disease [28]. Patients with more advanced and even metastatic forms of PCa are usually submitted to androgen deprivation therapy (ADT), a form of hormonotherapy whose objective is to block the production or the action of testosterone in the proliferation of cancer cells [11, 29]. In more advanced cases, besides ADT, RP and RT, patients can also be treated with courses of chemotherapy [11].

1.2. Advanced prostate cancer

Even though the majority of PCa cases are diagnosed at early stages of the disease, 10-20% of patients already present advanced metastatic disease at diagnosis [30]. Additionally, recurrence is a very common feature observed in this group of patients, after being submitted to the first line of treatment, which prompts the selection of alternative treatment schemes [11].

Many of the patients who die of PCa present prostate-confined disease at diagnosis. As so, it is important to assess the risk of death at diagnosis to differentiate low-risk from high-risk PCa, without relying solely on the TNM classification [31]. Although there are many different risk assessment schemes for PCa, that determine the risk according to different characteristics, as the Gleason score, clinical stage, serum PSA level and disease burden, the D'Amico Classification is the most used one. The American Urology Association (AUA) and the European Association of Urology (EAU) adopted the D'Amico and collaborators classification of risk of biochemical recurrence (BCR) after

surgery or EBRT, in which high-risk PCa was characterized as cases presenting PSA level >20 ng/mL, a Gleason score of 8 or higher, or clinical stage $\geq$ T2c (Table 1) [32]. Additionally, Huang *et al.* concluded that the percentage of positive biopsy cores (PPC) would also be a powerful predictor of outcome after RT and thus devised a new risk stratification, the PPC classification, defining high-risk PCa as those tumors with Gleason score $\geq$ 8 and/or PSA level >20 ng/mL plus PPC >50% [33].

Table 1 - European Association of Urology risk groups for biochemical recurrence of localized and locally advanced prostate cancer [24].

Low-Risk	Intermediate-Risk	High-Risk	
PSA <10ng/mL and Gleason score <7 and cT1-T2a	PSA 10-20 ng/mL or Gleason score 7 or cT2b	PSA >20ng/mL or Gleason score >7 or cT2c	Any PSA any Gleason score cT3-4 or cN+
Localized			Locally Advanced

As previously mentioned, patients with advanced PCa are commonly treated with ADT. This therapy is based on the premise that androgens are fundamental for the normal prostate epithelial cell growth and proliferation, and that their reduction leads to cell atrophy [12]. The main male androgen is testosterone, mostly produced by Leydig cells in the testicles, with a smaller contribution of the adrenal complex [34]. Its levels are maintained through a negative feedback mechanism between the hypothalamus–pituitary–gonadal axis. When testosterone levels are low, the hypothalamus releases gonadotropin-releasing hormone (GnRH) stimulating the release of luteinizing hormone (LH) and follicle-stimulating hormone (FSH) by the pituitary gland. In turn, these two latter hormones act on the Leydig cells to stimulate the production of testosterone [35]. Testosterone action on the prostate depends on its reduction by the 5 α -reductase enzyme to dihydrotestosterone (DHT), an active metabolite that binds to the androgen receptor (AR) with a higher affinity than free testosterone [36]. The AR is activated by high-affinity binding of testosterone or the more potent androgen DHT, leading to a cascade of events that culminate in the transcription of genes involved in cell growth and survival [37].

It is established that PCa growth is dependent, at least initially, on the relation between epithelial cells and androgens [12]. Therefore, it is to expect that when the circulating levels of androgens are drastically reduced, cell growth and proliferation are also reduced and cell apoptosis increases, resulting in the control of tumor growth [38].

Androgen deprivation can be achieved either surgically, through orchidectomy, by removing the major source of testosterone, or chemically, by suppressing the production of testosterone by the testicles or by suppressing the action of circulating testosterone in their receptors. There are several classes of androgen blockers: oestrogens, anti-androgens, luteinizing hormone-releasing hormone agonists (aLHRH) and LHRH antagonists [12]. The most commonly used classes are aLHRH, whose long-term exposure inhibits the LHRH receptors through down-regulation, and thus suppressing the secretion of LH and FSH and consequently the production of testosterone; and anti-androgens which compete with androgens at the receptor level. All of the classes of ADT can be used alone or in combination. Often these are used in combination to achieve complete androgen blockade (CAB) [24]. ADT can be used in association with RT for better results in intermediate and high-risk patients, and can also be applied as adjuvant therapy after localized treatment [39]. Moreover, ADT can be combined with chemotherapy in the treatment of metastatic PCa [40, 41]. Nonetheless, the use of this therapeutic option has several adverse effects, as a result of the depleted levels of androgens, ranging from erectile dysfunction, gynecomastia, hot flashes or osteoporosis, to more severe ones, as cardio- or hepatotoxicity [42]. Even though some of these side effects may not have a significant impact on cancer-related survival, they are a very important source of morbidity, especially when considering the usually advanced age of patients [39]. However, ADT is only effective for a short period of time. The majority of patients will recur in 18-24 months, evolving to a state of castration-resistant prostate cancer (CRPC), where tumor cells regain proliferative capacity despite the low levels of androgens (Figure 2) [43]. CRPC is characterized by castrate serum testosterone levels <50 ng/dL (or 1.7 nmol/L) combined with either evidence of biochemical (PSA) or radiological progression [24]. Patients with CRPC present a poor prognosis, as the estimated median overall survival (OS) stands between 12 and 35 months [44].

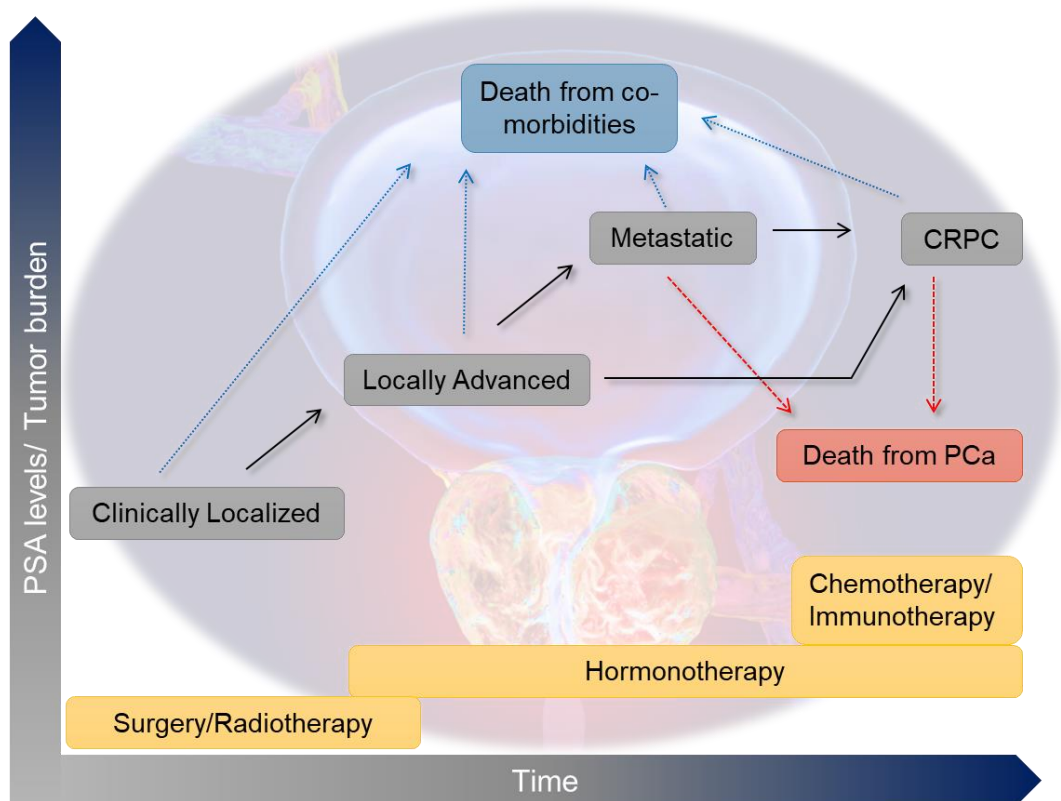


Figure 2 - Prostate cancer (PCa) natural history. In black shows the natural course of PCa, spanning around 10-15 years, and in yellow the more common treatments associated with each stage of the disease. PCa begins with a localized tumor in the prostate gland that can be treated with radical prostatectomy (RP) or radiotherapy (RT). When cases are only detected in advanced or metastatic stages or patients experience progression despite initial therapy, which is accompanied by rise in PSA and appearance of symptoms, PCa is commonly treated with hormonotherapy. However, hormonotherapy is only effective for 18-24 months before evolving to a castration-resistance state. While the majority of men diagnosed with PCa do not die from the disease, in the latter stages the median survival is very low.

The mechanisms that lead to CRPC are not completely understood, however, some theories have been proposed on how PCa cells can grow and proliferate even in the absence of androgens, as addressed in Table 2.

Table 2 - Mechanisms of castration resistance prostate cancer (CRPC) development.

	Pathway	Mechanism
Androgen receptor-dependent mechanisms	Hypersensitive Pathway	More production of AR, either by gene amplification and AR protein overexpression or by mutations; Increased sensitivity of AR to compensate for the lower levels of androgen; Higher rates of testosterone 5 α -reduction to dihydrotestosterone.
	Promiscuous Pathway	Wider specificity of the AR to other ligands other than testosterone, like progesterone, hydrocortisone, estradiol, and certain AR antagonists (acting as agonists) that can activate the receptor; Altered expression and function of AR co-regulators.
Androgen receptor-independent mechanisms	Outlaw Pathway	AR is activated without any ligand by phosphorylation from non-androgen sources such as Akt or the MAPK pathway.
	Bypass Pathway	Alternative survival pathways that annul the need for the AR or AR ligands: - Overexpression of <i>BCL2</i> - Activation of oncogenes - Inactivation of tumor suppressor genes
	Lurker Cell Pathway	Selection of androgen-independent tumor cells (possibly stem cells) through ADT.

AR – androgen receptor; *BCL2* – B-cell lymphoma 2; ADT – androgen deprivation therapy

Information gathered from: [37, 45, 46]

There is no definitive treatment for CRPC due to the high potential for progression and metastization; as so, any action is considered palliative, with special attention for the treatment of bone pain [24]. Treatment for CRPC includes chemotherapeutic agents, such as docetaxel and cabazitaxel, new generation ADT agents abiraterone acetate and enzalutamide, immunotherapeutic agent sipuleucel-T and radium-223, a bone-targeted radiopharmaceutical [12, 24].

1.3. Genetic variability in prostate cancer development

The fact that family history and the ethnic background are risk factors for PCa suggests a genetic component to the development of this disease [5]. In fact, the strong epidemiological evidence that arises from familial and twin studies have provided a better understanding of the genetic basis of PCa predisposition [47]. For example, *BRCA1* and *BRCA2*, genes widely known for their implication in the hereditary breast and ovarian

cancer syndrome, have also been implicated in hereditary PCa [48, 49]. Moreover, genetic alterations in *BRCA2* have additionally been reported to be associated with PCa aggressiveness and early-onset [50, 51].

Linkage studies have confirmed the association of the *HPC1* gene with PCa, and specifically connected it to early-onset hereditary PCa [52]. Also, genetic linkage studies have clearly reported the implication of the highly penetrant *HOXB13* gene on the predisposition to PCa, which have been shown to increase the risk of early-onset disease [49, 53, 54]. Additional genes such as *ATM*, *BRIP1*, *CHEK2*, *MUTYH*, *PALB2*, *PMS2*, mismatch repair genes (*MLH1*, *MLH3*, *MSH2*, *MSH6*) and the fusion gene *TMPRSS2-ERG* have also been implicated in PCa risk, although the replication of these findings has not been consistent [55-60]. Moreover, germline and somatic mutations in the DNA repair genes *BRCA1* and *BRCA2*, *PALB2*, *ATM*, *CDK12* or *P53* are also found in approximately 12% of CRPC patients [51, 61, 62]. Nevertheless, these rare susceptibility genes account only for a fraction of PCa heritability [63]. It is clear now that these highly penetrant mutations occur way too rarely in order to be accountable for the entire genetic background of PCa. As so, a polygenic model is likely to better elucidate the mechanisms underlying prostate tumorigenesis, where, alongside the less common but high penetrance susceptibility genes, smaller effect size variants may also impact disease susceptibility and/or outcome of PCa patients [64, 65].

One of the common but low penetrance variants is polymorphisms. As one of the most common genetic alterations presented in human DNA, polymorphisms correspond, by definition, to sequence variations where the minor allele is present in, at least, 1% of the population [66]. The simplest and most abundant form of these variations, single nucleotide polymorphism (SNP), leads to a substitution of only one nucleotide in the DNA sequence [67]. As they are scattered throughout the genome, the impact of this type of alteration on the functionality of a specific gene or protein depends on their location, ranging from potentially null effect variants, to gene expression or protein alteration-inducing variants [67].

Over the years, there has been intense debate regarding the type of variants that contribute to individual susceptibility to complex diseases [68]. In accordance, linkage analyses have been a suitable technique for the discovery of highly penetrant variants implied in rare Mendelian disorders, although this type of analysis has only limited success for the mapping of common, multigenic diseases [69]. In fact, the common disease/common variant hypothesis states that common diseases are most likely caused by genetic variations which are also common in the general population and their cumulative effects (Figure 3). Therefore, those alterations have small penetrance and the total risk due to common genetic variation must be spread across multiple determinants

[70]. Thus, the limited impact of linkage analysis in the study of the genetic background of common diseases was eventually overcome by the development of population-based analysis, as well as genome research techniques, particularly genome-wide association studies (GWAS) [71].

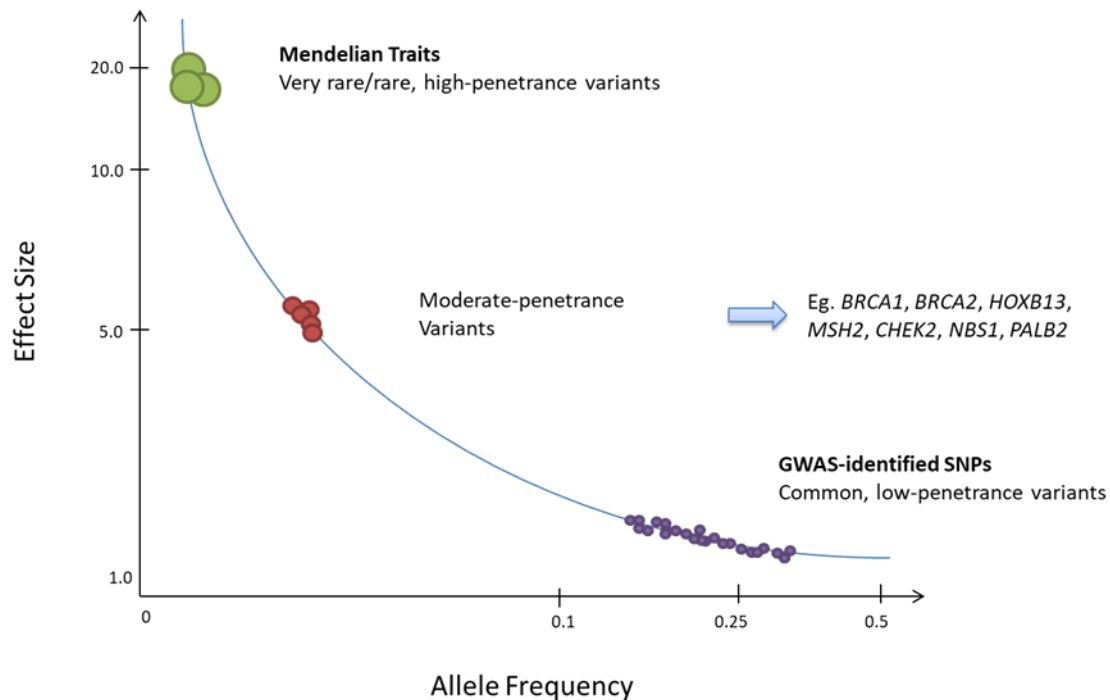


Figure 3 - Types of variants in cancer risk. Mendelian traits are influenced by variants that, despite being extremely rare, have a high penetrance and thus present large effect sizes. In contrast, variants identified through Genome-Wide Association Study (GWAS) have a much lower penetrance but are much more common in the general population. Variants like *BRCA1*, *BRCA2*, *HOXB13*, *MSH2*, *CHEK2*, *NBS1*, *PALB2*, that affect PCa risk lie as intermediate determinants, with a moderate penetrance and a common frequency in population. Thus, a polygenic model, encompassing rare and common variants, may likely explain PCa susceptibility. GWAS: Genome-wide association study; SNPs: Single-nucleotide polymorphisms. Adapted from: [72, 73].

The development of GWAS has revolutionized genetics and genomics research. Through an agnostic approach, this type of analysis allows the survey of DNA variations, specifically SNPs, throughout the human genome in order to identify genetic factors which may account for common traits' etiology [70]. To achieve this, a large number of genetic markers are tested simultaneously for the association with specific traits, in hundreds or thousands of samples [74]. Because of their whole-genome approach, GWAS' findings might unravel putatively associated markers within or around regions which were not previously known to have an impact on a particular trait, as well as the association of loci with several different traits not previously related, or the identification of markers in gene-desert regions [75].

It is known that a substantial part of the human genome is organized in haplotype blocks [76]. Thus, GWAS are based upon the principle of linkage disequilibrium (LD), which explains the non-random association between alleles at different loci [32, 33]. LD allows the selection of a specific set of SNPs, known as tagSNPs, which are representative of the majority of the markers presented in each block. Consequently, the use of tagSNPs may reduce the number of genetic variants that need to be genotyped in order to obtain the extended genetic spectrum of a particular region [69, 77].

Despite the innovative GWAS methodology, the identification of relevant disease-associated markers has been facing some drawbacks, partially because each causal variant only accounts for a small contribution to the overall heritability of a trait [69]. In fact, this feature, known as missing heritability, might be a reflection of the relatively small effect sizes inherent to GWAS' findings, with a median reported odds ratio (OR) of 1.33 [75]. Possible explanations for this feature may rely on (1) the existence of rare variants with larger effects on the phenotype that might not be detected by the available genotyping arrays, (2) common variants with smaller effects that have not been identified yet, and (3) structural variants, including copy number variations, translocations and inversions, that are not captured by the existent genotyping platforms [78]. In addition, other important features that are possibly being overlooked are the gene-gene and gene-environment interactions, which are frequently unaccounted for in family studies, resulting in an overestimation of the heritability. Furthermore, epigenetic factors, which can also be inherited, may also account for the phenomenon of missing heritability [79, 80].

In order to compensate for some of the limitations and drawbacks inherent to the GWAS' study design, some strategies have been implemented. One of those strategies is the use of the Bonferroni correction that intends to minimize the number of false-positive associations resulting from multiple testing, by establishing a genome-wide threshold at 5×10^{-8} [69, 81, 82]. However, the use of this strict threshold reduces the chances of finding positive associations in SNPs with lower effect sizes, unless the number of samples used is relevant enough to achieve adequate statistical power [83, 84]. Hence, it is possible to recruit patients from several international consortia, like PRACTICAL (Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome) and ICPCG (International Consortium for Prostate Cancer Genetics), to overcome sample size limitations and combine statistical power [73, 85, 86]. On the other hand, the population samples available in this consortia are very heterogeneous, as they pool together cohorts of patients from various ancestries [85]. Furthermore, in the presence of a genetic association study with low statistical power, upward bias in the estimated effect size is observed, meaning the reported results are frequently overestimated, in a phenomenon known as "winner's curse" [87]. Consequently, the need to confirm all GWAS findings is

mandatory, to eliminate false-positive findings and test the generalization of confirmed results through the accomplishment of studies in populations of different ancestries [88]. Replications and validation studies are a very important strategy to assess the credibility of these results [89].

1.3.1. GWAS in Aggressive Prostate Cancer

As previously mentioned, according to the common disease/common variant hypothesis, more common and complex diseases are more likely caused by more common variants [68]. Being cancer a common and complex disease, it is expected that more common but less penetrant variants, like SNPs, can be at least partly responsible for PCa susceptibility [72]. Therefore, GWAS plays an important part in uncovering the missing heritability of PCa [90].

In 2007, the first cancer GWAS was conducted by Gudmundsson and collaborators, who reported a new genetic variant at 8q24 locus, rs16901979, associated with PCa susceptibility [91]. This variant is located in a region with low gene content, with the nearest genes being *MYC* and *FAM84B*. However, no conclusions were drawn regarding the functional impact of rs16901979, as no significant differences were observed regarding *MYC* and *FAM84B* expression, both in normal and tumor prostate tissue [91]. Simultaneously, two other studies, one a PCa GWAS and the other a fine-mapping study conducted on the 8q24 locus, have identified additional variants associated with PCa in this region, establishing it as a putative important determinant of prostate carcinogenesis [92, 93]. Subsequently, 45 more PCa GWAS have been conducted, mostly in European populations, with approximately 170 susceptibility variants reported so far [71, 94].

To this date, no GWAS have been conducted in CRPC. Though, Gallagher and collaborators conducted a validation study on the clinical outcome of Ashkenazi Jewish-ancestry men, studying, among others, the association of previously identified PCa susceptibility markers with this phenotype [95]. They reported associations for rs4962416 (Hazard Ratio (HR) =0.69, 95% Confidence Interval (CI) =0.53–0.91), rs10486567 (HR=1.43, 95%CI=1.12–1.83), and rs6465657 (HR=1.35, 95%CI=1.07–1.70) with metastatic CRPC, at $P<0.05$. However, the authors stated that these results were not conclusive since they did not perform any correction for multiple testing, intending them to be “hypothesis-generating”. The genetic variant rs10486567 presented the most significant association with castration-resistant metastasis ($P=5.0\times 10^{-3}$) and was also strongly associated with BCR. This SNP is located within the *JAZF1* gene, which encodes

a three C2-H2-type zinc finger protein and acts as a transcriptional repressor of *NR2C2*, a nuclear receptor expressed in prostate tissue [96]. It has recently been established that *JAZF1* contributes to PCa progression, promoting cell proliferation and invasion, by up-regulating JNK/Slug expression, and this can potentially be used as a therapeutic target for both primary and CRPC [97]. In fact, Slug has been shown to inhibit E-cadherin expression, responsible for cell-cell adhesion, in prostate carcinogenesis [98].

Despite the lack of GWAS on CRPC, there have been a few studies conducted with the intent of discovering variants that could act as predictors of aggressive disease [99-106]. In 2007, Duggan and collaborators were the first to conduct a GWAS study for aggressive PCa [99]. In the discovery phase, combined data from two aggressive PCa GWAS, both with subjects of European ancestry, showed that 81 SNPs were statistically significantly associated with this phenotype ($P < 0.05$), with 11 markers reaching a P -value < 0.01 and 7 of which presenting consistent results in both studies [92, 107]. The authors further attempted to validate these 7 SNPs in two different populations, one of European and another of African-American ancestry. Only rs1571801 has maintained a significant association after the validation stage ($OR = 1.36$, $95\%CI = 1.13-1.63$; $P = 1.0 \times 10^{-3}$), when considering both European and African-American cases. Interestingly, the association of rs1571801 was stronger in patients with a Gleason Score > 8 , compared to those with a Gleason Score of 7 or 8 ($P = 0.001$ vs $P = 0.028$) in the population of European ancestry. This SNP is an intronic variant located in *DAB2IP*, a PCa tumor suppressor gene whose expression is usually reduced in PCa cells [108]. Since the accomplishment of this study, 7 additional GWAS on aggressive PCa susceptibility were performed [100-106]. In one of those, an association was reported for the variant rs4054823, at 17p12, when considering the combined results of the discovery stage and three different validation stages (Combined $OR = 1.13$, $95\%CI = 1.08-1.19$; $P = 2.1 \times 10^{-6}$) [101]. In contrast with the majority of GWAS, which presented a case-control design, this study used a case-case design, enrolling cases of aggressive and indolent PCa, in order to identify genetic markers that could act as predictors of the former. Significant differences between the two groups of PCa patients ($P_{het} = 2.1 \times 10^{-8}$) were also reported, suggesting rs4054823 may be differentially associated with disease aggressiveness. Moreover, the frequency of the minor allele T was reported to be increased in patients with poorly differentiated disease, which confers a higher chance of aggressive PCa development for homozygous individuals ($OR = 1.26$, $95\%CI = 1.16-1.36$; $P = 2.1 \times 10^{-8}$) [101]. This variant is located in an intergenic region, with the closest gene being *HS3ST3A1*, which encodes a heparan sulfate biosynthetic enzyme, a protein with no known relation to PCa [101].

In an effort to increase the study power and, hence, detect new susceptibility loci, Amin Al Olama *et al.* performed a meta-analysis study [105]. Through the inclusion of 4

GWAS, the authors were able to gather 11085 cases, 5953 of which with aggressive disease, and 11463 healthy controls, mostly of European ancestry. After analysing approximately 2.6 million SNPs in the discovery stage, rs11672691 was found to be significantly associated with aggressive PCa (OR=1.12, 95% CI=1.03–1.21; $P_{\text{combined}}=1.4 \times 10^{-8}$), although, when testing for overall PCa, the association was even stronger (OR=1.08, 95% CI=1.05–1.12; $P_{\text{comb}}=2.2 \times 10^{-12}$). In the replication stage, despite the effect size being higher in aggressive cases in relation to indolent ones (OR=1.12, 95% CI=1.03–1.21; $P=0.006$ vs OR=1.08, 95% CI=1.05–1.12; $P=8.2 \times 10^{-7}$), this difference was not statistically significant ($P_{\text{heterogeneity}}=0.18$). This intronic variant, located on chromosome 19q13, lies within *PCAT19*, a long non-coding RNA gene. Additionally, rs11672691 was recently shown to be linked with CRPC development, as well as some prognostic features, such as advanced tumor stage and PSA progression. This SNP may be also involved in the regulation of *HOXA2*, *PCAT19* and *CEACAM21* expression, playing a role in PCa cell growth, invasion, metastasis, and disease progression [109]. In line with the previous variant, a further identified SNP, rs11704416, was also associated with the studied phenotype ($P_{\text{comb}}=4.0 \times 10^{-6}$), presenting a stronger association with overall PCa as well ($P_{\text{comb}}=3.7 \times 10^{-7}$). This SNP is located in the *FAM83F* gene, a member of the FAM83 family that possesses oncogenic properties, but with no known relation to PCa [110, 111].

Altogether, aggressive PCa GWAS have reported 16 markers associated with this phenotype so far (ORs ranging from 0.94 to 2.90, and P -values ranging from 1.0×10^{-3} to 7.2×10^{-11}), with 10 variants reaching the genome-wide significance threshold (Table 3). Despite the small number of SNPs reported, the discovery stage of these studies have showed nominally significant associations ($P < 0.05$) for a much larger number of polymorphisms (Supplementary Table 1), that were ultimately disregarded in an effort to limit the number of findings that would be followed in further replication/validation stages.

Table 3 - Reported loci associated with prostate cancer aggressiveness through GWAS.

GWAS	SNP	Chr	Gene	OR	95%CI	P-value
Duggan <i>et al.</i> 2007	rs1571801	9q33.2	<i>DAB2IP</i>	1.36	1.13–1.63	1.00x10 ⁻³
Sun <i>et al.</i> 2009	rs9623117	22q13.1	<i>TNRC6B</i>	1.18	1.11–1.26	4.96x10 ⁻⁷
Xu <i>et al.</i> 2010	rs4054823	17p12	<i>HS3ST3A1</i> (nearest gene)	1.13	1.08–1.19	2.10x10 ⁻⁶
FitzGerald <i>et al.</i> 2011	rs6497287	15q13.1	<i>HERC2</i>	1.46	1.10–1.94	4.00x10 ⁻³
Nam <i>et al.</i> 2011	rs11199874	10q26.12	<i>WDR11, FGFR2</i> (nearest genes)	2.90	2.10–4.10	7.20x10 ⁻¹¹
	rs10749408	10q26.12	<i>WDR11, FGFR2</i> (nearest genes)	2.50	1.90–3.30	3.20x10 ⁻¹¹
	rs10788165	10q26.12	<i>WDR11, FGFR2</i> (nearest genes)	2.90	2.10–4.10	1.30x10 ⁻¹⁰
	rs4775302	15q21.1	<i>GATM</i> (nearest gene)	2.60	1.30–3.70	3.10x10 ⁻⁸
	rs1994198	15q21.1	<i>GATM</i> (nearest gene)	2.60	1.30–3.70	8.20x10 ⁻⁹
Schumacher <i>et al.</i> 2011	rs2292884	2q37.3	<i>MLPH</i>	1.14	1.07–1.20	7.80x10 ⁻⁶
	rs902774	12q13.13	<i>KRT8</i> (nearest gene)	1.21	1.13–1.29	1.40x10 ⁻⁸
Olama <i>et al.</i> 2012	rs11672691	19q13.2	<i>PCAT19</i>	1.12	1.03–1.21	1.40x10 ⁻⁸
	rs11704416	22q13.1	<i>FAM83F</i>	0.94	0.86–1.02	4.00x10 ⁻⁶
Berndt <i>et al.</i> 2015	rs35148638	5q14.3	<i>RASA1, CCNH</i> (nearest genes)	1.09	1.06–1.12	6.49x10 ⁻⁹
	rs78943174	3q26.31	<i>NAALADL2</i> (nearest genes)	1.34	1.20–1.48	4.18x10 ⁻⁸
	rs62113212	19q13.33	<i>KLK3</i> (nearest genes)	1.15	1.10–1.20	5.85x10 ⁻⁹

GWAS: Genome Wide Association Study; SNP: Single Nucleotide Polymorphism; OR: Odds Ratio; 95%CI: Confidence Interval at 95%.

Data taken from [99-106]

1.3.2. The challenge of post-GWAS research

GWAS has emerged as an essential tool in the search for associations between common variants and complex diseases, as the case of PCa. This approach may provide insight into new genetic information on aggressive disease, which can be potentially used in the clinical context [74]. So far, the obtained results for aggressive PCa GWAS have been inconclusive, due not only to the limitations associated with the GWAS' design but also to other reasons related to each particular study. Firstly, the small sample sizes used in these studies are a major limitation for achieving the suitable statistical power, which in combination with neglect on non-European cases hamper the translation of these findings to other populations [100, 102-104]. Thus, most authors struggle to reach enough statistical power to detect variants with smaller effect sizes or that are only marginally associated with the phenotype, which might also partially explain the phenomenon of missing heritability in PCa [78, 102]. Furthermore, the heterogeneity observed in the definition of aggressive PCa adopted across the various GWAS may be considered as another limitation (Supplementary Table 1). Gleason Score is an important tool in the assessment of PCa prognosis and most studies include this characteristic into their definition of aggressive disease, however, some consider Gleason score ≥ 8 while others Gleason score ≥ 7 [99-106]. This difference impairs the uniform stratification of cases across the different studies, as those using Gleason score ≥ 8 often consider Gleason score 7 to be of intermediate risk, thus diluting the subgrouping of aggressive PCa cases and possibly influencing the results. Moreover, a Gleason score 7 might encompass distinct prognosis features, ranging from well-differentiated tumors with a minor poor differentiated component to more poorly differentiated tumors with a minor well-differentiated component, inducing heterogeneity in case selection [112].

Another major limitation is the case-control design implemented by most GWAS [99, 100, 103-105]. This strategy allows the comparison of the allele frequencies between aggressive PCa cases and healthy controls, not considering the possibility that the studied variants may not only be associated with aggressive but also with indolent disease [101, 106]. Accordingly, one of these GWAS conducted by Sun and collaborators has showed that the reported marker was also associated with non-aggressive PCa [100]. In order to overcome this putative drawback, the execution of case-case studies, comparing aggressive cases to indolent ones, might be a better strategy to increase study power and specifically determine which genetic variants may underpin the PCa aggressive phenotype [101]. As so, it becomes imperative the implementation of approaches that allow the confirmation of these findings, as validation studies in independent populations, namely of different ancestries, and post-GWAS analysis, as fine mapping and functional

studies, have also emerged as powerful tools to help understand the biological plausibility of each locus and their involvement in PCa etiologic pathways [89, 113]. Moreover, validation studies tend to be more heterogeneous regarding ethnicity, which allows testing the generalizability of GWAS-findings on subjects of different genetic backgrounds [47].

Ultimately, the main aim of PCa GWAS is the translation of its findings into the clinical setting [72]. Genetic profiling based on GWAS' SNPs can identify individuals at higher risk for aggressive PCa, and thus allowing a better understanding of which patients to have under tighter surveillance. Aside from being biomarkers for susceptibility, SNPs can also be biomarkers for prediction of outcome [113]. This would allow to predict the response to therapy and recognize those who are more prone to present treatment-related toxicity [72, 114]. In the era of personalized medicine, they can also offer new information on PCa related pathways, possibly providing new therapeutic targets for aggressive or castration-resistant PCa, favoring a personalized treatment according to patients' genetic background (Figure 4) [72].

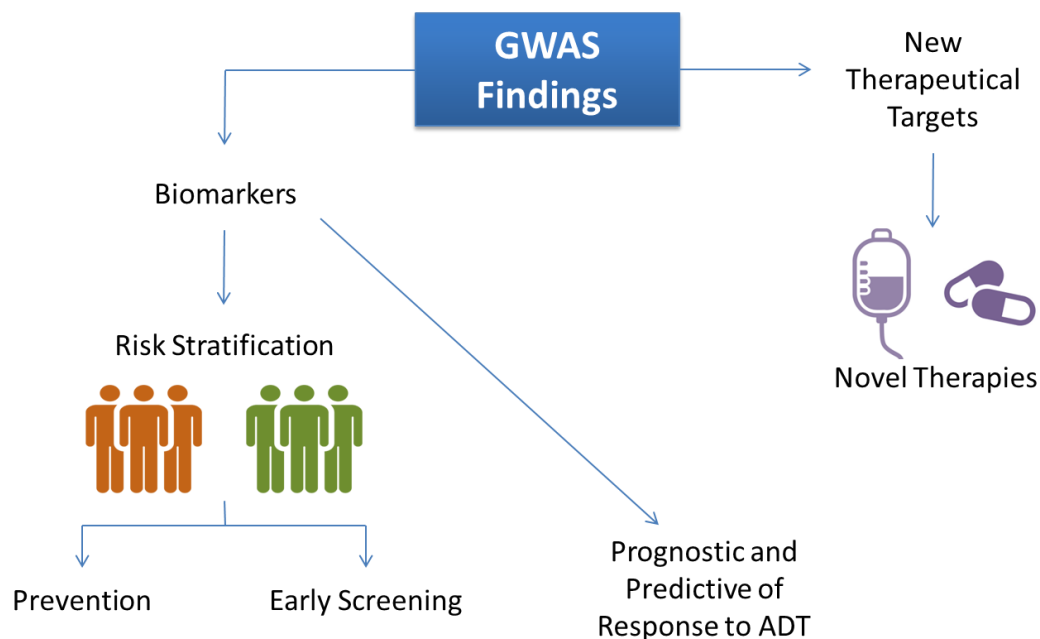


Figure 4 - Clinical application of GWAS findings. The identification of SNPs related to aggressiveness through GWAS prompts for their implementation into clinical practice. GWAS biomarkers can help, through genetic profiling, to stratify patients according to their risk of developing aggressive prostate cancer (PCa) and thus develop strategies to help prevent it and to improve early diagnosis of PCa through screening. These variants can also be used as biomarkers for prediction of response to therapy, especially to ADT. Also, GWAS can lead to the identification of novel pathways and help uncovering new therapeutical targets to benefit patients with aggressive disease. GWAS: Genome-Wide Association Study; ADT: Androgen Deprivation Therapy. Adapted From: [72, 114].

2. Objectives

2.1. Main Objective

To study the impact of aggressive prostate cancer (PCa) single nucleotide polymorphisms (SNPs) discovered through genome-wide association studies (GWAS) in the clinical outcome of a cohort of PCa patients from the North Region of Portugal.

2.2. Specific Objectives

- Conduction of a literature review on aggressive PCa GWAS;
- Selection of genetic variants associated with aggressive PCa, reported through GWAS;
- Assessment of the impact of the selected genetic variants *MLPH* rs2292884, *APBB1IP* rs1335540, *HERC2* rs6497287, *FOXO1/LINC00598* rs7327461, *RASA1/CCNH* rs35148638 and *FGF10* rs2121875 in the clinical outcome of PCa patients.

3. Materials and Methods

3.1. Study population characterization

We performed a retrospective hospital-based cohort study on 746 male patients from the North region of Portugal that were admitted in Instituto Português de Oncologia do Porto (IPO-Porto), with histopathologically confirmed PCa between 1990 and 2009. The clinicopathologic and follow-up data from patients obtained from medical records are described in Table 4. The mean age of the patients was 66 years (median = 66 years; minimum = 41 years; maximum = 88 years) and the mean follow-up time for each patient was 158.13 months (minimum = 150 months; Maximum = 166 months). At the time of diagnosis, tumor staging was performed according to the EAU and then grouped into two categories, the first presenting localized disease (T1-T2) which contained 45% of patients, and the second with advanced disease (T3-T4, N+ and/or M+) containing 55% of patients. Still at the time of diagnosis, for PSA levels, 70.5% of patients presented levels <20 ng/mL while 29.5% had levels ≥20 ng/mL, and for Gleason score, 74.8% of patients had a score <8 whereas 25.2% had a score ≥8. Of the entirety of patients in the study, 69.6% underwent definitive therapy, namely RP or RT, and 74.1% underwent treatment with ADT, immediately after diagnosis or after relapsing from first-line treatment, consisting of the following options: 77.1% of combination therapy with anti-androgens and aLHRH, 10.8% with aLHRH alone, 8.4% with anti-androgens alone and a small percentage (3.7%) with another option. Castration resistance was defined as the time from ADT initiation to two consecutive rises of PSA levels higher than the PSA nadir or radiological progression of the disease [115, 116].

Table 4 - Clinicopathological characteristics of the study population.

Characteristic	Prostate Cancer (n=746)
Age at diagnosis, yrs.	
Median (min-max)	66.09 (41-88)
Follow up time, mths.	
Mean (min-max)	158.13 (150.49-165.78)
PSA at diagnosis, ng/mL	
Median	10.9
Mean $\pm$ SD	77.9 $\pm$ 326.0
Gleason Score (%)	
<8	74.8
$\geq$ 8	25.2
Clinical Stage (%)	
T1-T2	45.0
T3-T4, N+ and/or M+	55.0
Definitive Therapy (%)	
RP/RT	69.6
No	30.4
Hormonotherapy (%)	
Yes	74.1
No	25.9

All patients signed a written informed consent prior to their enrolment in this study, according to Helsinki Declaration principles. This study was approved by the ethics committee at IPO-Porto (CES. 198/018).

3.2. SNP selection

To select the SNP variants that were used in this study, we performed a review and collected all SNPs reported through GWAS that were statistically associated with aggressive PCa. The variants gathered were submitted to the GWAS4D software that provided a ranking based on the *P*-value reported in the original study and the possible functional consequences of the variants. Taking in consideration the ranking provided by the software and the minor allele frequency (MAF) in the Iberian population (>15%) and the potential biological relevance of the genes in prostate pathways, 6 SNPs were selected: *MLPH* rs2292884, *APBB1IP* rs1335540, *HERC2* rs6497287,

FOXO1/LINC00598 rs7327461, *RASA1/CCNH* rs35148638 and *FGF10* rs2121875 (Table 5).

Table 5 - Characteristics used in the selection of the study polymorphisms.

Polymorphisms	Iberian MAF	P-value [ref]	Functional Effect	Consequence/ Gene	Gene Function
rs2292884 (A>G)	22% (G)	7.80x10 ⁻⁶ [104]	GG genotype was associated with lower gene expression in PCa tissue [117]; G allele was correlated with decreased levels of protein in primary prostate tumors, and decreased RNA levels in tumor tissue [118].	Missense variant <i>MLPH</i>	Complex <i>MLPH/Rab27/myosin</i> is involved in the polarized transport of melanosomes, in the secretion pathway and involved in hair and skin pigmentation; Deregulation of <i>MLPH</i> has been found in various tumors.
rs1335540 (A>T)	38.8% (T)	4.64x10 ⁻⁵ [100]	-	Regulatory region variant (2KB Upstream) <i>APBB1IP</i>	Involved in RET and integrin alpha-IIb beta-3 signaling pathway; Appears to function in the signal transduction from RAS activation to cytoskeletal remodeling and mediates Rap1-induced adhesion.
rs6497287 (T>C)	15.9% (C)	4.00x10 ⁻³ [102]	-	Intron variant <i>HERC2</i>	Involved in regulation of p53 transcriptional activity and in DNA damage response and repair.
rs7327461 (T>G)	19.6% (G)	9.11 x10 ⁻⁵ [100]	-	Intron variant <i>FOXO1</i> <i>LINC00598</i>	<i>FOXO1</i> plays a role in glucose metabolism, cell survival, proliferation and differentiation and oxidative stress resistance. <i>LINC00598</i> is a long non-coding RNA that regulates the expression of various target genes.
rs35148638 (A>C)	21.5% (C)	6.49x10 ⁻⁹ [106]	-	Intron variant <i>RASA1</i> (~79kb downstream of <i>CCNH</i>)	<i>RASA1</i> acts as a RAS function suppressor, is involved in the regulation of cell differentiation and proliferation and controls blood vessel growth. <i>CCNH</i> is involved in RNA transcription and cell cycle control, and in androgen receptor signaling and nucleotide excision repair (NER) pathways.
rs2121875 (G>T)	32.7% (G)	7.15x10 ⁻⁵ [119]	-	Intron variant <i>FGF10</i>	Involved in cell proliferation, differentiation and invasion.

Minor allele frequency values for were obtained on "Ensembl" database.

Gene functions were obtained on "GeneCards" database.

OR – odds ratio; MAF – minor allele frequency; (-) – data not applicable

3.3. Laboratory procedures

Peripheral venous blood samples were collected by forearm venipuncture using Ethylenediamine tetraacetic acid (EDTA)-containing tubes and genomic DNA was extracted from the samples using the extraction kit Qiagen®, QIAmp DNA Blood Mini Kit (Qiagen® 51106).

The genotyping of the variants was performed by Real-Time Polymerase Chain Reaction (Real-Time PCR) using the TaqMan® Allelic discrimination Assay methodology that uses probes with different fluorescent reporters (VIC and FAM) and thus allowing to differentiate homozygous from heterozygous samples [120]. For each variant, the used assays were as shown in Table 6.

Table 6 - Context sequence of the study polymorphisms.

Polymorphism	[VIC/FAM] targeting allele
rs2292884 (C__25939813_10)	CTGCAGATCTTTGAGCTGAATAAGC[A/G]TATTTTCAGCTGTGGAATG CCTGCTG
rs1335540 (C__8868603_10)	CCAGAAAGTGCTGAAGTAAACACTT[A/T]GGGAGACCGTTGCAACA TAAAGCGG
rs6497287 (C__31479154_10)	TTAAAGAGAACTAAGGTTTTTGCAT[C/T]GTTTAGGTGGAGAATATA CTAATAA
rs7327461 (C__30473622_20)	CTAGAGATTTCTCTGCTAACACAGG[G/T]GAGGGGCCCAGCACACA AGTGGCCC
rs35148638 (C__3298775_10)	GCTAACAGGCCAAAACTTCAGTCTT[A/C]GTCCATTTTCTATTACTG CAGCAGA
rs2121875 (C__16105569_10)	GATGCTGTAATGCTATTTAAAAATA[A/C]ATTTAGTAGGAAACTCAAA ACCTTC

Real-time PCR was performed according to an optimized protocol developed by our group, and the reactions were carried out using 6µL of a reaction mix containing 2.5µL of TaqMan® SNP Genotyping Assay mix, 2.375µL of sterile water, 0.125µL of the corresponding TaqMan® SNP Genotyping Assay and 1µL of genomic DNA. Thermal conditions were set to optimize enzyme activity, with an initial stage of 10 minutes at 95°C for the activation of the Taq DNA polymerase, followed by 45 cycles of 15 sec at 92°C to denature the DNA double-strand plus 1 minute at 60°C for primer annealing and extension. Amplification was detected through the StepOne Plus Real-Time PCR system and StepOne Software (version 2.3 Applied Biosystems) (Figure 5). Two negative controls

were used in each reaction in order to guarantee that there was no contamination, and replicas were amplified for 10% of the samples. The analysis of the genotyping results was independently conducted by two researchers who were unacquainted with patients' clinical information.

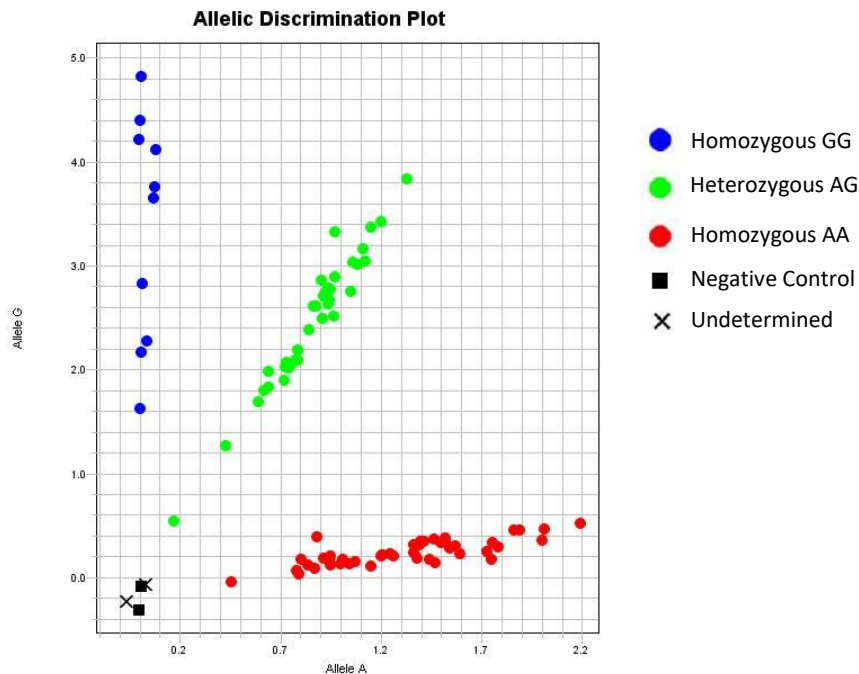


Figure 5 - Example of an allelic discrimination plot for rs2292884

3.4. Statistical analysis

Statistical analysis was performed with the IBM® SPSS® Statistics for Windows™ software (version 24.0, SPSS Inc, 2016).

Associations between patients' clinicopathologic characteristics and the genetic variants were tested using a chi-square test (χ^2) for categorical variables, namely PSA levels at diagnosis (PSA <20ng/mL vs. PSA ≥20ng/mL), Gleason score (<8 vs. ≥8) and clinical stage (localized vs. advanced), and a student's t-test was used for the continuous variable age.

The survival analysis and the time to castration resistance analysis were performed using the Kaplan-Meier method to obtain survival curves, and the log-rank test to compare survival probabilities. After an initial analysis between survival curves under the log-additive model, the most suitable genetic model was established for each variant.

Subgroup stratifications were performed according to PSA levels at diagnosis, Gleason score and D'Amico Classification (non-high-risk vs. high-risk) (Table 1) [24].

The OS outcome was defined as the time interval between diagnosis and the patients' death (by PCa or any other cause) or the last clinical evaluation. Risk of death was determined using a Cox proportional hazard model, with a 95% CI.

The time to castration resistance outcome was defined as the time interval between the beginning of ADT until castration resistance (or the last clinical evaluation), in patients who were submitted to ADT. Risk of castration resistance was also determined using a Cox proportional hazard model, with a 95% CI, and adjusted for covariates Gleason score (Gleason <8 vs. Gleason ≥8) and clinical stage (localized vs. advanced) in a multivariate model. The Cox proportional HR model was validated using a bootstrap resampling to investigate the stability of risk estimates (1000 replications).

All statistical analysis were two-sided, with an established level of significance of $P<0.05$.

4. Results

4.1. Descriptive statistics of the polymorphisms and the association with clinicopathologic characteristics

The obtained genotypic distributions, MAF and genotyping failure for each variant are described in Table 7.

Table 7 - Descriptive statistics obtained for each polymorphism.

Genetic polymorphism	MAF (%)	Genotype distribution	Genotyping failure (%)
<i>MLPH</i> rs2292884	25.0	AA 56.3% (n=420) AG 37.4% (n=279) GG 6.3% (n=47)	0.0
<i>APBB1IP</i> rs1335540	38.5	AA 38.1% (n=265) AT 46.8% (n=326) TT 15.1% (n=105)	6.7
<i>HERC2</i> rs6497287	14.5	TT 72.1% (n=533) TC 26.7% (n=197) CC 1.2% (n=9)	0.9
<i>FOXO1/LINC00598</i> rs7327461	16.6	TT 69.4% (n=518) TG 28.1% (n=209) GG 2.5% (n=19)	0.0
<i>RASA1/CCNH</i> rs35148638	24.2	AA 58.6% (n=435) AC 34.4% (n=255) CC 7.0% (n=52)	0.5
<i>FGF10</i> rs2121875	37.7	TT 38.3% (n=285) TG 48.0% (n=357) GG 13.7% (n=102)	0.3

MAF – minor allele frequency

There were no significant differences between genotypes of the studied polymorphisms and clinicopathologic characteristics of the patients, such as age, PSA at diagnosis, Gleason score and clinical stage (data not shown), except for variant *MLPH* rs2292887 and age ($P=0.006$), and variant *HERC2* rs6497287 and PSA levels at diagnosis ($P=0.030$).

4.2. Overall survival analysis

The survival analysis showed no statistically significant differences between survival time and the majority of the variants in a log-additive model ($P=0.511$, $P=0.084$, $P=0.285$, $P=0.822$, $P=0.941$, for *MLPH*, *HERC2*, *FOXO1*, *RASA1/CCNH* and *FGF10* respectively), only showing a significant association for *APBB1IP* rs1335540 ($P=0.042$). The association with *APBB1IP* rs1335540 was even more significant when considering a dominant genetic model (TT/AT genotype vs. AA genotype) ($P=0.017$), with AA genotype carriers presenting a decreased survival when compared to T allele. Furthermore, when stratifying patients for PSA levels at diagnosis, the *APBB1IP* rs1335540 variant continued to show an association with OS in the dominant genetic model for patients with PSA <20ng/mL ($P_{\text{PSA}<20}=0.002$ vs. $P_{\text{PSA}\geq 20}=0.763$). Similarly, in a recessive genetic model (TT/GT genotype vs. GG genotype), *FOXO1/LINC00598* rs7327461 also showed statistically significant differences in the group with PSA <20ng/mL, when stratified for PSA levels at diagnosis. In this subgroup, there was a significant decrease in OS in individuals with the GG genotype when compared to carriers of the T allele ($P_{\text{PSA}<20}=0.019$ vs. $P_{\text{PSA}\geq 20}=0.111$). None of the remaining variants showed association when stratifying for PSA levels at diagnosis (data not shown). An additional analysis stratified by Gleason score was performed but no significant results were found (data not shown).

A subsequent analysis stratified according to D'Amico classification showed statistically significant differences in 3 variants, all in the non-high-risk group. In the dominant genetic model, *APBB1IP* rs1335540 AA homozygous presented a significant decrease in OS when comparing to carriers of the T allele ($P_{\text{non-high-risk}}=0.020$ vs. $P_{\text{high-risk}}=0.104$) (Figure 6).

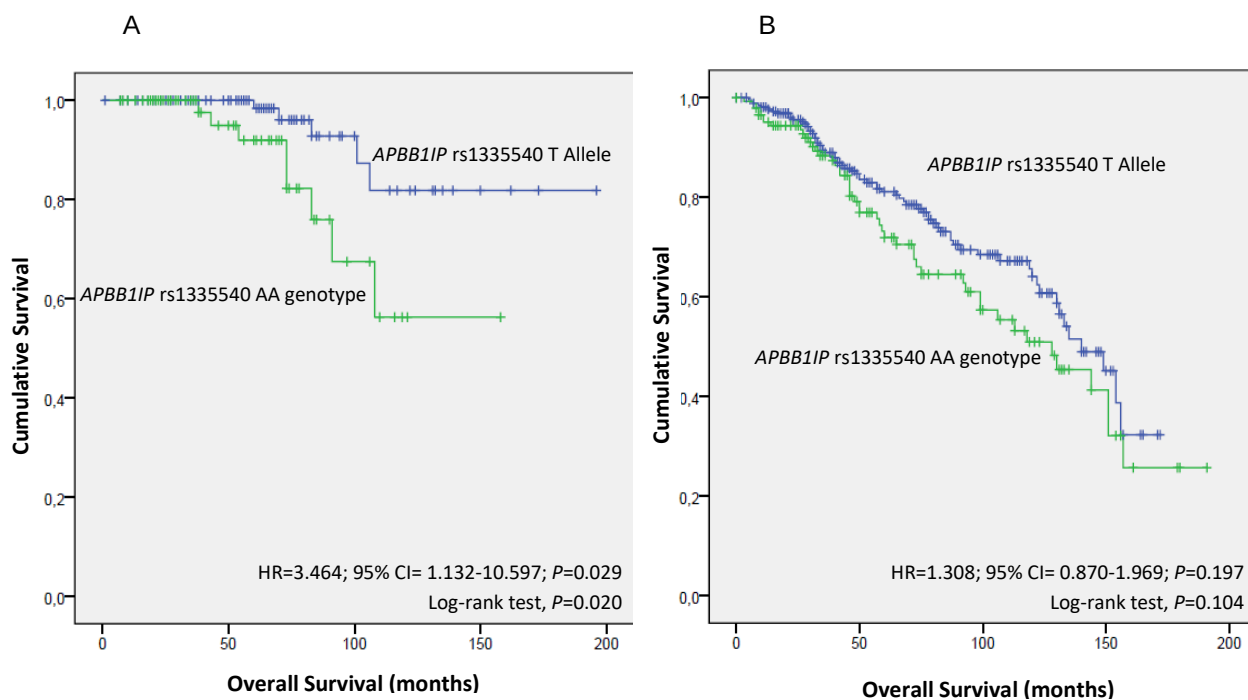


Figure 6 - Overall survival by Kaplan-Meier and log-rank test according to *APBB1IP* rs1335540 polymorphism (dominant genetic model), stratifying by D'Amico classification. A) Subgroup of patients with non-high-risk D'Amico classification; B) Subgroup of patients with high-risk D'Amico classification.

Statistically significant differences were also found for *MLPH* rs2292884 in the recessive genetic model (AA/AG genotype vs. GG genotype). In the subgroup of patients with non-high-risk classification, GG genotype carriers had a significantly decreased survival time ($P_{\text{non-high risk}}=0.019$ vs. $P_{\text{high risk}}=0.670$) than A allele carriers (Figure 7).

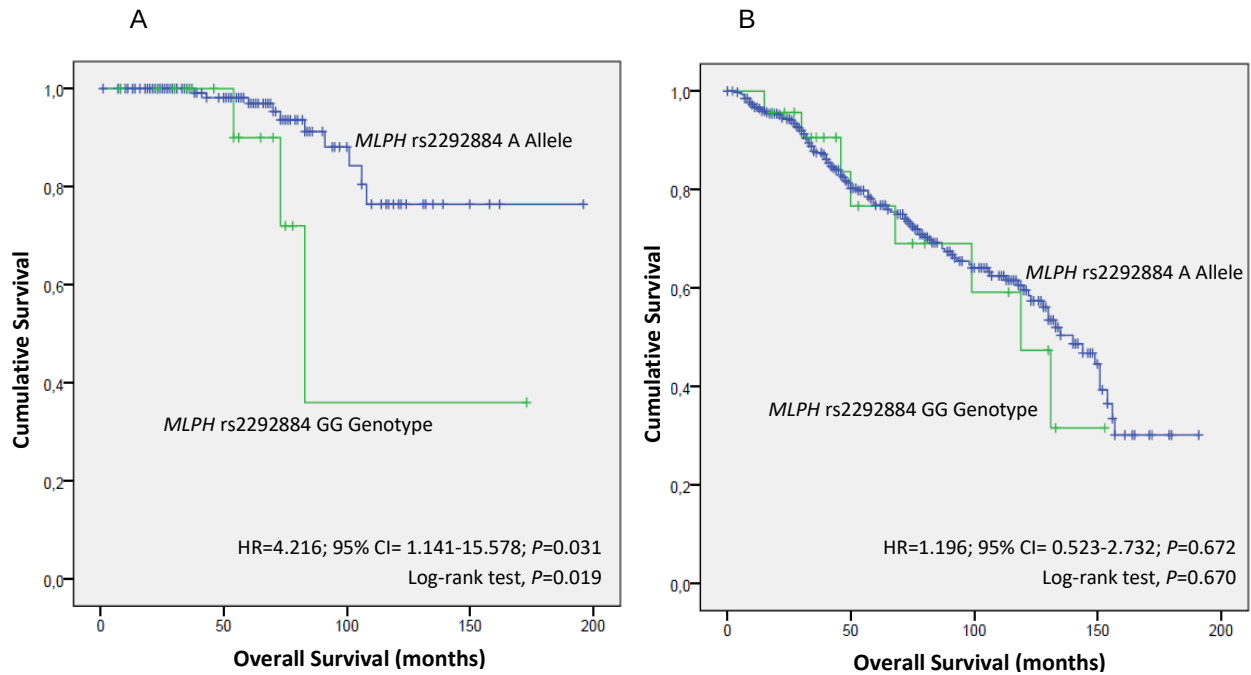


Figure 7 - Overall survival by Kaplan-Meier and log-rank test according to *MLPH* rs2292884 polymorphism (recessive genetic model), stratifying by D'Amico classification. A) Subgroup of patients with non-high-risk D'Amico classification; B) Subgroup of patients with high-risk D'Amico classification.

Regarding *FGF10* rs2121875, in the dominant genetic model, we observed a significant increase in survival in individuals with the G allele, in comparison with carriers of the TT genotype ($P_{\text{non-high risk}}=0.010$ vs. $P_{\text{high risk}}=0.340$), in the non-high-risk classification subgroup of patients (Figure 8).

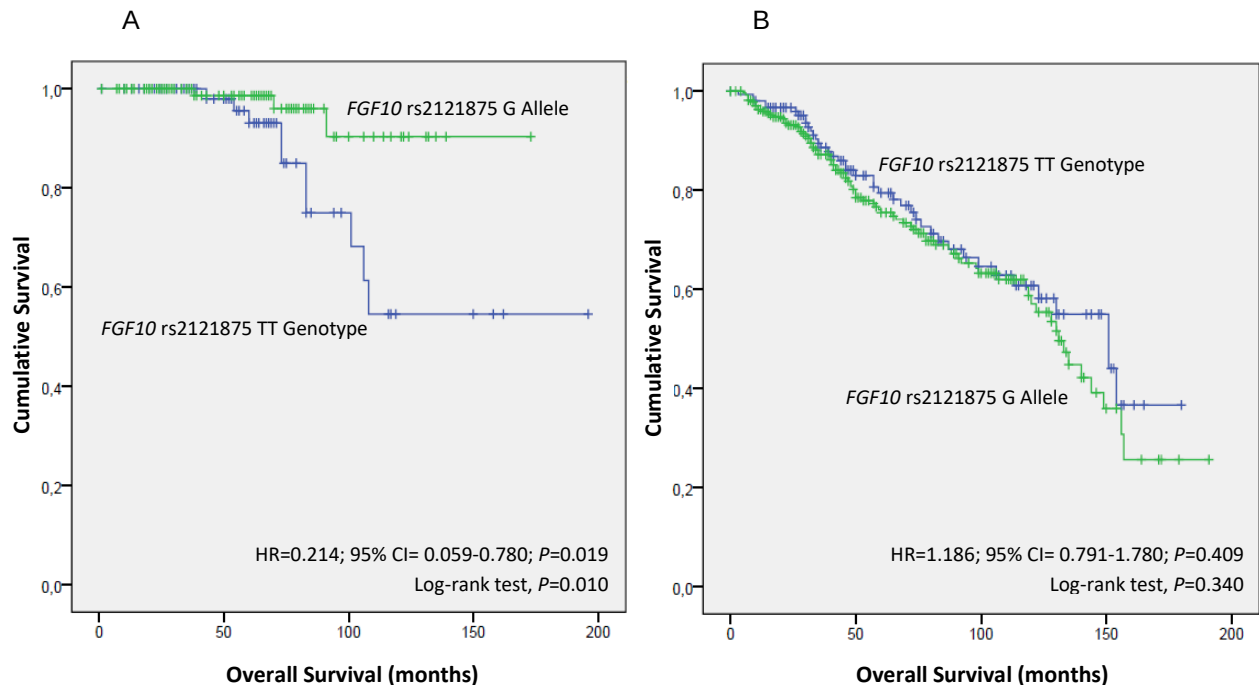


Figure 8 - Overall survival by Kaplan-Meier and log-rank test according to *FGF10* rs2121875 polymorphism (dominant genetic model), stratifying by D'Amico classification. A) Subgroup of patients with non-high-risk D'Amico classification; B) Subgroup of patients with high-risk D'Amico classification.

4.3. Time to castration resistance analysis

Regarding castration resistance, we observed that there were statistically significant differences for three of the variants in the subgroup of patients with PSA <20ng/mL: *HERC2* rs6497287, *FOXO1/LINC00598* rs7327461 and *RASA1/CCNH* rs35148638. Analysis stratifying for Gleason score and D'Amico classification did not result in any significant differences in any of the variants (data not shown).

For the *HERC2* rs6497287 variant we obtained significant association with time to castration resistance in the log-additive model ($P<0.001$). Using the dominant genetic model (TT genotype vs. CC/TC genotype), although the association was not significant, there was a tendency for a decrease in time to castration resistance in carriers of the C allele ($P=0.055$). Upon further stratification with PSA levels at diagnosis, this association resumed significance for individuals with PSA levels <20 ng/mL, but not for individuals with PSA ≥ 20 ng/mL ($P_{\text{PSA}<20}=0.009$ vs. $P_{\text{PSA}\geq 20}=0.630$) (Figure 9). In the subgroup of patients with PSA levels <20 ng/mL, carriers of the C allele had a median time to

castration resistance of 70 months compared to 92 months observed for patients with the GG genotype.

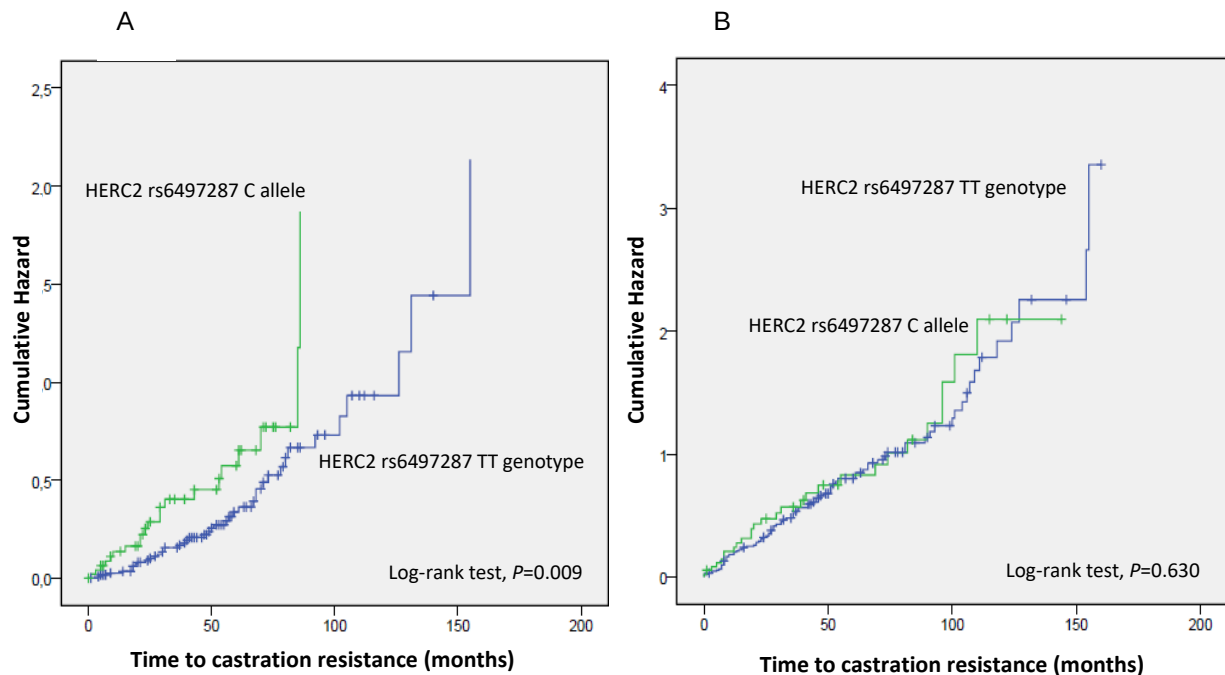


Figure 9 - Time to castration resistance by Kaplan-Meier and log-rank test according to HERC2 rs6497287 polymorphism (dominant genetic model), stratified by PSA levels at diagnosis. A) Subgroup of patients with PSA <20ng/mL; B) Subgroup of patients with PSA ≥ 20 ng/mL.

Furthermore, a multivariate analysis was performed to estimate the risk of castration resistance for patients with PSA levels <20ng/mL, adjusting for the *HERC2* rs6497287 variant with Gleason score and clinical stage, known indicators of aggressive PCa, as covariates (Table 8). We observed that C allele carriers presented a 2.1-fold increased risk to earlier castration resistance when compared to TT homozygous genotypes (HR=2.08, 95% CI=1.20–3.61, $P=0.009$, and $P_{\text{bootstrap}}=0.011$).

Table 8 - Multivariate Cox regression analysis for *HERC2* rs6497287 on the risk of earlier castration resistance in patients with PSA levels <20ng/mL at diagnosis, considering Gleason score and clinical stage.

	HR	95% CI	P-value
<i>HERC2</i> rs6497287			
TT genotype vs. C Allele	2.08	1.20–3.61	0.009*
Gleason score	1.96	1.14–3.39	0.016
Localized vs. Advanced	2.04	1.07–3.91	0.032

* $P=0.011$ after bootstrap based on 1000 samples

The *FOXO1/LINC00598* rs7327461 polymorphism did not show significant differences in time to castration resistance in a log-additive model ($P=0.262$). However, when stratifying by PSA levels at diagnosis in a recessive genetic model, in the subgroup with PSA levels <20 ng/mL, we observed significantly reduced time to castration resistance for patients with the GG genotype ($P_{\text{PSA}<20}=0.001$ vs. $P_{\text{PSA}\geq 20}=0.147$) (Figure 10). These patients, when compared to carriers of the T allele, presented a median time to castration resistance shorter by 45 months.

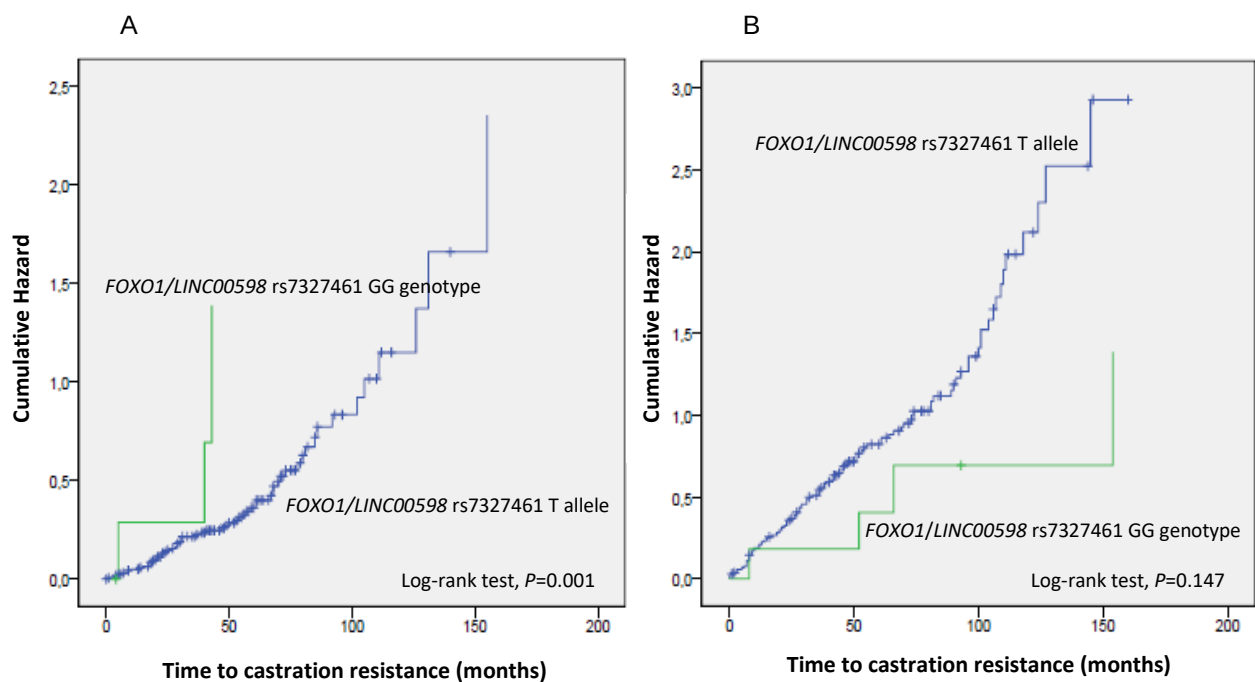


Figure 10 - Time to castration resistance by Kaplan-Meier and log-rank test according to *FOXO1/LINC00598* rs7327461 polymorphism (recessive genetic model), stratifying for PSA levels at diagnosis; A) Subgroup of patients with PSA <20ng/mL; B) Subgroup of patients with PSA ≥ 20 ng/mL.

Using the multivariate Cox regression analysis, we found that carriers of *FOXO1/LINC00598* rs7327461 GG homozygous genotypes presented 5.1-fold increased risk of earlier castration resistance, when compared to T allele carriers, using Gleason score and clinical stage as covariates (HR=5.08, 95% CI=1.67–15.44, $P=0.004$ and $P_{\text{bootstrap}}=0.004$) (Table 9).

Table 9 - Multivariate Cox regression analysis for *FOXO1/LINC00598* rs7327461 on the risk of earlier castration resistance in patients with PSA levels <20ng/mL at diagnosis, considering Gleason score and clinical stage.

	HR	95% CI	P-value
<i>FOXO1/LINC00598</i> rs7327461			
T Allele vs. GG genotype	5.08	1.67–15.44	0.004*
Gleason score	1.62	0.92–2.86	0.094
Localized vs. Advanced	2.36	1.22–4.58	0.011

* $P=0.004$ after bootstrap based on 1000 samples

No statistically significant difference was observed considering the variant *RASA1/CCNH* rs35148638 in the log-additive model ($P=0.374$). However, in the recessive genetic model stratified for PSA levels at diagnosis there were significant differences between the carriers of CC homozygous genotype and A allele ($P_{\text{PSA}<20}=0.042$ vs. $P_{\text{PSA}\geq 20}=0.331$). In this model, the median time to castration resistance development in CC genotype carriers was shorter than A allele carriers (53 months vs. 85 months) (Figure 11).

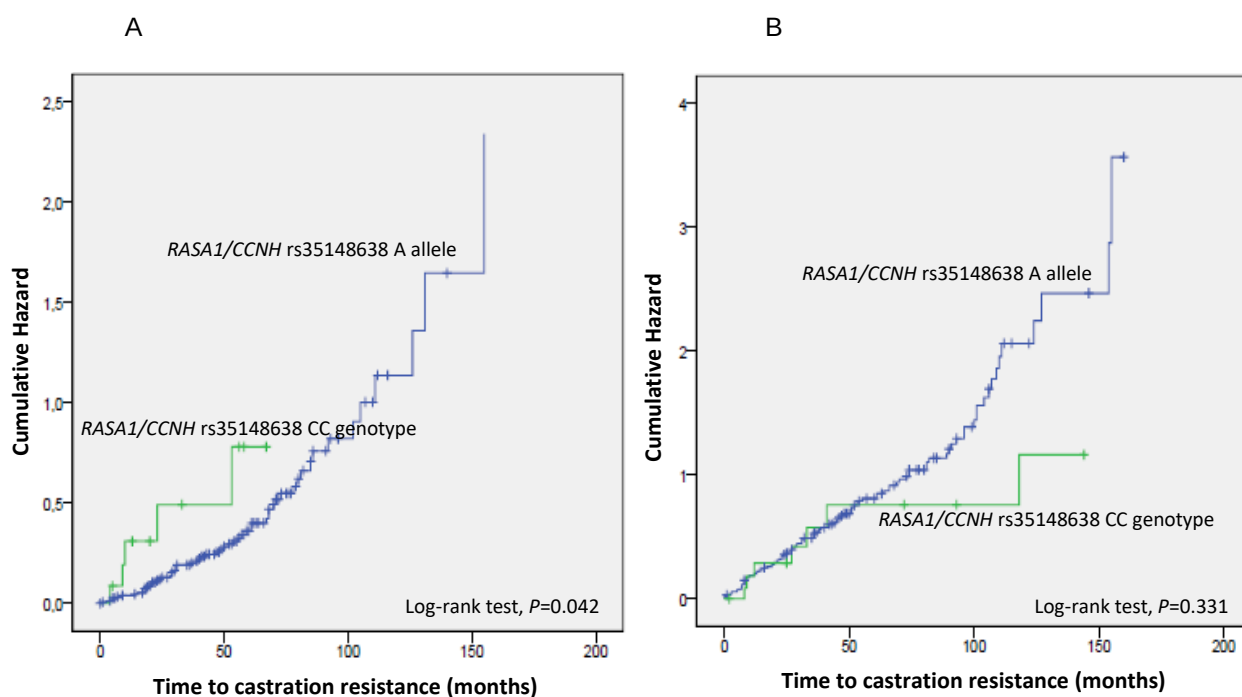


Figure 11 - Time to castration resistance by Kaplan-Meier and log-rank test according to *RASA1/CCNH* rs35148638 polymorphism (recessive genetic model), stratified by PSA levels at diagnosis in a. A) Subgroup of patients with PSA <20ng/mL; B) Subgroup of patients with PSA ≥20ng/mL.

In the multivariate Cox regression, we observed that patients who carry the *RASA1/CCNH* rs35148638 CC homozygous have a 2.6-fold increased risk of earlier castration resistance than patients that carry the A allele, after adjusting for Gleason score and clinical stage (HR=2.60, 95% CI=1.01–6.66, $P=0.047$ and $P_{\text{bootstrap}}=0.056$) (Table 10).

Table 10 - Multivariate Cox regression analysis for *RASA1/CCNH* rs35148638 on the risk of earlier castration resistance in patients with PSA levels <20ng/mL at diagnosis, considering Gleason score and clinical stage.

	HR	95% CI	P-value
<i>RASA1/CCNH</i> rs35148638			
A Allele vs. CC genotype	2.60	1.01–6.66	0.047*
Gleason score	1.84	1.06–3.20	0.031
Localized vs. Advanced	1.99	1.03–3.84	0.040

* $P=0.056$ after bootstrap based on 1000 samples

Additionally, a multivariate Cox regression combining all of the risk genotypes of the variants demonstrating significant results was conducted. We observed a 2.5-fold increased risk of earlier castration resistance in individuals who present at least one of the risk genotypes (HR=2.50, 95% CI=1.46–4.28, $P=0.001$, $P_{\text{bootstrap}}=0.002$) (Table 11). In this analysis, this group of individuals had a median time to castration resistance of 70 months, compared to 102 months for patients without any risk genotypes.

Table 11 - Multivariate Cox regression analysis for combined *HERC2* rs6497287, *RASA1/CCNH* rs35148638 or *FOXO1/LINC00598* rs7327461 on the risk of earlier castration resistance in patients with PSA levels<20ng/mL at diagnosis, considering Gleason score and clinical stage.

	HR	95% CI	P-value
rs6497287 C allele or rs35148638 CC genotype or rs7327461 GG genotype	2.50	1.46–4.28	0.001*
Gleason score	1.89	1.09–3.26	0.023
Localized vs. Advanced	2.11	1.10–4.04	0.024

* $P=0.002$ after bootstrap based on 1000 samples

Lastly, in the multivariate Cox regression combining pairs of risk genotypes of the former three variants showed a 4.3-fold increased risk for earlier castration resistance in individuals that carry two of the risk genotypes comparing with those who present one or no risk genotype (HR=4.30, 95% CI=1.51–12.20, $P=0.006$ and $P_{\text{bootstrap}}=0.002$) (Table 12). The median estimated time to castration resistance in patients with two risk genotypes was 43 months, compared to the 86 months of the patients who only presented 1 or none risk genotypes.

Table 12 - Multivariate Cox regression analysis for combined *HERC2* rs6497287 and *FOXO1/LINC00598* rs7327461 or *HERC2* rs6497287 and *RASA1/CCNH* rs35148638 on the risk of earlier castration resistance in patients with PSA levels<20ng/mL at diagnosis, considering Gleason score and clinical stage.

	HR	95% CI	P-value
rs6497287 C allele and rs7327461 GG genotype or rs6497287 C allele and rs35148638 CC genotype[#]	4.30	1.51–12.20	0.006*
Gleason score	1.89	1.10–3.25	0.021
Localized vs. Advanced	2.08	1.09–3.98	0.026

*P=0.002 after bootstrap based on 1000 samples

[#]Combination of rs7327461 GG genotype and rs35148638 CC genotype was not present in any patient

5. Discussion

Despite the decrease in PCa mortality rates over the last couple of years and the extended life expectancy observed in the majority of patients, there is still a group of patients who present with aggressive disease that have a worst prognosis [1, 11]. Hormonotherapy helps to treat these patients with more advanced cancer, but the development of castration resistance, which eventually presents in the majority of individuals who are subjected to ADT, is a major factor for the mortality observed. This resistance translates in a shorter survival expectancy for patients, as there is no definitive treatment [43]. It is difficult to predict those who are at higher risk of progressing to a castration resistance state. Clinicians rely on clinicopathological characteristics, such as PSA and Gleason score, to help define the course of treatment more suited for each case, but these variables have some flaws regarding specificity and sensitivity that can cause clinicians to overlook the aggressiveness of the disease [24, 32]. Additionally, there is not a well-defined marker for risk of castration resistance making it difficult to assess prognosis.

This prompts a need for biomarkers that can help stratify patients according to their risk of relapse in order to better tailor treatment schemes and dosages. Especially genetic markers that allow predicting the risk of relapse based on patients' genetic background. GWAS has been an important tool in the search for new genetic markers associated with aggressive PCa, but the majority of the resulting associations have yet to be considered conclusive. To achieve this, further and more extensive studies need to be conducted in order to confirm those associations so the variants can be implemented in clinical practice [70]. Hence, this study aims to assess the impact of aggressive PCa GWAS variants on the clinical outcome of Portuguese patients.

Unlike with candidate gene studies, for which a gene is chosen *a priori* based on the pathways it is involved in and the possible mechanisms leading to certain phenotype, GWAS allow for the discovery of markers in genes with no previous connections with the phenotype, offering a wider view of the genetics involved. However, this advantage can also be a disadvantage because the associations resulting from GWAS are often SNPs not previously studied that fall on genes with no known implication on the phenotype, hampering the understanding of their functional mechanisms [75].

The functionality of a genetic variant depends on where they are located in the genome and which genes they affect. Functional consequences can vary from having no effect to altering gene expression or protein structure [67]. In fact, the majority of variants, like intronic ones, are located in non-coding regions which difficult the understanding of their functional consequences [121]. If the variants are lying within regulatory regions they can (1) affect the regulation of gene expression, (2) theoretically influence the methylation status of a gene, (3) lead to the alteration of transcription factor (TF) binding

sites, (4) affect the stability of mRNA, and (5) lead to alternative splicing that can result in alternative isoforms of the protein with variations in degree of functionality [122]. Contrarily, variants located in coding regions, namely in exons like missense variants, have the potential to cause not only alterations in protein structure and consequently in protein function, but also alterations in mRNA or protein quantity [122].

For this study, 4 intronic variants (*HERC2* rs6497287, *FGF10* rs2121875, *RASA1/CCNH* rs35148638 and *FOXO1/LINC00598* rs7327461), a regulatory region variant (*APBB1IP* rs1335540) and a missense variant (*MLPH* rs2292887) were studied (Supplementary Table 1) [100, 102, 104, 106, 119, 123]. Some of these variants are located within or close to genes that have tumor-suppressing qualities like *RASA1* and *FOXO1*, while others are located on genes like *APBB1IP*, *CCNH* and *FGF10*, with tumorigenic and tumor progressing properties [124-128]. Concerning *MLPH*, *HERC2* and *LINC00598* there is not much information available on their implications in cancer.

The first outcome evaluated was OS, as it is one of the most relevant measures used in the clinical setting. Statistical significant associations with OS stratified according to D'Amico classification were obtained for variants *APBB1IP* rs1335540, *MLPH* rs2292884 and *FGF10* rs2121875, according to the most suited genetic model.

For variant *APBB1IP* rs1335540 the association with OS was observed only in the subgroup of patients with non-high-risk D'Amico classification, where the AA genotype resulted in a significantly decreased survival time compared to the T allele ($P_{\text{non-high-risk}}=0.020$ vs. $P_{\text{high-risk}}=0.104$). The Cox analysis in this subgroup confirmed the association, stating that AA genotype carrier had a 3.5 higher risk of death than carriers of the T allele ($P=0.029$). This variant is located in the promoter region of *APBB1IP*, which can alter the promoter's activity and consequently affect gene expression and RNA stability [129]. *APBB1IP*, also known as *RIAM*, encodes for Rap1-GTP-interacting adapter molecule that interacts with RAP1, a GTPase whose activation is related to migration and metastasis in multiple cancer models, including PCa [126, 130, 131]. More specifically, RIAM-RAP1 complex is involved in integrin-mediated cell adhesion and migration, important in driving metastasis formation, like proved in a study that showed that silencing RIAM led to delayed metastization in melanoma cells, while also leading to increased apoptosis [126]. Furthermore, this complex is also important for integrin-mediated T cells activation [132-134]. We hypothesize that the AA genotype influences promoter activity and ultimately culminates in overexpression of *APBB1IP*. A higher level of RIAM can translate in higher RIAM-RAP1 complex formation, which would contribute to metastization, and thus to a worst clinical outcome, as observed in the survival curves and in the Cox analysis.

The results for variant *MLPH* rs2292884 pointed for a reduced OS in GG genotype patients in the non-high-risk subgroup ($P_{\text{non-high risk}}=0.019$ vs. $P_{\text{high risk}}=0.670$). The Cox

analysis showed a 4.2-fold increased risk of death for carriers of the GG genotype when compared to A allele carriers ($P=0.031$). The *MLPH* gene encodes for the Melanophilin protein, responsible for the pigmentation of skin and hair, through its involvement in melanosome transportation. In estrogen receptor-positive breast cancer, higher expression of *MLPH* was shown to be significantly associated with longer disease-free survival [135]. There are no known mechanisms that could explain the association of *MLPH* with PCa, however, higher *MLPH* protein levels were correlated with a more advantageous risk profile, including lower Gleason score, clinical stage, PSA level, reduced tumor mass and no biochemical recurrence [118]. These evidence point towards a tumor-suppressing ability of the *MLPH* gene. Regarding the rs2292884 variant, two studies have correlated the risk allele (G) or genotype (GG) with decreased *MLPH* expression in PCa tissue [117, 118]. Furthermore, lower expression of *MLPH* was associated with a higher risk for PCa [136]. Moreover, this variant culminates in the substitution (CAT>CGT) leading to the replacement of the of a Histidine (HIS) to an Arginine (ARG), which could affect protein functionality [137]. As so, with the knowledge that the G allele of rs2292884 leads to a decrease in *MLPH* protein levels, and with the assumption that *MLPH* plays a role in tumor suppression, the negative impact of the SNP on our results is in concordance with what is described in the literature. Additionally, rs2292884 is in moderate LD with rs11891426 ($r^2=0.53$), located within an AR binding site of *MLPH*, and its risk allele is also correlated to lower *MLPH* expression, possibly forming a haplotype whose cumulative negative effect leads to a worst prognosis [118].

In the last variant associated with OS, *FGF10* rs2121875, the minor allele G conferred a significant increase in survival when compared with the TT genotype, also in the subgroup of patients with non-high-risk disease ($P_{\text{non-high risk}}=0.010$ vs. $P_{\text{high risk}}=0.340$). This was confirmed through the Cox regression, with G allele carriers showing a protective effect against death (HR=0.214; $P=0.019$). *FGF10* is a gene encoding for fibroblast growth factor 10, a protein that belongs to the fibroblast growth factor family, involved in embryonic development, cell growth and survival, and tumor growth and invasion [138]. *FGF10* was shown to promote endometrial cancer cells growth through the phosphorylation of ERK1/2 and consequent activation of the MAPK pathway, a pathway involved in cell proliferation [139]. This gene is important in prostate organogenesis, acting as a mesenchymal paracrine regulator of epithelial growth [140]. A few studies in mouse models have demonstrated the potential role of *FGF10* in PCa [128, 141, 142]. Memarzadeh and co-workers showed that overexpression of *FGF10* in prostate mesenchyme led to hyperproliferation of the prostate epithelium and formation of adenocarcinoma through activation of its receptor, FGFR1. Moreover, paracrine *FGF10* increased the levels of AR, even with low levels of testosterone available, resulting in

progression to CRPC [128]. Other authors demonstrated that depletion of AR resulted in reduced *FGF10* expression, which implies that both are involved in a feedback mechanism [143]. The only setback is that overexpression of *FGF10* has not yet been detected in human PCa cell lines, but *in vivo* analysis may prove to be different [144]. Despite having no knowledge on the functional implications of *FGF10* rs2121875, the fact that it is located in a TF binding site may alter the ability to bind to TFs. We can hypothesize that the G allele is related to a lower *FGF10* expression comparing to the TT genotype. Taking into consideration the fact that paracrine FGF10 augments levels of AR, lower levels of FGF10 would mean lower expression of AR, and thus a less aggressive phenotype with a better prognosis. This evidence suggests that the inhibition of FGF10 or its receptors could be a potential therapeutical target to delay the progression of PCa, providing a better outcome for patients.

The possible involvement of these three variants in metastization and castration resistance is correlated with their association to OS. However, in the case of *MLPH* rs2292884, despite being the only variant with a known functional consequence, its implication in PCa tumorigenesis or progression is yet to be discovered.

The other outcome evaluated was time to castration resistance. Patients on ADT are bound to acquire resistance to this therapy and experience biochemical and/or radiological progression, in which tumor cells regain the capacity to grow and proliferate despite the absence of testosterone [43]. However, some recur sooner than others, and while that may be predicted by the clinicopathological characteristic inherent to each patient, the genomic background may also be helpful to predict the response to ADT [61]. Therefore, it is also of clinical interest to assess intermediate outcomes like time to castration resistance, as they are impactful in disease management.

The results according to the most suited genetic model showed that intronic variants *HERC2* rs6497287 C allele ($P=0.009$), *FOXO1/LINC00598* rs7327461 GG genotype ($P=0.001$) and *RASA1/CCNH* rs35148638 CC genotype ($P=0.042$) were, individually, significantly associated with lower time to castration resistance in the subgroup of patients with PSA at diagnosis <20ng/mL. The multivariate Cox analysis showed that a panel with the relevant SNPs combining the risk genotypes in pairs, would significantly predict the risk of earlier castration resistance in the subgroup of patients with PSA levels <20ng/mL (HR=4.30, $P=0.006$). In this subgroup, a pair of risk genotypes leads to a decrease in time to castration resistance by half when compared to patients with only one or none (43 months vs. 86 months). None of the variants on this risk panel have been the focus of functional studies, which hampers the link between the risk genotypes and a PCa related pathway.

Considering the role of *HERC2* and *CCNH* in DNA damage response, it is expected that downregulation of these genes results in an accumulation of DNA damage. *HERC2*, an E3 ubiquitin ligase, is responsible for the recruitment of BRCA1 to DNA damage sites [102, 145]. To accomplish that, *HERC2* helps in facilitating the assembly of the RNF8-UBC13 complex to recruit BRCA1 to sites of DNA damage response to DNA double-strand breaks [146]. Also, at DNA damage sites, BRCA1 forms a complex with BRCA1-associated RING domain protein1 (BARD1) essential for BRCA1 stability [147]. *HERC2* is involved in BRCA1 stability by ubiquitinating BARD1-uncoupled BRCA1 and thus targeting it for degradation [145]. Aside from the interaction with BRCA1, *HERC2* can also bind to p53, regulating its activity. It has been showed that the depletion of *HERC2* decreases the transcriptional activity of p53, affecting DNA damage response, cellular regulation, the expression of target genes, and its tumor suppressor characteristics [148]. Our results showed that the *HERC2* rs6497287 C allele resulted in shorter time to castration resistance, so it is possible that it leads to a lower expression of *HERC2* or to an isoform of the protein with reduced functionality and less affinity to p53. Also, this SNP is in linkage disequilibrium with rs7183877 ($r^2 = 1$), also located in *HERC2*, which could, instead, be the causal variant [102].

Regarding *CCNH*, it encodes for Cyclin H, involved in the NER pathway, but also in cell cycle regulation and AR signaling. Together with CDK7 and MAT1, *CCNH* forms the cyclin-activated kinase (CAK) complex, and CAK interacts with AR increasing its transcriptional activity [149]. CAK can also phosphorylate several nuclear receptors, like the AR and the estrogen receptor (ER), and can affect the regulation of p53 when down-regulated [150, 151]. Furthermore, in a study that analysed transcriptomes of the androgen-sensitive LNCaP cell line during their progression in vivo, Romanuik and colleagues concluded that *CCNH* exhibited increased levels of expression in castration resistance and that these results may lead to increased AR signaling or altered levels of AR expression [127]. In our study, the *RASA1/CCNH* rs35148638 CC genotype is responsible for an earlier castration resistance. The location of this variant in the promoter flanking region could indicate a possible implication in transcription regulation. Combining these results with the studies available in the literature, it is possible that this variant may lead to early CRPC through the hypersensitive pathway: by increasing the expression of *CCNH*, the subsequent higher expression of *AR* can translate into more receptors to compensate for the low levels of testosterone. In this case, the involvement of this gene in the NER pathway was not sufficient to compensate for its involvement in the AR pathway. Also, this variant is intronic to *RASA1*, a tumor suppressor gene that is responsible for the suppression of RAS function, and consequently responsible for the regulation of cell proliferation and survival, and has been implicated in angiogenesis and vascularization

[152]. Moreover, down-regulation or inactivation of *RASA1* was found to be correlated with triple-negative breast cancer, with metastasis in melanoma and with worst prognosis in human hepatocellular carcinoma and breast invasive ductal carcinoma [124, 153-155]. Thus, aside from the possible implication in the AR pathway through upregulation of *CCNH*, the rs35148638 CC genotype may also be responsible for the down-regulation of *RASA1*, leading to more invasive nature of PCa.

The other variant present in this risk panel was *FOXO1/LINC00598* rs7327461. The *FOXO1* gene encodes for transcription factor FOXO1, a tumor suppressor that is frequently downregulated in PCa [156]. FOXO1 has also been correlated to the inhibition of ERG-mediated tumorigenesis, cell migration and invasion in PCa, as well as being a downstream effector of PTEN [157]. FOXO1 is also responsible for the inhibition of RUNX2-mediated cell migration, invasion and metastasis formation in PCa, by binding to this transcription factor and inhibiting its transcriptional activity [125]. More directly involved in the AR pathway, it was shown that FOXO1 has inhibitory effects on the AR transcriptional activity and that reversely, AR can also bind to the *FOXO1* promoter, inhibiting its expression [158, 159]. We hypothesize that the association of *FOXO1/LINC00598* rs7327461 GG genotype with a shorter time to castration resistance may be due to a down-regulation of this tumor suppressor gene, and consequently activation of AR transcription which would, in a feedback loop, further inhibit the expression of *FOXO1*. In line with this assumption, the lack of *FOXO1* expression would lead to an increase in the proto-oncogene ERG and the transcription factor RUNX2 action in metastization. The occurrence of metastasis following ADT is evidence that the patient is no longer responding to therapy, progressing to a phase of CRPC [24]. Accordingly, these patients would probably benefit from treatments targeting ERG or RUNX2, like FTY720, a drug that was found to be capable of suppressing *RUNX2* expression in patients with metastatic CRPC [160]. Furthermore, the GG genotype can also be responsible for the regulation of *LINC00598* target genes transcription. Jeong and co-workers showed that *LINC00598* regulates transcription of cell cycle regulator *CCND2* through the inhibition of the recruitment of FOXO1 to the promoter of *CCND2*, and also showed that a *LINC00598* knockout led to G0/G1 cell cycle arrest which in turn inhibited proliferation in various human cell lines [161].

DNA repair mechanisms are important to maintain DNA and cellular stability, while involvement in AR pathway is an obvious mechanism involved in progression to CRPC (Figure 12). Taking all these hypotheses together, it is to expect that patients that have more than one of the risk genotypes have an increased risk for shorter time to castration resistance than patients with one or none as SNPs' effects are cumulative [69].

All of the significant associations for either OS or time to castration resistance were observed in patients with theoretically less aggressive disease (PSA <20ng/mL and non-high-risk D'Amico classification). We can consider that in these patients, the impact of these variants may play a more important role than in patients with more high risk disease, since in the high-risk group, usually patients have higher Gleason score, clinical stage and possibly metastazition, which in itself, already compromises the outcome in these patients, and thus the effect of the genetic component is masked. Furthermore, it has been proposed by Qin *et al.* that cells expressing little or no PSA (PSA^{-/lo}) presented resistance to ADT and drugs, and lead to higher expression of anti-stress and survival genes. Also, these PSA^{-/lo} cells exhibited cancer stem cell-like properties, and generated cells that had high levels of PSA, suggesting that PSA^{-/lo} cells may be the CRPC stem cell population, remaining quiescent following ADT and eventually gaining resistance [162].

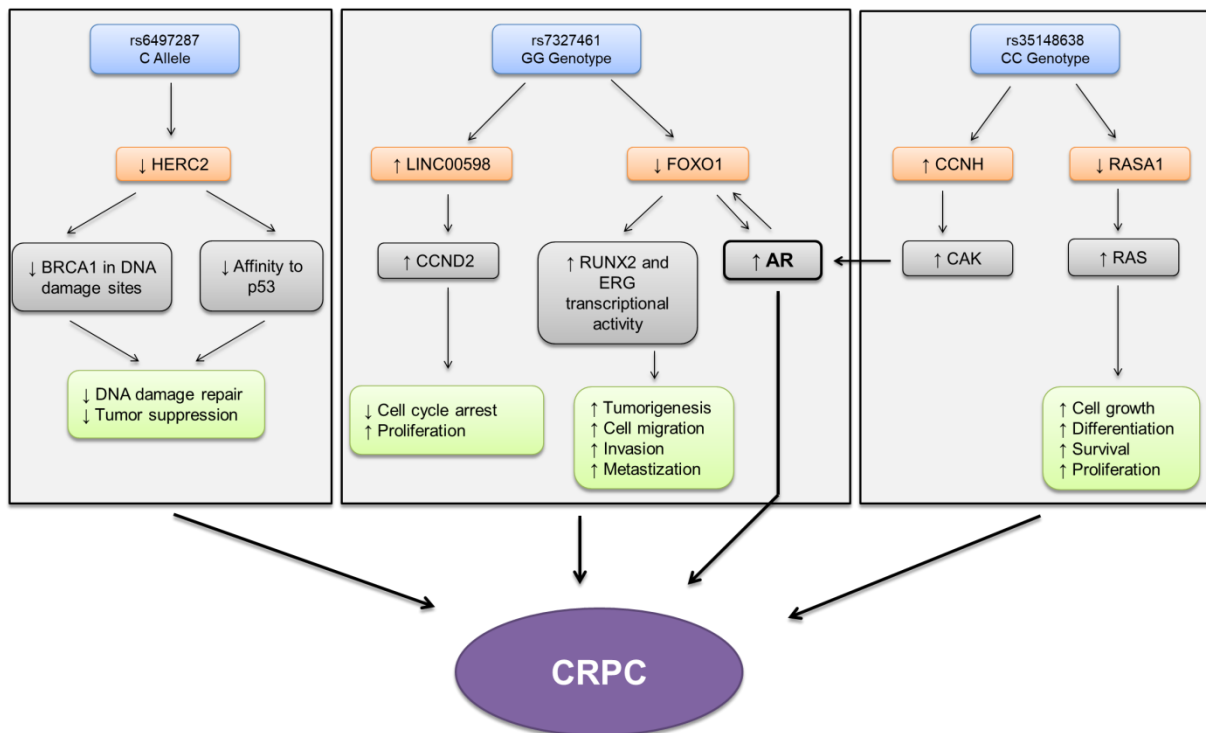


Figure 12 - Proposed model for the impact of the risk SNP panel in time to castration resistance.

6. Conclusions and Future Perspectives

The high incidence of PCa and the poor prognosis of patients with advanced disease prompt the identification of variants that can help identify individuals at higher risk of progression to CRPC or at higher risk of death. Accordingly, GWAS could be a helpful tool to discover SNPs that are associated with aggressive PCa and that have an impact on patients' outcome. GWAS' associations result from large cohorts of patients which provides great statistical power but are frequently overlooked for their frequent association with genes that have no apparent implication in the phenotype. These variants, most lying in non-coding regions, might have a cumulative effect through the regulation of genes and pathways that could be implicated in PCa. As so, it is important to further investigate GWAS' associations to better understand PCa genetics.

This study was set up to assess the impact of six GWAS-identified SNPs associated with aggressive PCa on patients' outcome. Our results showed that *APBB1IP* rs1335540, *MLPH* rs2292884 and *FGF10* rs2121875, were associated with OS and *HERC2* rs6497287, *FOXO1/LINC00598* rs7327461 and *RASA1/CCNH* rs35148638, were associated with time to castration resistance, all in the less aggressive subgroups. Although all of these associations were statistically significant, the minor genotypes for some variants were slightly underpowered within the subgroups, especially in the case of *FOXO1/LINC00598* rs7327461 GG genotype, meaning a careful interpretation of these results should be conducted. Nonetheless, apart from rs7327461, all of the remaining associations with time to castration resistance had narrow 95% CI in the multivariate Cox analysis for the PSA levels <20ng/mL subgroup, implying a higher accuracy for these associations. The result from the risk panel that combined pairs of SNPs was of particular interest for its utility in the stratification of patients with PSA levels <20ng/mL according to their risk of early progression to CRPC. This subgroup includes patients with lower PSA levels that could make clinicians take on a more conservative approach, and that, if those patients are not properly monitored, lead to the early progression of PCa to more advanced stages. Thus, in these cases, the genotyping of this SNP risk panel could be performed in a liquid biopsy alongside with the PSA analysis, to help predict response to ADT.

For future studies, we propose more extensive cohorts to improve statistical power and more ethnical diverse backgrounds to verify if our results can be generalized into other populations. Also, post-GWAS studies can provide a better and more detailed understanding of the putative functional mechanisms of the studied variants. Fine mapping would allow the discovery of other SNPs in high LD and prioritize them according to which is more likely to be the causal variant. Moreover, functional studies assessing the regulation of gene expression, transcription or translation at RNA or protein level, are of great importance to evaluate the functional consequences of these variants, both in

androgen-dependent and androgen-independent PCa cell lines. Afterwards, it is important to perform *in vitro* studies to evaluate if pathways involving the genes affected by the variants could be potential therapeutical targets for PCa.

Ultimately, the identification of these variants and genetic profiles may help to establish new biomarkers for outcome prediction, allowing the evaluation of therapy response, namely ADT. Furthermore, in the era of personalized medicine, they can provide additional information on new therapeutical targets, offering more effective and individualized treatment options based on patients' genetic profile.

7. Appendix

Appendix 1

Supplementary Table 1 - Overview of aggressive PCa susceptibility GWAS.

Study	Population	Definition of Aggressive PCa	SNP	Cases/Controls	MAF	Locus	Consequence and Gene	OR95%	P-value
Duggan <i>et al.</i> 2007	European Ancestry and African Americans	Stage T3/T4 or N+ or M+ or Gleason Score ≥ 7 or pre-operative PSA value ≥ 20 ng/mL.	rs1571801	1242/917	0.16	9q33.2	intron variant <i>DAB2IP</i>	1.36	1.0×10^{-3}
Sun <i>et al.</i> 2009	European Ancestry – Swedish Population (CAPS)	Stage T3/T4 or N+ or M+ or differential grade III or Gleason Score ≥ 8 or preoperative PSA value ≥ 50 ng/mL. (CAPS)	rs11580380	498/494 (CAPS)	0.09	1p21.1	intergenic variant	x / 1.58	x / 8.99×10^{-5}
			rs9684740		0.26	4p16.2	intron variant <i>C4orf50</i>	0.65 / x	1.63×10^{-5} / x
			rs4466060		0.13	4q28.3	intergenic variant	x / 1.45	x / 4.30×10^{-5}
			rs10032223		0.14	4q35.2	intergenic variant	1.68 / x	4.05×10^{-5} / x
			rs2548035		0.36	5p13.2	Intron variant <i>ADAMTS12</i>	0.66 / x	2.47×10^{-5} / x
	European Ancestry – European Americans (CGEMS)	Stage T3/T4 or Gleason Score ≥ 7 (CGEMS)	rs6580317	737/1105 (CGEMS)	0.14	5q31.3	intron variant <i>KCTD16</i>	0.76 / 0.71	3.0×10^{-2} / 1.15×10^{-4}
			rs4448013		0.49	5q34	intron variant <i>LOC105377700</i>	0.73 / 0.71	2.0×10^{-2} / 3.47×10^{-4}
			rs221770		0.10	7q22.1	intergenic variant	1.77 / 1.38	9.75×10^{-5} / 3.31×10^{-3}
			rs1535453		0.12	9p24.1	intron variant <i>PLGRKT</i>	3.04 / 1.46	1.84×10^{-4} / 2.83×10^{-2}
			rs4559327		0.33	9p13.1	intergenic variant	x / 1.37	x / 4.16×10^{-5}

Study	Population	Definition of Aggressive PCa	SNP	Cases/Controls	MAF	Locus	Consequence and Gene	OR95%	P-value
			rs10123553		0.34	9q22.2	intergenic variant	0.68 / x	7.57×10^{-5} / x
			rs1335540		0.31	10p12.1	5 prime UTR variant <i>APBB1IP</i>	0.66 / 0.84	1.85×10^{-4} / 1.84×10^{-2}
			rs7913838		0.04	10p11.21	intergenic variant	1.97 / 1.53	3.28×10^{-3} / 1.82×10^{-3}
			rs4576883		0.17	12q24.23	intron variant <i>AC084880.2</i>	0.65 / 0.70	4.88×10^{-3} / 1.06×10^{-3}
			rs7327461		0.33	13q14.11	intron variant <i>FOXO1 and LINC00598</i>	1.63 / 1.28	5.20×10^{-4} / 1.36×10^{-2}
			rs2332389		0.29	14q24.2	intron variant <i>AL357153.2</i>	x / 1.63	x / 6.61×10^{-6}
			rs17127296		0.08	14q32.11	intron variant <i>CCDC88C</i>	1.70 / 1.77	3.0×10^{-2} / 2.45×10^{-4}
			rs743672		0.15	16p13.12	intergenic variant	2.05 / 1.77	7.91×10^{-3} / 7.28×10^{-5}
			rs7271186		0.07	20q12	intergenic variant	1.82 / 1.52	2.35×10^{-3} / 1.30×10^{-3}
			rs6586243		0.19	21q22.3	intron variant <i>PRDM15</i>	0.54 / 0.66	9.45×10^{-4} / 5.32×10^{-4}
			rs7291691		0.36	22q13.1	intron variant <i>TNRC6B</i>	1.37 / 1.34	1.0×10^{-2} / 4.05×10^{-4}
			rs9623117	4755/7200	0.36	22q13.1	intron variant <i>TNRC6B</i>	1.18	4.96×10^{-7}

Study	Population	Definition of Aggressive PCa	SNP	Cases/Controls	MAF	Locus	Consequence and Gene	OR95%	P-value
Xu <i>et al.</i> 2010	European Ancestry (mainly from Sweden)	Stage T3/T4 or Gleason Score ≥ 7 , N+ or M+,	rs4054823	4829/12205	0.42	17p12	intergenic variant Closest annotated gene (~120 kb telomeric) is HS3ST3A	1.13	2.1×10^{-6}
FitzGerald <i>et al.</i> 2011	Caucasian American (Washington residents)	Gleason score ≥ 7 (4+3) or regional or distant stage or diagnostic PSA value of 50 ng/mL or more or evidence of progression/recurrence; or prostate cancer-specific death	rs6497287	202/100 (527/1167 in validation)	0.17	15q13.1	intron variant <i>HERC2</i>	8.92 (1.46 in validation)	5.20×10^{-5} (0.004 in validation)
			rs3774315		0.18	3q26.31	intron variant <i>TNFSF10</i>	2.50	7.34×10^{-5}
			rs8015211		0.26	14q22.3	intron variant <i>ATG14</i>	2.25	1.64×10^{-5}
			rs12433615		0.34	14q22.3	intergenic variant	2.16	4.06×10^{-5}
			rs1188069		0.30	14q22.3	intron variant <i>KTN1-AS1</i>	2.52	6.38×10^{-5}
			rs2341883		0.50	14q22.3	intron variant <i>KTN1-AS1</i>	2.06	1.01×10^{-4}
			rs11150069		0.47	16q23.1	intron variant <i>WVOX</i>	0.42	9.43×10^{-6}
			rs4628973		0.43	16q24.3	intron variant <i>URAHP</i>	0.35	4.62×10^{-7}
			rs225061		0.10	Xp21.3	intron variant <i>IL1RAPL1</i>	0.59	9.71×10^{-4}
Nam <i>et al.</i> 2011	Caucasian North Americans	D'Amico Risk Criteria - High Risk	rs12699509	316/229 (1663/1776 in validation)	0.29	7p21.3	intron variant <i>AC011287.1</i>	2.7	6.0×10^{-3}
			rs3114316		0.12	7q11.23	intron variant <i>CCDC146</i> and <i>LOC102723791</i>	3.4	4.0×10^{-3}
			rs3852402		0.38	9q23.32	intron variant <i>FANCC</i>	2.2	1.0×10^{-2}
			rs2226016		0.34	11p14.1	intergenic variant	2.3	1.0×10^{-2}
			rs4281668		0.42	15q25.3	intron variant <i>AKAP13</i>	2.9	4.0×10^{-3}
			rs17018760		0.11	2p12	intron variant <i>CTNNA2</i>	4.4	1.7×10^{-6}
			rs410259		0.12	2p22.1	intron variant <i>SLC8A</i>	4.0	1.0×10^{-5}

Study	Population	Definition of Aggressive PCa	SNP	Cases/Controls	MAF	Locus	Consequence and Gene	OR95%	P-value
			rs10749408		0.34	10q26.12	intergenic variant	2.5/ 1.18 (in validation)	1.0×10^{-5} / 2.0×10^{-2} (in validation)
			rs2429763		0.18	11p14.3	intron variant <i>AC006299.1</i>	2.9	1.0×10^{-5}
			rs595018		0.19	11q12.2	intergenic variant	2.9	1.0×10^{-5}
			rs2347306		0.05	12q24.32	intron variant <i>AC006065.3</i> and <i>AC006065.4</i>	5.5	1.0×10^{-5}
			rs11199874		0.20	10q26.12	intergenic variant	1.27 (in validation)	1.0×10^{-3} (in validation)
			rs10788165		0.31	10q26.12	regulatory region variant (located in a TF binding site)	1.29 (in validation)	5.0×10^{-4} (in validation)
			rs4775302		0.41	15q21.1	intergenic variant	1.31 (in validation)	1.0×10^{-3} (in validation)
			rs1994198		0.39	15q21.1	intergenic variant	1.33 (in validation)	9.0×10^{-4} (in validation)
Schumacher <i>et al.</i> 2011	European Ancestry	Gleason Score ≥ 8 or Stage C/D	rs2292884	2782/4458	0.37	2q37.3	missense variant <i>MLPH</i>	1.14	7.8×10^{-6}
			rs902774		0.08	12q13.13	regulatory region variant (located in a TF binding site) Closest gene (7 kb telomeric) is <i>KRT8</i>	1.21	1.4×10^{-8}
Olama <i>et al.</i> 2012 (Meta-analysis)	Mostly European Ancestry, also Asian and African-American	Gleason Score ≥ 8	rs11672691	5953/11463	0.48	19q13.2	intron variant <i>PCAT19</i>	1.12	1.4×10^{-8}
			rs11704416		0.20	22q13.1	3 prime UTR variant <i>FAM83F</i>	0.94	4.0×10^{-6}
Berndt <i>et al.</i> 2015	European Ancestry	Gleason Score ≥ 8	rs35148638	1638/11060	0.22	5q14.3	intron variant <i>RASA1</i> and <i>CCNH</i>	1.18	8.85×10^{-5}

Study	Population	Definition of Aggressive PCa	SNP	Cases/Controls	MAF	Locus	Consequence and Gene	OR95%	P-value
			rs78943174	12485 (case only analysis)	<0.01	3q26.31	intron variant <i>NAALADL2</i>	1.34	4.18x10 ⁻⁸
			rs62113212		0.03	19q13.33	Intron variant <i>KLK3</i>	1.15	5.85x10 ⁻⁹
			rs4054823	423/2941 (validation)	0.42	17p12	intergenic variant	0.83	1.0x10 ⁻²
Vaidyanathan <i>et al.</i> 2017 *	New Zealand Population	Clinical T stage ≥cT2c, and/or the Gleason score ≥8, and/or the serum prostate serum antigen (PSA) level >20ng/ml	Interaction between rs2121875 and rs4809960	254/369	0.47 and 0.19	5p12 and 20q13	<i>FGF10</i> (intron variant) (rs2121875) and <i>CYP24A1</i> (intron variant) (rs4809960)	2.918 (int)	7.15x10 ⁻⁵ (int)

The results shown in this table are from the discovery phase, the combined phases results of each study (when applicable) or the validation phases results (when conducted). Minor allele frequency values for all populations obtained on "Ensembl" database. CAPS: Cancer of the Prostate in Sweden study; CGEMS: The Cancer Genetic Markers of Susceptibility prostate cancer GWAS; int: interaction; X: no data available.

*This variant was firstly discovered through GWAS as being associated with overall PCa in a study by Kote-Jarai *et al.* [123].

Appendix 2

A paper titled “Aggressive prostate cancer phenotype and genome-wide association studies: where are we now?” has been accepted for publication in the scientific journal Pharmacogenomics.

8. References

1. Bray, F., et al., 2018, *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. CA: a cancer journal for clinicians. **68**(6): p. 394-424.
2. Center, M.M., et al., 2012, *International variation in prostate cancer incidence and mortality rates*. European urology. **61**(6): p. 1079-1092.
3. Ferlay, J., et al., 2018, *Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018*. European Journal of Cancer. **103**: p. 356-387.
4. Ferlay, J., et al., 2015, *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012*. International journal of cancer. **136**(5): p. E359-E386.
5. Attard, G., et al., 2016, *Prostate cancer*. Lancet. **387**(10013): p. 70-82.
6. Wong, M.C., et al., 2016, *Global incidence and mortality for prostate cancer: analysis of temporal patterns and trends in 36 countries*. European urology. **70**(5): p. 862-874.
7. Perdana, N.R., et al., 2016, *The risk factors of prostate cancer and its prevention: a literature review*. Acta Med Indones. **48**(3): p. 228-238.
8. Lilja, H., 1985, *A kallikrein-like serine protease in prostatic fluid cleaves the predominant seminal vesicle protein*. The Journal of clinical investigation. **76**(5): p. 1899-1903.
9. Catalona, W.J., et al., 1991, *Measurement of prostate-specific antigen in serum as a screening test for prostate cancer*. New England Journal of Medicine. **324**(17): p. 1156-1161.
10. Pezaro, C., H.H. Woo, and I.D. Davis, 2014, *Prostate cancer: measuring PSA*. Internal medicine journal. **44**(5): p. 433-440.
11. Litwin, M.S. and H.-J. Tan, 2017, *The diagnosis and treatment of prostate cancer: a review*. Jama. **317**(24): p. 2532-2542.
12. Calais da Silva, F. and G.P. Génito-Urinário, 2016, *Tumores malignos : tracto génito-urinário*, 1 editor, editor. Calais da Silva, Fernando, Génito-Urinário, Grupo Português, and Oncologia, Sociedade Portuguesa de. Vol. 2.
13. Jansen, F.H., et al., 2010, *Prostate-specific antigen (PSA) isoform p2PSA in combination with total PSA and free PSA improves diagnostic accuracy in prostate cancer detection*. European urology. **57**(6): p. 921-927.
14. Catalona, W.J., et al., 1998, *Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: a prospective multicenter clinical trial*. Jama. **279**(19): p. 1542-1547.

15. Sokoll, L.J., et al., 2010, *A prospective, multicenter, NCI EDNRN study of [-2] proPSA: improving prostate cancer detection and correlating with cancer aggressiveness*. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. **19**(5): p. 1193.
16. Bussemakers, M.J., et al., 1999, *Dd3:: A new prostate-specific gene, highly overexpressed in prostate cancer*. Cancer research. **59**(23): p. 5975-5979.
17. Lee, C.H., O. Akin-Olugbade, and A. Kirschenbaum, 2011, *Overview of prostate anatomy, histology, and pathology*. Endocrinology and Metabolism Clinics. **40**(3): p. 565-575.
18. McNeal, J.E., et al., 1988, *Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread*. The American journal of surgical pathology. **12**(12): p. 897-906.
19. Eble, J., et al., 2004, *WHO classification of tumors of the urinary system and male genital organs*. Lyon: IRAC.
20. Humphrey, P.A., 2012, *Histological variants of prostatic carcinoma and their significance*. Histopathology. **60**(1): p. 59-74.
21. Stark, J.R., et al., 2009, *Gleason score and lethal prostate cancer: does 3+ 4= 4+ 3?* Journal of Clinical Oncology. **27**(21): p. 3459.
22. Pierorazio, P.M., et al., 2013, *Prognostic Gleason grade grouping: data based on the modified Gleason scoring system*. BJU international. **111**(5): p. 753-760.
23. Epstein, J.I., et al., 2016, *The 2014 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma*. The American journal of surgical pathology. **40**(2): p. 244-252.
24. Mottet, N., et al., 2017, *EAU-ESTRO-ESUR-SIOG guidelines on prostate cancer*. European Association of Urology. p. 146.
25. Brierley, J.D., M.K. Gospodarowicz, and C. Wittekind, 2016, *TNM classification of malignant tumours*. John Wiley & Sons.
26. Facchini, G., et al., 2018, *Optimal management of prostate cancer based on its natural clinical history*. Current cancer drug targets. **18**(5): p. 457-467.
27. Barry, M.J. and L.H. Simmons, 2017, *Prevention of prostate cancer morbidity and mortality: primary prevention and early detection*. Medical Clinics. **101**(4): p. 787-806.
28. Zaorsky, N.G., et al., 2017, *The evolution of brachytherapy for prostate cancer*. Nature Reviews Urology. **14**(7): p. 415.
29. Schröder, F., et al., 2012, *Androgen deprivation therapy: past, present and future*. BJU international. **109**: p. 1-12.

30. Dasgupta, S., S. Srinidhi, and J.K. Vishwanatha, 2012, *Oncogenic activation in prostate cancer progression and metastasis: Molecular insights and future challenges*. Journal of carcinogenesis. **11**.
31. Chang, A.J., et al., 2014, *High-risk prostate cancer—classification and therapy*. Nature reviews Clinical oncology. **11**(6): p. 308.
32. D'amico, A.V., et al., 1998, *Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer*. Jama. **280**(11): p. 969-974.
33. Huang, J., et al., 2012, *Percentage of positive biopsy cores: a better risk stratification model for prostate cancer?* International Journal of Radiation Oncology* Biology* Physics. **83**(4): p. 1141-1148.
34. Behre, H.M. and S. Nieschlag, 2012, *Testosterone*. Springer.
35. Tilbrook, A. and I.J. Clarke, 2001, *Negative feedback regulation of the secretion and actions of gonadotropin-releasing hormone in males*. Biology of reproduction. **64**(3): p. 735-742.
36. Carson III, C. and R. Rittmaster, 2003, *The role of dihydrotestosterone in benign prostatic hyperplasia*. Urology. **61**(4): p. 2-7.
37. Lamb, A.D. and D.E. Neal, 2013, *Role of the androgen receptor in prostate cancer*. Trends in Urology & Men's Health. **4**(3): p. 26-30.
38. Denmeade, S.R., X.S. Lin, and J.T. Isaacs, 1996, *Role of programmed (apoptotic) cell death during the progression and therapy for prostate cancer*. The Prostate. **28**(4): p. 251-265.
39. Pagliarulo, V., et al., 2012, *Contemporary role of androgen deprivation therapy for prostate cancer*. European urology. **61**(1): p. 11-25.
40. Vale, C.L., et al., 2016, *Addition of docetaxel or bisphosphonates to standard of care in men with localised or metastatic, hormone-sensitive prostate cancer: a systematic review and meta-analyses of aggregate data*. The lancet oncology. **17**(2): p. 243-256.
41. Ritch, C. and M. Cookson, 2018, *Recent trends in the management of advanced prostate cancer*. F1000Research. **7**.
42. Sharifi, N., J.L. Gulley, and W.L. Dahut, 2005, *Androgen deprivation therapy for prostate cancer*. Jama. **294**(2): p. 238-244.
43. Sen, A., et al., 2012, *Paxillin mediates extranuclear and intranuclear signaling in prostate cancer proliferation*. The Journal of clinical investigation. **122**(7): p. 2469-2481.

44. Cornford, P., et al., 2017, *EAU-ESTRO-SIOG guidelines on prostate cancer. Part II: treatment of relapsing, metastatic, and castration-resistant prostate cancer*. European urology. **71**(4): p. 630-642.
45. Feldman, B.J. and D. Feldman, 2001, *The development of androgen-independent prostate cancer*. Nature Reviews Cancer. **1**(1): p. 34.
46. Chandrasekar, T., et al., 2015, *Mechanisms of resistance in castration-resistant prostate cancer (CRPC)*. Translational andrology and urology. **4**(3): p. 365.
47. Benafif, S. and R. Eeles, 2016, *Genetic predisposition to prostate cancer*. British medical bulletin. **120**(1): p. 75-89.
48. Ford, D., et al., 1994, *Risks of cancer in BRCA1-mutation carriers*. The Lancet. **343**(8899): p. 692-695.
49. Consortium, B.C.L., 1999, *Cancer risks in BRCA2 mutation carriers*. Journal of the National Cancer Institute. **91**(15): p. 1310-1316.
50. Castro, E., C. Goh, and D. Olmos, 2013, *Germline BRCA mutations are associated with higher risk of nodal involvement, distant metastasis, and poor survival outcomes in prostate cancer*. Journal of Clinical Oncology. **31**(14): p. 1748.
51. Pritchard, C.C., et al., 2016, *Inherited DNA-repair gene mutations in men with metastatic prostate cancer*. New England Journal of Medicine. **375**(5): p. 443-453.
52. Grönberg, H., et al., 1999, *In Swedish families with hereditary prostate cancer, linkage to the HPC1 locus on chromosome 1q24-25 is restricted to families with early-onset prostate cancer*. The American Journal of Human Genetics. **65**(1): p. 134-140.
53. Ewing, C.M., et al., 2012, *Germline mutations in HOXB13 and prostate-cancer risk*. New England Journal of Medicine. **366**(2): p. 141-149.
54. Easton, D., et al., 1997, *Cancer risks in two large breast cancer families linked to BRCA2 on chromosome 13q12-13*. American journal of human genetics. **61**(1): p. 120.
55. Leongamornlert, D., et al., 2012, *Germline BRCA1 mutations increase prostate cancer risk*. British journal of cancer. **106**(10): p. 1697.
56. Leongamornlert, D., et al., 2014, *Frequent germline deleterious mutations in DNA repair genes in familial prostate cancer cases are associated with advanced disease*. British journal of cancer. **110**(6): p. 1663.
57. Raymond, V.M., et al., 2013, *Elevated risk of prostate cancer among men with Lynch syndrome*. Journal of Clinical Oncology. **31**(14): p. 1713.
58. Cybulski, C., et al., 2013, *An inherited NBN mutation is associated with poor prognosis prostate cancer*. British journal of cancer. **108**(2): p. 461.

59. Tomlins, S.A., et al., 2005, *Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer*. science. **310**(5748): p. 644-648.
60. Giri, V.N. and J.L. Beebe-Dimmer. *Familial prostate cancer*. in *Seminars in oncology*. 2016. Elsevier.
61. Robinson, D., et al., 2015, *Integrative clinical genomics of advanced prostate cancer*. Cell. **161**(5): p. 1215-1228.
62. Grasso, C.S., et al., 2012, *The mutational landscape of lethal castration-resistant prostate cancer*. Nature. **487**(7406): p. 239.
63. Wu, X. and J. Gu, 2016, *Heritability of prostate cancer: a tale of rare variants and common single nucleotide polymorphisms*. Annals of translational medicine. **4**(10).
64. Group, I.S.M.W., 2001, *A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms*. Nature. **409**(6822): p. 928.
65. Dias, A., et al., 2018, *Prostate cancer germline variations and implications for screening and treatment*. Cold Spring Harbor perspectives in medicine. **8**(9): p. a030379.
66. Brookes, A.J., 1999, *The essence of SNPs*. Gene. **234**(2): p. 177-186.
67. Collins, F.S., L.D. Brooks, and A. Chakravarti, 1998, *A DNA polymorphism discovery resource for research on human genetic variation*. Genome research. **8**(12): p. 1229-1231.
68. Schork, N.J., et al., 2009, *Common vs. rare allele hypotheses for complex diseases*. Current opinion in genetics & development. **19**(3): p. 212-219.
69. Hirschhorn, J.N. and M.J. Daly, 2005, *Genome-wide association studies for common diseases and complex traits*. Nature Reviews Genetics. **6**(2): p. 95.
70. Bush, W.S. and J.H. Moore, 2012, *Genome-wide association studies*. PLoS computational biology. **8**(12): p. e1002822.
71. Benafif, S., Z. Kote-Jarai, and R.A. Eeles, 2018, *A Review of Prostate Cancer Genome-Wide Association Studies (GWAS)*. Cancer Epidemiology and Prevention Biomarkers. **27**(8): p. 845-857.
72. Sud, A., B. Kinnersley, and R.S. Houlston, 2017, *Genome-wide association studies of cancer: current insights and future perspectives*. Nature Reviews Cancer. **17**(11): p. 692.
73. Ahmed, M. and R. Eeles, 2016, *Germline genetic profiling in prostate cancer: latest developments and potential clinical applications*. Future science OA. **2**(1).
74. Manolio, T.A., 2010, *Genomewide association studies and assessment of the risk of disease*. New England journal of medicine. **363**(2): p. 166-176.

75. Hindorff, L.A., et al., 2009, *Potential etiologic and functional implications of genome-wide association loci for human diseases and traits*. Proceedings of the National Academy of Sciences. **106**(23): p. 9362-9367.
76. Mueller, J.C., et al., 2005, *Linkage disequilibrium patterns and tagSNP transferability among European populations*. The American Journal of Human Genetics. **76**(3): p. 387-398.
77. Johnson, G.C., et al., 2001, *Haplotype tagging for the identification of common disease genes*. Nature genetics. **29**(2): p. 233.
78. Manolio, T.A., et al., 2009, *Finding the missing heritability of complex diseases*. Nature. **461**(7265): p. 747.
79. Ku, C.S., et al., 2010, *The pursuit of genome-wide association studies: where are we now?* Journal of human genetics. **55**(4): p. 195.
80. McCarthy, M.I. and J.N. Hirschhorn, 2008, *Genome-wide association studies: potential next steps on a genetic journey*. Human molecular genetics. **17**(R2): p. R156-R165.
81. Yang, Q., et al. *Power and type I error rate of false discovery rate approaches in genome-wide association studies*. in *BMC genetics*. 2005. BioMed Central.
82. Risch, N. and K. Merikangas, 1996, *The future of genetic studies of complex human diseases*. Science. **273**(5281): p. 1516-1517.
83. Stringer, S., et al., 2011, *Underestimated effect sizes in GWAS: fundamental limitations of single SNP analysis for dichotomous phenotypes*. PloS one. **6**(11): p. e27964.
84. Park, J.-H., et al., 2010, *Estimation of effect size distribution from genome-wide association studies and implications for future discoveries*. Nature genetics. **42**(7): p. 570.
85. Kote-Jarai, Z., et al., 2008, *Multiple novel prostate cancer predisposition loci confirmed by an international study: the PRACTICAL Consortium*. Cancer Epidemiology and Prevention Biomarkers. **17**(8): p. 2052-2061.
86. Schaid, D.J., B.L. Chang, and I.C.F.P.C. Genetics, 2005, *Description of the International Consortium For Prostate Cancer Genetics, and failure to replicate linkage of hereditary prostate cancer to 20q13*. The Prostate. **63**(3): p. 276-290.
87. Zöllner, S. and J.K. Pritchard, 2007, *Overcoming the winner's curse: estimating penetrance parameters from case-control data*. The American Journal of Human Genetics. **80**(4): p. 605-615.
88. Igl, B.-W., I.R. König, and A. Ziegler, 2009, *What do we mean by 'replication' and 'validation' in genome-wide association studies?* Human heredity. **67**(1): p. 66-68.

89. König, I.R., 2011, *Validation in genetic association studies*. Briefings in bioinformatics. **12**(3): p. 253-258.
90. Walsh, P.C., 2017, *The search for the missing heritability of prostate cancer*. European urology. **72**(5): p. 657-659.
91. Gudmundsson, J., et al., 2007, *Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24*. Nature genetics. **39**(5): p. 631.
92. Yeager, M., et al., 2007, *Genome-wide association study of prostate cancer identifies a second risk locus at 8q24*. Nature genetics. **39**(5): p. 645.
93. Haiman, C.A., et al., 2007, *Multiple regions within 8q24 independently affect risk for prostate cancer*. Nature genetics. **39**(5): p. 638.
94. Buniello, A., et al., 2018, *The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019*. Nucleic acids research. **47**(D1): p. D1005-D1012.
95. Gallagher, D.J., et al., 2010, *Susceptibility loci associated with prostate cancer progression and mortality*. Clinical Cancer Research. p. 1078-0432. CCR-10-0028.
96. Thomas, G., et al., 2008, *Multiple loci identified in a genome-wide association study of prostate cancer*. Nature genetics. **40**(3): p. 310.
97. Sung, Y., et al., 2018, *Jazf1 promotes prostate cancer progression by activating JNK/Slug*. Oncotarget. **9**(1): p. 755.
98. Xie, Y., et al., 2014, *Slug regulates E-cadherin repression via p19Arf in prostate tumorigenesis*. Molecular oncology. **8**(7): p. 1355-1364.
99. Duggan, D., et al., 2007, *Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP*. Journal of the National Cancer Institute. **99**(24): p. 1836-1844.
100. Sun, J., et al., 2009, *Sequence variants at 22q13 are associated with prostate cancer risk*. Cancer research. **69**(1): p. 10-15.
101. Xu, J., et al., 2010, *Inherited genetic variant predisposes to aggressive but not indolent prostate cancer*. Proceedings of the National Academy of Sciences. **107**(5): p. 2136-2140.
102. FitzGerald, L.M., et al., 2011, *Genome-wide association study identifies a genetic variant associated with risk for more aggressive prostate cancer*. Cancer Epidemiology and Prevention Biomarkers.
103. Nam, R.K., et al., 2011, *New variants at 10q26 and 15q21 are associated with aggressive prostate cancer in a genome-wide association study from a prostate biopsy screening cohort*. Cancer biology & therapy. **12**(11): p. 997-1004.

104. Schumacher, F.R., et al., 2011, *Genome-wide association study identifies new prostate cancer susceptibility loci*. Human molecular genetics. **20**(19): p. 3867-3875.
105. Amin Al Olama, A., et al., 2012, *A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease*. Human molecular genetics. **22**(2): p. 408-415.
106. Berndt, S.I., et al., 2015, *Two susceptibility loci identified for prostate cancer aggressiveness*. Nature communications. **6**: p. 6889.
107. Zheng, S.L., et al., 2004, *Sequence variants of toll-like receptor 4 are associated with prostate cancer risk: results from the CAnceR Prostate in Sweden Study*. Cancer research. **64**(8): p. 2918-2922.
108. Chen, H., et al., 2003, *Epigenetic regulation of a novel tumor suppressor gene (hDAB2IP) in prostate cancer cell lines*. Journal of Biological Chemistry. **278**(5): p. 3121-3130.
109. Gao, P., et al., 2018, *Biology and clinical implications of the 19q13 aggressive prostate cancer susceptibility locus*. Cell. **174**(3): p. 576-589. e18.
110. Baillat, D. and R. Shiekhataar, 2009, *Functional dissection of the human TNRC6 (GW182-related) family of proteins*. Molecular and cellular biology. **29**(15): p. 4144-4155.
111. Stelzer, G., et al., 2016, *The GeneCards suite: from gene data mining to disease genome sequence analyses*. Current protocols in bioinformatics. **54**(1): p. 1.30. 1-1.30. 33.
112. Epstein, J.I., et al., 2016, *A contemporary prostate cancer grading system: a validated alternative to the Gleason score*. European urology. **69**(3): p. 428-435.
113. Farashi, S., et al., 2018, *Post-GWAS in prostate cancer: from genetic association to biological contribution*. Nature Reviews Cancer. p. 1.
114. McCarthy, M.I., et al., 2008, *Genome-wide association studies for complex traits: consensus, uncertainty and challenges*. Nature reviews genetics. **9**(5): p. 356.
115. Scher, H.I., et al., 2008, *Design and end points of clinical trials for patients with progressive prostate cancer and castrate levels of testosterone: recommendations of the Prostate Cancer Clinical Trials Working Group*. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. **26**(7): p. 1148.
116. Heidenreich, A., et al., 2014, *EAU guidelines on prostate cancer. Part II: treatment of advanced, relapsing, and castration-resistant prostate cancer*. European urology. **65**(2): p. 467-479.

117. Penney, K.L., et al., 2015, *Association of prostate cancer risk variants with gene expression in normal and tumor tissue*. Cancer Epidemiology and Prevention Biomarkers. **24**(1): p. 255-260.
118. Bu, H., et al., 2016, *Putative Prostate Cancer Risk SNP in an Androgen Receptor-Binding Site of the Melanophilin Gene Illustrates Enrichment of Risk SNPs in Androgen Receptor Target Sites*. Human mutation. **37**(1): p. 52-64.
119. Vaidyanathan, V., et al., 2017, *SNP-SNP interactions as risk factors for aggressive prostate cancer*. F1000Research. **6**.
120. Woodward, J., 2014, *Bi-allelic SNP genotyping using the TaqMan® assay*, in *Crop Breeding*. Springer. p. 67-74.
121. Tak, Y.G. and P.J. Farnham, 2015, *Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome*. Epigenetics & chromatin. **8**(1): p. 57.
122. Maney, D.L., 2017, *Polymorphisms in sex steroid receptors: from gene sequence to behavior*. Frontiers in neuroendocrinology. **47**: p. 47-65.
123. Kote-Jarai, Z., et al., 2011, *Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study*. Nature genetics. **43**(8): p. 785.
124. Sung, H., et al., 2016, *Inactivation of RASA1 promotes melanoma tumorigenesis via R-Ras activation*. Oncotarget. **7**(17): p. 23885.
125. Zhang, H., et al., 2011, *FOXO1 inhibits Runx2 transcriptional activity and prostate cancer cell migration and invasion*. Cancer research. **71**(9): p. 3257-3267.
126. Hernández-Varas, P., et al., 2011, *Rap1-GTP-interacting adaptor molecule (RIAM) protein controls invasion and growth of melanoma cells*. Journal of Biological Chemistry. **286**(21): p. 18492-18504.
127. Romanuik, T.L., et al., 2010, *LNCaP Atlas: gene expression associated with in vivo progression to castration-recurrent prostate cancer*. BMC medical genomics. **3**(1): p. 43.
128. Memarzadeh, S., et al., 2007, *Enhanced paracrine FGF10 expression promotes formation of multifocal prostate adenocarcinoma and an increase in epithelial androgen receptor*. Cancer cell. **12**(6): p. 572-585.
129. Shastry, B.S., 2009, *SNPs: impact on gene function and phenotype*, in *Single Nucleotide Polymorphisms*. Springer. p. 3-22.
130. McSherry, E.A., et al., 2011, *Breast cancer cell migration is regulated through junctional adhesion molecule-A-mediated activation of Rap1 GTPase*. Breast Cancer Research. **13**(2): p. R31.
131. Bailey, C.L., P. Kelly, and P.J. Casey, 2009, *Activation of Rap1 promotes prostate cancer metastasis*. Cancer research. **69**(12): p. 4962-4968.

132. Patsoukis, N., et al., 2009, *RIAM regulates the cytoskeletal distribution and activation of PLC- γ 1 in T cells*. *Sci. Signal.* **2**(99): p. ra79-ra79.
133. Su, W., et al., 2015, *Rap1 and its effector RIAM are required for lymphocyte trafficking*. *Blood*. **126**(25): p. 2695-2703.
134. Klapproth, S., et al., 2015, *Loss of the Rap1 effector RIAM results in leukocyte adhesion deficiency due to impaired β 2 integrin function in mice*. *Blood*. **126**(25): p. 2704-2712.
135. Thakkar, A., et al., 2015, *High expression of three-gene signature improves prediction of relapse-free survival in estrogen receptor-positive and node-positive breast tumors*. *Biomarker insights*. **10**: p. BMI. S30559.
136. Mancuso, N., et al., 2018, *Large-scale transcriptome-wide association study identifies new prostate cancer risk regions*. *Nature communications*. **9**(1): p. 4079.
137. Ensembl. Available from: <http://www.ensembl.org/index.html>.
138. GeneCards. Available from: <https://www.genecards.org/>.
139. Taniguchi, F., et al., 2003, *Activation of mitogen-activated protein kinase pathway by keratinocyte growth factor or fibroblast growth factor-10 promotes cell proliferation in human endometrial carcinoma cells*. *The Journal of Clinical Endocrinology & Metabolism*. **88**(2): p. 773-780.
140. Thomson, A. and G. Cunha, 1999, *Prostatic growth and development are regulated by FGF10*. *Development*. **126**(16): p. 3693-3701.
141. Li, Q., et al., 2018, *Pharmacologically targeting the myristoylation of the scaffold protein FRS2 α inhibits FGF/FGFR-mediated oncogenic signaling and tumor progression*. *Journal of Biological Chemistry*. **293**(17): p. 6434-6448.
142. Memarzadeh, S., et al., 2011, *Role of autonomous androgen receptor signaling in prostate cancer initiation is dichotomous and depends on the oncogenic signal*. *Proceedings of the National Academy of Sciences*. **108**(19): p. 7962-7967.
143. Yu, S., et al., 2013, *Androgen receptor in human prostate cancer-associated fibroblasts promotes prostate cancer epithelial cell growth and invasion*. *Medical oncology*. **30**(3): p. 674.
144. Ropiquet, F., et al., 2000, *FGF-10 is expressed at low levels in the human prostate*. *The Prostate*. **44**(4): p. 334-338.
145. Bonanno, L., et al., 2016, *Combinatory effect of BRCA1 and HERC2 expression on outcome in advanced non-small-cell lung cancer*. *BMC cancer*. **16**(1): p. 312.
146. Bekker-Jensen, S., et al., 2010, *HERC2 coordinates ubiquitin-dependent assembly of DNA repair factors on damaged chromosomes*. *Nature cell biology*. **12**(1): p. 80.

147. Brzovic, P.S., et al., 2003, *Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex*. Proceedings of the National Academy of Sciences. **100**(10): p. 5646-5651.
148. Cubillos-Rojas, M., et al., 2014, *The E3 ubiquitin protein ligase HERC2 modulates the activity of tumor protein p53 by regulating its oligomerization*. Journal of Biological Chemistry. **289**(21): p. 14782-14795.
149. Lee, D.K., H.O. Duan, and C. Chang, 2000, *From Androgen Receptor to the General Transcription Factor TFIID IDENTIFICATION OF cdk ACTIVATING KINASE (CAK) AS AN ANDROGEN RECEPTOR NH2-TERMINAL ASSOCIATED COACTIVATOR*. Journal of Biological Chemistry. **275**(13): p. 9308-9313.
150. Keriell, A., et al., 2002, *XPD mutations prevent TFIID-dependent transactivation by nuclear receptors and phosphorylation of RAR α* . Cell. **109**(1): p. 125-135.
151. Ko, L.J., et al., 1997, *p53 is phosphorylated by CDK7-cyclin H in a p36MAT1-dependent manner*. Molecular and cellular biology. **17**(12): p. 7220-7229.
152. Eerola, I., et al., 2003, *Capillary malformation–arteriovenous malformation, a new clinical and genetic disorder caused by RASA1 mutations*. The American Journal of Human Genetics. **73**(6): p. 1240-1249.
153. Suárez-Cabrera, C., et al., 2017, *A transposon-based analysis reveals RASA1 is involved in triple-negative breast Cancer*. Cancer research. **77**(6): p. 1357-1368.
154. Chen, Y.-L., et al., 2017, *Down-regulation of RASA1 is associated with poor prognosis in human hepatocellular carcinoma*. Anticancer research. **37**(2): p. 781-785.
155. Liu, Y., et al., 2015, *Downregulation of Ras GTPase-activating protein 1 is associated with poor survival of breast invasive ductal carcinoma patients*. Oncology reports. **33**(1): p. 119-124.
156. Hornsveid, M., et al. *Re-evaluating the role of FOXOs in cancer*. in *Seminars in cancer biology*. 2018. Elsevier.
157. Yang, Y., et al., 2017, *Loss of FOXO1 cooperates with TMPRSS2–ERG overexpression to promote prostate tumorigenesis and cell invasion*. Cancer research. **77**(23): p. 6524-6537.
158. Zhao, Y., D.J. Tindall, and H. Huang, 2014, *Modulation of androgen receptor by FOXA1 and FOXO1 factors in prostate cancer*. International journal of biological sciences. **10**(6): p. 614.
159. Liu, P., et al., 2008, *A transcription-independent function of FOXO1 in inhibition of androgen-independent activation of the androgen receptor in prostate cancer cells*. Cancer research. **68**(24): p. 10290-10299.

160. Chua, C.-W., et al., 2009, *Suppression of androgen-independent prostate cancer cell aggressiveness by FTY720: validating Runx2 as a potential antimetastatic drug screening platform*. Clinical cancer research. **15**(13): p. 4322-4335.
161. Jeong, O.-S., et al., 2016, *Long noncoding RNA linc00598 regulates CCND2 transcription and modulates the G1 checkpoint*. Scientific reports. **6**: p. 32172.
162. Qin, J., et al., 2012, *The PSA-⁻/_{lo} prostate cancer cell population harbors self-renewing long-term tumor-propagating cells that resist castration*. Cell stem cell. **10**(5): p. 556-569.