جامعة سيدي محمد بن عبد الله بفاس
UNIVERSITE SIDI MOHAMED BEN ABDALLAH DE FES
كلية العلوم ظهر المهراز فاس
FACULTE DES SCIENCES DHAR EL MEHRAZ FES



N° d'Ordre : **58 /2022**

Formation Doctorale : Sciences & Techniques

Discipline : Biologie & Santé

Spécialité : Biologie Moléculaire & Bioinformatique

Laboratoire : Laboratoire de Biotechnologie, Environnement, Agro-alimentaire et santé

THESE DE DOCTORAT

Presented By

Chaimae SAMTAL

GENETIC AND GENOMIC STUDIES OF PROSTATE CANCER IN THE MOROCCAN POPULATION

Thesis presented and defended in Fez on November, 25th 2022 in front of the jury composed of:

Pr. Dalila BOUSTA	***Faculté Des Sciences Dhar El Mahras. Fès***	***President***
Pr Noureddine BOUKHATEM	***University Mohammed Premier, Oujda***	***Referee***
Pr. Khalil HAMMANI	***Polydisciplinary Faculty of Taza***	***Referee***
Pr. Nabil ISMAILI	***University Mohammed VI of Health Sciences, Casablanca***	***Referee***
Pr. Norman H. LEE	***George Washington University, USA***	***Examiner***
Pr. Karim OULDIM	***Faculty of medicine and Pharmacy and Dentistry, Fez***	***Examiner***
Pr. Laila BOUGUENOUCH	***Faculty of medicine and Pharmacy and Dentistry, Fez***	***Examiner***
Pr. Keith A. CRANDALL	***George Washington University, USA***	**Examiner**
Pr. Hassan GHAZAL	***National Center for Scientific and Technical Research, Rabat***	**Co-supervisor**
Pr. Hicham BEKKARI	***Faculté Des Sciences Dhar El Mahras. Fès***	**Supervisor**

Année universitaire : 2021-2022

FOREWARD AND SCIENTIFIC PRODUCTION

Place of wet laboratory & Bioinformatic analysis:

- Laboratory of Biotechnology, Environment, Agri-food and Health, Faculty of Sciences Dhar El Mahraz–Sidi Mohammed Ben Abdellah University, Fez, Morocco
- Laboratory of Medical Oncogenetics, Hospital University Hassan II Fez, Morocco
- Institute of Pasteur, Casablanca, Morocco
- National Center for Scientific and Technical Research, Rabat, Morocco
- Laboratory of Pharmacology & Physiology, George Washington University School of Medicine & Health Sciences, Washington DC, USA.
- Computational Biology Institute, George Washington University, USA
- National Institute of Cancer, NIH, USA

The thesis work was achieved in collaboration with:

- The biopsies were collected in generous corporation with Professor Ali Barki, urology service, University Hospital Mohammed VI, Oujda, Morocco
- The blood was collected in corporation with Professor Moulay Hassan Farih and Professor Touria Bouhafa, Urology department, university Hospital Hassan II, Fez, Morocco

SCIENTIFIC PRODUCTION

N	Title of publication	Authors	Journal	Year of publication/volume,pages	Indexed in	Impact factor
1	*A review of clinical*	*Fouzia Radouani, Lyndon Zass, Yosr Hamdi, Jorge da Rocha,*	*Personalized medicine*	*3 Mars 2020/ volume 17, N*	*Elsevier/*Scopus	*2.119*

	pharmacogenetics Studies in African populations	*Reem Sallam, SoniaAbdelhak, Samah Ahmed, Maryame Azzouzi1, Ichrak Benamri1, Alia Benkahla,Balkiss Bouhaouala-Zahar, Melek Chaouch, Haifa Jmel,RymKefi, Ayoub Ksouri,Judit Kumuthini, Phumlani Masilela, Collen Masimirembwa, Houcemeddine Othman, Sumir Panji, Lilia Romdhane, Chaimae Samtal,RaniaSibira,KaisGhedira, Faisal Fadlelmola, Samar Kamal Kassim, Nicola Mulder*		*2, pages (155-170)*		
2	*Human OMICs and Computational Biology Research in Africa: Current Challenges and Prospects*	*Hamdi Y, Zass L, Othman H, Radouani F, Allali I, Hanachi M, Okeke CJ, Chaouch M, Tendwa MB, Samtal C, Mohamed Sallam R, Alsayed N, Turkson M, Ahmed S, Benkahla A, Romdhane L, Souiai O, Tastan Bishop Ö, Ghedira K, Mohamed Fadlelmola F, Mulder N, Kamal Kassim S*	*OMICS*	*1 avril 2021, volume 25, N4, pages(213-233)*	*PubMed/MEDLINE, PubMed Central, Web of Science: Science, Citation Index, Scopus, Current Contents/Life Sciences…*	*3.978*
3	*Data Management Plans in the genomics research revolution of Africa: Challenges and recommendations*	*Fadlelmola FM, Zass L, Chaouch M, Samtal C, Ras V, Kumuthini J, Panji S, Mulder*	*Journal of Biomedical Informatics*	*8 Septembre 2021, volume 122,*	*Embase, PubMed/Medline, Scopus, Science Citation Index, Science Citation, Index Expanded…*	*8*
4	*H3ABioNet genomic medicine and microbiome data portals hackathon proceedings*	*Fadlelmola FM, Ghedira K, Hamdi Y, Hanachi M, Radouani F, Allali I, Kiran A, Zass L, Alsayed N, Fassatoui M, Samtal C, Ahmed S, Da Rocha J, Chaqsare S, Sallam RM, Chaouch M, Farahat M, Ssekagiri A, Parker Z, Adil M, Turkson M, Benchaalia A, Benkahla A, Panji S, Kassim S, Souiai O, Mulder N*	*Database (Oxford)*	*17 avril 2021, volume 2021, pages(1-13)*	*Journal Citation Reports /Science Edition, MEDLINE/PubMed, PubMed Central, Asian Science Citation Index, Chemical Abstracts, Directory of Open Access Journals (DOAJ), The Standard Periodical Directory*	*4.776*

5	*Performing post-genome-wide association study analysis: overview, challenges and recommendations*	*Adam, Y., Samtal, C., Brandenburg, J., Falola, O., & Adebiyi, E.*	*F1000 Research*	*4 Octobre 2022, volume 10, pages (1-22)*	*UGC CARE, Scopus, DOAJ, PubMed*	*2.297*
6	African Genomic Medicine Portal: A Web Portal forBiomedical Applications	Othman H, Zass L, da Rocha JEB, Radouani F, Samtal C, Benamri I, Kumuthini J, Fakim YJ, Hamdi Y, Mezzi N, Boujemaa M, Okeke CJ, Tendwa MB, Sanak K, Chaouch M, Panji S, Kefi R, Sallam RM, Ghoorah AW, Romdhane L, Kiran A, Meintjes AP, Maturure P, Jmel H, Ksouri A, Azzouzi M, Farahat MA, Ahmed S, Sibira R, Turkson MEE, Ssekagiri A, Parker Z, Fadlelmola FM, Ghedira K, Mulder N, Kamal Kassim S	Journal of Personalized Medicine	11 February 2022, volume 12, N 2	Scopus, SCIE (Web of Science), PubMed, PMC, Embase, and other databases.	3.508
7	SARS-CoV-2 virus genome structure and evolution	Chaimae SAMTAL, Nihal HABIB, Abdellah IDRISSI AZAMI	Moroccan Journal of Public Health	2020, Volume 1, N1	-	-
8	Review of Prostate Cancer Genomic Studies in Africa	Chaimae Samtal, Islam El Jaddaoui, Salsabil Hamdi, Laila Bouguenouch, Karim Ouldim, Chakib Nejjari, Hassan Ghazal, Hicham Bekkari	Frontiers in Genetics	Accepted on September 2022- published on 11 October 2022		4.37
9	Update on Prostate Cancer Epidemiology in Morocco	Chaimae Samtal, Laila Bouguenouch, Nabil Ismaili, Marwa El baldi, Badreddine El makhzen, Karima El rhazi, Mohammed Mzyiene, Moulay Hassan Farih, Karim Ouldim, Hassan Ghazal, Hicham Bekkari	Pending Review	Pending review		
10	Epigenetic Analysis of Circulating Tumor DNA in	Samtal et al.	In preparation			

	Prostate Cancer of Moroccan Patients					
11	TMPRSS2-ERG Gene Fusion in Prostate Cancer of Moroccan Patients	Samtal et al.	In Preparation			
12	Whole Exome Sequencing of Moroccan Prostate Cancer Tumors	Samtal et al.	In Preparation			

HONORS & FELLOWSHIPS

N	Name	Type	Purpose	Dates
1	Fulbright Joint Supervision Doctoral program	Scholarship	Visiting scholar at School of Medicine and Health Sciences, and the Computational Biology Institute, George Washington University, DC, USA.	30 August 2017- 2 January 2019
2	International Center for Theoretical Physics (ICTP), Trieste, Italy.	Travel Fellowship	To attend "The CODATA-RDA Research Data Science Advanced Workshops on Bio-informatics, Climate Data Sciences, Extreme Sources of Data and Internet of Things (IoT)/Big-Data Analytics"	19-23August 2019
3	H3ABioNet	Travel Fellowship	To attend the Precision Medicine Hackathon, Pasteur Institute, Tunisia,	12-18 April 2019
4	Elife community Ambassador		Elife community Ambassador program, 2019	2019/2020
5	National Human Genome Research Institute (NHGRI), at NIH, USA.	Travel Fellowship	Was accepted to attend the International Summit in Human Genetics and Genomics	30 August -30 September 2017
6	New York University Abu Dhabi, Abu Dhabi, UAE	Travel fellowship	To attend the Winter Institute in Statistical Genetics Abu Dhabi 2017	22- 26 January 2017
7	Institute of Infectious Disease and Molecular Medicine Workshop, University of Cape Town, South Africa	Travel Fellowship	To attend the Analysis of NGS data for Pharmacogenomics Workshop	4-6 April 2016
8	"My Thesis in 180 seconds_ Edition 2016" National Competition, Mai		Won the 2nd Prize and Audience Prize	March 2016

	2016, Rabat, Morocco.			
9	"My Thesis in 180 seconds_ Edition 2016" International Competition, September 2016, Rabat, Morocco		Won the Audience Prize	September 2016
10	University of the Witwatersrand, Johannesburg, South Africa	Travel Fellowship	Short term Pharmacogenetics and Bioinformatics Internship	1 November-24 December 2021
11	H3ABioNet		H3ABioNet Ambassador	2017
12	George Washington University, Washington DC, USA		Director of Academic support, GWISA (George Washington International Students Association)	January-November 2018

INTERNSHIPS

N	Host Institution	Timeframe	Purpose/Subject of the Internship	Program/Funding
1	George Washington University, Washington DC, USA	1 year and 4 Months	Conduct part of my thesis research work	Fulbright Joint Supervision Doctoral program
2	Sydney Brenner Institute for Molecular Bioscience (SBIMB), University of the Witwatersrand, Johannesburg, South Africa	2 Months	Short term Pharmacogenetics and Bioinformatics Internship	NA/Wits university

WORKSHOPS ORGANIZATION

N	Name	Nature of the event	Date
1	The First Pan African Workshop on Newborn Screening – National Center of Scientific and technological research, Rabat, Morocco	Workshop	12-14 June 2019
2	International colloquium: Vahvistus-Bio (Valorization of natural and synthetic biomolecules in novel drug-vector nanostructures) Faculty of Sciences, Fez, Morocco	Colloquium	April 2019
3	"Medical Information" Workshop, Rabat, Morocco	Workshop	11-12 July 2016
4	"e-Health and Telemedicine in Mediterranean Countries" Conference, Rabat, Morocco	Conference	13-15 July 2016
5	Colloquium Microbiome for Health II	Coloquium	29-31 October 2018

WORKSHOPS & CONFERENCES PARTICIPATION

N	Name	Nature of the event	Date
1	*Analysis of NGS Data for Pharmacogenomics*, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, South Africa	Workshop	April 2016
2	Pharmacogenomics &Precision Medicine, University of Cape Town, South Africa	Conference	April 2016
3	*European Human Genetic*, Barcelona, Spain	Conference	May 2016
4	*African Society of Human Genetics (AFSHG)*, Dakar, Senegal	Conference	May 2016
5	*Education for a Better Morocco*, American University of Leadership, Rabat, Morocco	Conference	June 2016
6	*Scientific Writing*: Faculty of Medicine and Pharmacy, Fez, Morocco	Workshop	June 2016
7	*Medical Information*, ENSP, Rabat, Morocco	Workshop	July 2016
8	" *Uro-Oncological Pathology Tutorial: A 3-day "meet the expert"* , Spanish National Cancer Research Centre (CNIO), Madrid, Spain	Conference	November 2016
9	*Genome Surgery and Gene Therapy,* Faculty of Medicine and Pharmacy, Rabat, Morocco	Workshop	December 2016
10	'Hackital –a hackathon in the Capitol', George Washington University, Washington DC, USA	Hackathon	November 2017
11	Precision Medicine Hackathon, Pasteur Institute, Tunisia	Hackathon	April 2019
12	The CODATA-RDA Research Data Science Advanced Workshops on Bio-informatics, Climate Data Sciences, Extreme Sources of Data and Internet of Things(IoT)/Big-Data Analytics" at the ICTP, Trieste, Italy	Workshop	19-23 August 2019
13	The 16th H3Africa Virtual Consortium Meeting	Meeting	7-10 September 2020
14	the Advances in Genome Biology and Technology (AGBT) meeting	Meeting	10-11 September 2020
15	*NCBI Molecular Biology Resources &Tools*, Rabat, Morocco	Workshop	October 2015
16	*Ethics of Research,* Faculty of Medicine and Pharmacy, Fez, Morocco	Workshop	December 2015
17	*ScienceDirect*, *Scopus*, *Mendeley*, Faculty Sciences, Fez, Morocco	Workshop	March 2016
18	the H3Africa Rare Diseases Working Group (RDWG)/ClinGen workshop	Workshop	3-5 February 2021
19	Data Management Plans Workshop	Workshop	7-11 September 2020
20	online Genomic Data Carpentry workshop	Workshop	16, 17, 23 and 24th of September 2020

21	Introduction to HTML & CSS, Washington DC, USA	Workshop	8 January 2018
22	Professional Development Workshop for Educators, Georgetown University, Washington DC, USA	Workshop	November 18, 2017
23	Online GitHub Workshop	Workshop	27 May 2020
24	virtual Data Carpentry Genomics workshop for introduction to the best practices in the introductory computational skills needed for genomic data management and analysis, @TUDelft	Workshop	September 16-24, 2020
25	RedCap 101 and 102, George Washington University, Washington DC, USA	Workshop	December 2017
26	25th Congres SMC, Marrakech	Oral Communication	28,29 April 2017
27	'La chirurgie du Génome & Thérapie Génique', Rabat, Morocco	Journée Internationale de Biotechnologie Médicale	20 December 2016

TRAINING AND COURSES

N	**Host Organization/ Institution**	**Name**	**Nature of the event**	**Date/ Timeframe**
1	H3ABioNet/Wellcome connecting science	Next Generation Sequencing Bioinformatics Training	Training	24 April 2017 – 4 August 2017 (Total of 96 hours)
2	New York University Abu Dhabi, UAE	Abu Dhabi Winter Institute in Statistical Genetics	Winter School	January 22-26, 2017
3	H3ABioNet/ Oregon Health & Science University, USA	Health informatics course	Course	1 April 2020 to 16 July 2020 (Total of 80 hours)
4	The carpentries project	The Carpentries Instructor Training	Training	28 June – 2 July 2021
5	Carpentry training	Bonus Module for Online Teaching during Carpentry Connect South Africa	Training	14 September 2021
6	The Arctic University of Norway	CodeRefinery, Online course	course	Oct 20-22, 27-29 October 2020 (Total of 18 hours)
7	University of Gothenburg Sweden	REPRODUCIBILITY IN SCIENTIFIC RESEARCH (Online)	course	18- 22 January 2021
8	George Washington University, Washington DC, USA	Biosafety and bloodborne pathogen training	Training	2018
9	George Washington University, Washington DC, USA	the American Association for Laboratory Animal Science training and rodent handling techniques,	Training	2018
10	Collaborative Institutional Training	Biomedical Investigators & Human Research	course	2018

	Initiative (CITI Program), George Washington University, Washington DC, USA			
11	Collaborative Institutional Training Initiative (CITI Program), George Washington University, Washington DC, USA	Social & Behavioral Research & Human Research	course	2018
12	Wellcome connecting science	TRAIN THE TRAINER: DESIGN GENOMICS AND BIOINFORMATICS TRAINING	Course	3 weeks, 5 hours per week, December 2021
13	George Washington University, Washington DC, USA	The Basic Science of Oncology	Course	For one semester in 2018
14	George Washington University, Washington DC, USA	Genomics, Proteomics, and Bioinformatics	Course	For one semester in 2018
15	George Washington University, Washington DC, USA	Introduction to Bioinformatics	Course	For one semester in 2018
16	George Washington University, Washington DC, USA	Next Generation Technologies—Principles and Applications	Course	For one semester in 2018
17	George Washington University, Washington DC, USA	Public Health Genomics	Course	For one semester in 2018
18	George Washington University, Washington DC, USA	Molecular Oncology	Course	For one semester in 2018

TEACHING EXPERIENCE

N	**Host organization/institution**	**Name of the event (course/workshop)**	**Timeframe**	**Role**
1	H3ABioNet IBT course 2017	Introduction to Bioinformatics (IBT)	3 Months	Teaching Assistant
2	National Institute for Theoretical and Computational Sciences	online Software Carpentry (Shell, Git, Programming with Python) workshop	4-8 April 2022	Supporting Instructors

	(NITheCS), South Africa			
3	University of Cape Town, Cape Town, South Africa	Online Unix/Command Line Workshop, UCT carpentries Workshop	5-7 October 2021	Supporting Instructor
4	Faculty of Sciences Dhar Al Mahras, Fez, Morocco	Practical work of the module of genetics	3 years	Helper

SUPERVISION EXPERIENCE

N	**Type of supervision**	**Number of students**
1	Co-supervise PFE S6 Genetics	2

PROFESSIONAL MEMBERSHIPS

2019- IRC member associated

2019- Cancer Epigenetics Society

2019- MFAA (the Moroccan Fulbright Alumni Association) board

2018- Member, Moroccan Association for Telemedicine and eHealth

2016- American Society for Microbiology

2016- Member of the H3 African Bioinformatics Network

2016- Member, African Society of Human Genetics

DEDICATION

To the strong and brave fighting cancer

To my rock, to my eternal love

To my Mother....

You inspire me, You encourage me, You strengthen me.....Love you to the moon and back

ACKNOWLEDGEMENTS

First, I would like to express my deepest gratitude to my supervisor Prof. Hicham BEKKARI for his continuous support, guidance, understanding and encouragement throughout my Ph.D. He is not only great professor with immense knowledge, but also excellent advisor that support and drive students to the right direction. Without his patient guidance and support, I would not have been able to finish this dissertation.

I would like to thank my co-supervisor Prof Hassan Ghazal. Thank you for challenging me and for all the life lessons that I have learned through you and with you. Thank you for everything I learned from you since day 1. Thank you for your advice your help and your enthusiasm.

I would like to thank Prof Karim OULDIM for believing in me and for being such an inspiring person. Your continued support and enthusiasm for my project has meant the world to me.

I have had the privilege of having amazing mentors during my PhD study who have supported and inspired me. my American co-supervisors Prof. Norman H. Lee and Prof Keith Crandall. Thank you for welcoming me into your labs and introducing me to the world of genomics and bioinformatics. Thank you for showing me how great working in a collaborative environment with amazing scientists and people can be. I am grateful for your patience, guidance, understanding, Kindness and support before, during and after I left the US, that made my Fulbright experience a life changing and an incredible learning experience for me.

I am extremely grateful to Prof Laila Bouguenouch for her support, advice, encouragement and most importantly her boundless enthusiasm. Thank you for being approachable and for your modesty, kindness and your endless willing to help.

I would like to express my deepest gratitude and appreciation to Prof Dalila Bousta (FSDM Fez), for accepting to preside the thesis defense committee of my dissertation. I'm really thankful and grateful for your time and your consideration.

My Honourable gratitude and thanks go to Prof Khalil Hammani (Faculty polydisciplinary of Taza), Prof Nabil Ismaili (University Hospital Mohammed VI/University Hospital Cheikh Khalifa/ University Mohammed VI of Sciences and Health), Prof Noureddine Boukatem (Faculty of Sciences Oujda), for accepting to report and evaluate my thesis and provide their valuable feedback.

A big thank to all of the collaborators who have contributed to this study by collecting samples, performing next generation sequencing and provide valuable discussions and instructions, especially, Dr. Stefan Ambs, Dr Mohamed Khan, Dr Tiffany Dorsey, Dr Wei Tang, Dr Huaitian Liu, Prof Barki Ali, Dr Salsabil Hamdi, Dr Amine, Dr Aynaoui Mohammed, Prof Moulay Hassan Farih, Dr Mohamed Mezyiene : This work would not have been done without you.

I would like to thank the Fulbright program, MACECE and AMIDEAST for giving me the opportunity to conduct great academic research, providing opportunities to share and discuss with the international community.

Thank you for my lab mates, in Morocco and the US, with whom I have had the opportunity to learn from and collaborate and establish a friendship with them including Jacqueline Moy, Matthew Bendall, Jimmy Bernot, Kaitlin Garofano, Tessa Knox, Emily Sun, Hanane Touijer, Ihssane el Bouchikhi…etc.

I wish to thank all the team members in Oncogenetics lab of Fez, Morocco, National institute of Cancer, NIH, US, Prof Norman Lee lab's members and Prof Keith Crandall lab's members.

I am also grateful to those who gave me help, support and encouragement on my work especially: Jacqueline Moy, Ihssane Elbouchikhi, Matthew Bendall, Jimmy Bernot.

I have had the privilege of being a part of an amazing network "H3ABioNet" that introduced me to the most cutting-age research area of genomics and bioinformatics and provided me with the opportunities to interact and learn from an international research scientist community.

Finally, but most importantly, my Ph.D. would not have been possible without my family' support and encouragement. I would like to express my heartfelt thanks to my family for giving me motivation and love to finish my PhD study. Especially, I owe my greatest acknowledgements to my parents, my sweet little brother and especially my mom for all their unwavering support in everything I do, for their unconditional love and support. thank you for always being there for me, for believing in me and encouraging me to pursue my dreams.

LIST OF ABREVIATION

A

AJCC: American Joint Committee on Cancer

ADT: Androgen deprivation therapy

AR: androgen receptor

ATBF1: Known as ZFHX3 (Zinc finger homeobox protein 3)

ARE: androgen reaction element

AML: acute myeloid leukaemia

AREs: androgen-responsive elements

APC: adenomatous polyposis coli protein

AURKA : Aurora kinase A

ADAMTS1: ADAM Metallopeptidase With Thrombospondin Type 1 Motif 1

AK2: adenylate kinase 2

B

BPH: Benign prostatic hyperplasia

BRCA1/2: Breast Cancer 1/2 Early Onset

BCR: biochemical recurrence

BRAF: B-Raf

BMP3: bone morphogenetic protein 3

BCAT1: branched- chain-amino- acid aminotransferase, cytosolic

BAM file: compressed binary version of a SAM file

BWA: based alignment algorithm

C

CNAs: copy number alterations

CNV: Copy Number Variations

CRs: Chromosomal Rearrangements

CZ: central zone

CHEK2: Human gene Checkpoint Kinase 2

ctDNA: circulating tumor DNA

CRPC: castrate-resistant prostate cancer

CHD1: Chromodomain-Helicase DNA-binding 1

CYP17A1: cytochrome P450 17A1

CTCs: circulating tumor cells

CYP3A4: cytochrome P450 family 3 subfamily A member 4

CXCR4: C-X-C Motif Chemokine Receptor 4

CUTL2: CUT-LIKE HOMEOBOX 2

CROCC: ciliary rootlet coiled-coil, rootletin

CFAP74: cilia and flagella associated protein 74

D

DRE: Digital Rectal Exam

DHT: dihydrotestosterone

E

ELAC2: ElaC Ribonuclease Z 2

EPHB2: Ephrin type-B receptor 2

ETS: erythroblast transformation-specific

ERG: ETS-related gene

EGFR: epidermal growth factor receptor

ETV1: ETS Variant Transcription Factor 1

ETV4: ETS Variant Transcription Factor 4

EZH2: Enhancer of zeste homolog 2

EIF1AX: Eukaryotic translation initiation factor 1A, X-chromosomal

F

FOXA1: Forkhead box protein A1

FHA: Filamentous hemagglutinin

FLI1: Friend leukemia integration 1 transcription factor

FFPE: Formalin-fixed paraffin-embedded

FKBP5: FKBP prolyl isomerase 5

FAM13C: family with sequence similarity 13 member C

G

GnRH: Gonadotropin Releasing Hormone

GATK: Genome Analysis Toolkit

GEPIA: Gene Expression Profiling Interactive Analysis

GSTP1: glutathione *S*- transferase P1

GNMT: Glycine N-Methyltransferase

H

HOXB13: Homeobox B13

HNF1B: Hepatocyte Nuclear Factor 1-Beta

HER2: human epidermal growth factor rec

HPC: hereditary prostate cancer

HGPIN: high-grade prostatic intraepithelial neoplasia

HRR: homologous recombination-related

HPGD: 15-Hydroxyprostaglandin Dehydrogenase

HSPG2: heparan sulfate proteoglycan 2

I

INDEL: INsertions/DELetions

IGFBP2: insulin-like growth factor binding protein-2

IKZF1: Ikaros Zinc finger protein 1

IDH1: Isocitrate dehydrogenase 1

IL1R2: Interleukin 1 Receptor Type 2

IDH1/2: isocitrate dehydrogenase 1/2

K

KLK3: Kallikrein-Related Peptidase 3

KRAS: Kirsten rat sarcoma virus

KLF6: Krueppel-like factor 6

KIT: KIT proto-oncogene

KCNS3: Potassium Voltage-Gated Channel Modifier Subfamily S Member 3

L

LHRH: luteinizing hormone-releasing hormone

LH: luteinizing hormone

LAMC2: Laminin Subunit Gamma 2

LUZP1: leucine zipper protein 1

M

MSP: methylation-specific polymerase chain reaction

mCRPC: metastatic castration-resistant prostate cancer

mtDNA: Mitochondrial DNA

MSI: microsatellite instability

MGMT: methylated-DNA-protein- cysteine methyltransferase

MLL2: also known as Histone-lysine N-methyltransferase 2D (KMT2D)

MYC : MYC proto-oncogene

MMP3: Matrix Metallopeptidase 3

MMP9: Matrix Metallopeptidase 9

MLL: currently known as lysine methyltransferase 2A

MYOM3: myomesin 3

N

NBN: Nibrin

NCOA2: Nuclear receptor coactivator 2

NGS: Next-Generation Sequencing

NKX3: Homeobox protein Nkx-3

NRP: neuropilin

NDRG4: N- myc downregulated gene 4

NDRG1: N- myc downregulated gene 1

NBPF1: neuroblastoma breaking point family member 1

O

OATP1B3: organic anion transporting polypeptide 1B3

OTX1: homeobox protein OTX1

ONECUT2: one cut domain family member 2

OPN: also known as secreted phosphoprotein 1

P

PCa : Prostate cancer

PLXN: Plexin

PZ: peripheral zone

PSA: Prostate Specific Antigen

PCA3: Prostate Cancer Associated 3

PARP: Poly [ADP-ribose] polymerase

PI3K/Akt: phosphoinositide 3-kinase-RAC-alpha serine/threonine-protein kinase

PTEN: phosphatase and tensin homolog

PIK3CA: phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha

PLAU: Plasminogen Activator, Urokinase

PLA1A: Phospholipase A1 Member A

PLAT: Plasminogen Activator, Tissue Type

PSMA: also known as folate hydrolase 1

R

RT-PCR: reverse transcription polymerase chain reaction

RNASEL: Ribonuclease L

RT-PCR: reverse transcriptase polymerase chain reaction

RP: Radical prostatectomy

RASSF1: ras association domain- containing protein 1

RET: rearranged during transfection

RAMED : Régime d'Assistance Médicale

RTK-RAS: Ras/Mitogen-Activated Protein Kinase

S

SNVs: Single Nucleotide Variants

SRD5A2: Steroid-5-Alpha-Reductase 2

STAG1: Stromal Antigen 1

SPOP: Speckle-type POZ protein

SPINK1: serine peptidase inhibitor, Kazal type 1

SVs: Structural variants

SOX2: sex determining region Y box2

SLC45A3: Solute carrier family 45 member 3

SEPT9: septin-9

SHOX2: short stature homeobox protein 2

SAM file: Sequence Alignment Map

SARDH : Sarcosine Dehydrogenase

SLC43A1: Solute Carrier Family 43 Member 1

SPINT1: Serine Peptidase Inhibitor, Kunitz Type 1

SEMA: semaphorins

T

TMPRSS2-ERG: transmembrane protease serine 2:V-Ets erythroblastosis virus E26 oncogene like

TZ: transition zone

TMPRSS2: transmembrane protease serine 2

TNM: tumor, node, metastasis

TP53: Tumour Protein 53

TCGA: The Cancer Genome Atlas

TIMER: Tumor Immune Estimation Resource

TWIST1: twistrelated protein 1

TFF3: Trefoil Factor 3

TOPO cloning: Topoisomerase based cloning

TAS1R2: taste 1 receptor member 2

TCGA: The Cancer Genome Atlas

U

UBXN11: UBX domain protein 11

W

WES: whole exome sequencing

WHO: World Health Organization

WGS: whole-genome sequencing

WNT signaling: Wnt/β-catenin

Z

ZNF783: Zinc Finger Family Member 783

ZFHX3: zinc finger homeobox 3

ZBTB16: zinc finger and BTB domain containing 16

ZEB1: Zinc Finger E-Box Binding Homeobox 1

LIST OF TABLES

Chapter 1: Literature Review

Table 1. 1: molecular genetic tests for prostate cancer (36) 14
Table 1. 2: Genomic diagnostic tests for prostate cancer (37).....15
Table 1. 3: Prostate cancer treatments (41) 16
Table 1. 4 : Future Therapies (43) 17
Table 1. 5: Role of the main somatic mutations in the principal target genes in prostate cancer (46) 18
Table 1. 6: Genetic abnormalities in prostate cancer and potential therapies (37) 25

Chapter 2: Update on Prostate Cancer Epidemiology in Morocco

Table 2. 1: Clinical and pathological characteristics of Moroccan PCa patients.....41
Table 2. 2: Comparison of clinical and pathological characteristics by age group and by healthcare insurance type 43

Chapter 3: Epigenetic Analysis of Circulating Tumor DNA In Prostate Cancer: Evaluation of GSTP1 Methylation

Table3. 1: PCR mix reaction of the Methylated and Unmethylated alleles.....59
Table3. 2: PCR conditions of the Methylated and Unmethylated alleles 60
Table3. 3: Clinical Characteristics of Prostate Cancer Patients.....60

Chapter 4: TMPRSS2-ERG Gene Fusion in Prostate Cancer of Moroccan Patients

Table 4. 1: Clinical Characteristics of Prostate Cancer Patients.....79
Table 4. 2: RT-PCR primers for the detection of TMPRSS2-ERG fusions 80
Table 4. 3: Reverse transcription reaction mix and incubation protocol 83
Table 4. 4: PCR reaction mix and incubation protocol.....83

Chapter 5: Whole Exome Sequencing of Moroccan Prostate Cancer Patients

Table 5. 1 : Major web resources for cancer genomics research (53) 107
Table 5. 2: Clinical Characteristics of Prostate Cancer Patients.....109

LIST OF FIGURES

Chapter I: Literature Review

Figure 1. 1: Drawing shows anterior and posterior prostate glands (5).....6
Figure 1. 2: Estimated age-standardized incidence rates (world) in 2020, prostate, males, all ages....7
Figure 1. 3: Estimated new cases of prostate cancer in North African countries for all ages in 2020. 9
Figure 1. 4: New cases of prostate cancer in all the African countries, for all ages in 2020.....9
Figure 1. 5: The 10 frequent cancers in males, Hospital registry of cancers-CHU Hassan II. Fez (2017) 10
Figure 1. 6: (A) Adenocarcinoma of the prostate, Gleason grade 3 + 3 = score of 6, with single, separate, well-formed glands in needle core biopsy from prostate. (B) Adenocarcinoma of the prostate, Gleason grade 4 + 4 = score of 8, with cribriform and fused glands in needle core biopsy from prostate. (C) Adenocarcinoma of the prostate, Gleason grade 5 + 5 = score of 10, with solid sheet-like growth in needle core biopsy from prostate (35). 13
Figure 1. 7: Evolution of prostate cancer (89).....21
Figure 1. 8 : Highlighting the different strategies used for precision medicine. (131).....27
Figure 1. 9: Translating genomics for prostate cancer trials (129).....27

Chapter 2: Update On Prostate Ccancer Epidemiology in Morocco

Figure 2. 1: Clinical characteristics of PCa patients......42
Figure 2. 2 : Percentage of diagnosed PCa patients in Fez area from 2015 to 2021 45

Chapter 3: Epigenetic Analysis of Circulating Tumor DNA in Prostate Cancer: Evaluation of GSTP1 methylation

Figure 3. 1: Origins of Cell Free Tumor DNA (ctDNA) and Various Genomic Alterations Utilized in Liquid Biopsies. (4).....52
Figure 3. 2: Hypermethylation of the glutathione S-transferase P1 promoter as early cancer biomarker (48).....55
Figure 3. 3: DNA methylation within a gene promoter region. Unmethylated CpG islands enable gene transcription (59).....56
Figure 3. 4: Flow chart of the Bisulfite Conversion Reaction.....58
Figure 3. 5: Agarose gels showing MSP screening of GSTP1 methylation.....61

Chapter 4: TMPRSS2-ERG Gene Fusion in Prostate Cancer of Moroccan Patients

Figure 4. 1: Schematic model for TMPRSS2-promoted prostate cancer cell invasion, tumor growth,

and metastasis (8). 73
Figure 4. 2: The Arrangement of TMPRSS2-ERG Fusion Gene by Translocation (29,30).....75
Figure 4. 3: Different mRNA Splice Variants of the TMPRSS2-ERG Fusion Gene (31). 75
Figure 4. 4: Schematic diagram of TMPRSS2-ERG gene fusion on chromosome 21 and the ethnic prevalence of TMPRSS2-ERG gene fusion in prostate cancer patients (39). 76
Figure 4. 5 : Biology of TMPRSS2-ERG gene fusions in prostate cancer (54) 78
Figure 4. 6: Flow chart of the method section 79
Figure 4. 7: Agarose gels showing RT-PCR screening of TMPRSS2-ERG fusion transcripts.....87
Figure 4. 8: Representative chromatograms of TMPRSS2-ERG fusion transcript sequencing.. 88

Chapter 5: Whole Exome Sequencing of Moroccan Prostate Cancer Patients

Figure 5. 1: Etiologic evolution of prostate cancer and key steps affected by germline variants and somatic fusions. 101
Figure 5. 2: Bioinformatics workflows for NGS 104
Figure 5. 3: Scheme of a comprehensive bioinformatics pipeline to analyze personalized genomic information.. 105
Figure 5. 4: The future of cancer research 107
Figure 5. 5: WES Analysis Workflow. 113
Figure 5. 6: Mutational spectrum of the prostate cancer samples. 114
Figure 5. 7: Overview of the somatic mutations in 12 PCa patients.. 115
Figure 5. 8 Oncoplot showing the heatmap of variant classification for the top 20 genes across 12 PCa patients. 116
Figure 5. 9: Most affected pathways across the 12 PCa samples. 117

TABLE OF CONTENTS

FOREWARD AND SCIENTIFIC PRODUCTION .....

SCIENTIFIC PRODUCTION .....

HONORS & FELLOWSHIPS .....

INTERNSHIPS .....

WORKSHOPS ORGANIZATION .....

WORKSHOPS & CONFERENCES PARTICIPATION .....

TRAINING AND COURSES.....

TEACHING EXPERIENCE.....

SUPERVISION EXPERIENCE .....

PROFESSIONAL MEMBERSHIPS .....

DEDICATION.....

ACKNOWLEDGEMENTS .....

LIST OF ABREVIATION.....

LIST OF TABLES .....

LIST OF FIGURES .....

TABLE OF CONTENTS.....

.....

ABSTRACT OF DISSERTATION.....

RESUME .....

ملخص.....

GENERAL INTRODUCTION..... 1

STATEMENT OF PROBLEM..... 3

Chapter 1: Literature Review..... 6

I. Prostate Anatomy and Function 6
II. Prostate Cancer Epidemiology worldwide 7
II.1 Prostate cancer in Africa 7
II.2 Prostate Cancer in Morocco 9
III. Prostate Cancer Diagnosis 10
III.1 Prostate Specific Antigen (PSA) 11
III.2 Histopathology 12
III.3 Molecular Genetic Tests 13
IV. Prostate Cancer Treatment 15
V. Driver genes of Prostate Cancer 17
VI. Molecular Pathways 20
VII. TMPRSS2-ERG gene fusion in prostate cancer and health disparities 21
VIII. Genomics of Prostate Cancer 22
VIII.1 Whole genome sequencing and whole exome sequencing 22
VIII.2 Bioinformatics Tools on Cancer Genomics 23
VIII.3 Metagenomics of Prostate Cancer 24
IX. Precision medicine applications in prostate cancer through genomics 25

Reference 28

Chapter 2: Update on Prostate Cancer Epidemiology in Morocco 38
I. Background 38
II. Methods 39
A. Cohort constitution 39
B. Statistical analysis 40
III. Results 40
IV. Discussion 45
V. Conclusions 47
References 49

Chapter 3: Epigenetic Analysis of Circulating Tumor DNA in Prostate Cancer: Evaluation of GSTP1 Methylation 52
I. Introduction 52
II. Background 54
1. Cell-free DNA as biomarkers 54
2. GSTP1 55
3. Hypermethylation events 56
III. Materials and Methods 57
1. Sample collection 57
2. Nucleic Acid Isolation 57
3. Bisulfite treatment 57
4. Methylation Specific PCR (MSP) 58

5. Clinical Characteristics 60

IV. Results and Discussion 61
V. Conclusion 64
References 65

Chapter 4: TMPRSS2-ERG Gene Fusion in Prostate Cancer of Moroccan Patients 72
I. Introduction 72
II. Background 72
II.1 TMPRSS2 Gene 72
II.2. ERG Gene 73
II.3. TMPRSS2-ERG Gene fusion 74
II.4. Role of *TMPRSS2-ERG* fusion gene in PCa 76
II.5 Molecular Pathway of *TMPRSS2-ERG* gene fusion 77
III. Materials and Methods 78
A. Cohort constitution 79
B. Clinical Characteristics 79
C. Reagents 80
D. RNA Extraction 81
E. Reverse transcription-polymerase chain reaction 82
F. Molecular cloning and sequencing of TMPRSS2-ERG variants 84
IV. Results and Discussion 86
Conclusion 89
References 91

Chapter 5: Whole Exome Sequencing of Moroccan Prostate Cancer Patients 97
I. Introduction 97
II. Background 98
II.1 Prostate Cancer Genomics 98
I.2 Next-generation sequencing technologies and applications 101
I.3 Whole Exome Sequencing 102
I.4 Bioinformatics and cancer genomics 103
I.5 Databases and Web Tools for Cancer Genomics Study 106
III. Materials and Methods 108
A. Cohort constitution 108
B. Clinical Characteristics 109
C. DNA Extraction 109
D. Exome library preparation 110
E. Exome Sequencing 111
F. Data Processing 112

G. Variant Calling 112

IV. Results 113
V. Discussion 117
VI. Conclusion 121
References 122

GENERAL LIMITATIONS OF THESE STUDIES 129

GENERAL CONCLUSIONS & PERSPECTIVES 130

APPENDIX 132
Appendix 1: Overview of the RecoverAll Total Nucleic Acid Isolation Procedure 132
Appendix 2: Preparation of 2% Agarose Gel for Electrophoresis 133
Appendix 3: Thermo Scientific EpiJET Bisulfite Conversion Kit: REAGENT PREPARATION 133

ABSTRACT OF DISSERTATION

Prostate cancer (PCa) is the second leading cause of cancer death among Moroccan men, with the number of patients increasing over the last decade. The increase is attributed to different changes in lifestyle. In recent years, the enormous advances in the field of genomics and bioinformatics alongside the development of next-generation sequencing technologies (NGS) have widely contributed to the molecular characterization and diagnostics of prostate cancer. This knowledge has been instrumental in the development of personalized treatments for this disease. This thesis was divided into four parts. In the first part, we conducted an epidemiological study on PCa in Moroccan men to determine the specificities and the needs of this population as this cancer remains an important and alarming growing health problem. This study covered detailed clinical information about symptoms, treatments and other information related to diagnoses and clinical notes. It is expected that this pilot study will contribute to generating more epidemiological information about PCa, and justify the need to establish a National Prostate Cancer Registry and to conduct further extensive research into PCa epidemiology in Morocco.

In the second part, we investigated glutathione S-transferase pi 1 (*GSTP1)* promoter hypermethylation in DNA isolated from the plasma of Moroccan PCa patients. *GSTP*1 is the most promising biomarker candidate for PCa diagnosis and its methylation thus plays a role in the oncogenic process. Analysis of *GSTP1* promoter hypermethylation by methylation-specific PCR (MSP) thus provides a specific tool for molecular diagnosis and has a prognostic potential as a DNA methylation marker for PCa. This study represents the first study of its kind (for GSTP1 in prostate cancer) for the Moroccan population.

The third part, was devoted to identifying transmembrane protease serine 2:V-Ets erythroblastosis virus E26 oncogene like (*TMPRSS2-ERG*) gene fusion transcript variants and the potential correlation of such fusions to clinical outcome and treatment responses. Formalin-fixed paraffin-embedded tissues was collected. RNA was extracted for reverse transcription polymerase chain reaction (RT-PCR) to evaluate the *TMPRSS2-ERG* gene fusion among PCa biopsies from Moroccan patients. The analysis of *TMPRSS2-ERG* fusions highlighted the prevalence of this fusion gene in the Moroccan population, which represents North African men, and confirmed that alternative genomic mechanisms might explain the disproportionately high PCa burden in African populations. This study represents the fourth study of its kind for the African continent.

The purpose of the fourth part was to cover all possible biomarkers using whole exome sequencing (WES) to obtain a good overview of prostate cancer genomics and identify unique genetic features of PCa in the Moroccan population. DNA was extracted from tissue samples for WES, and genomic analysis was performed to determine the somatic and germline variants in addition to copy number variations (CNVs) and other alterations to identify new markers in order to characterize the abnormal exome regions using bioinformatics tools. This was done in Moroccan prostate patients with, intermediate-risk to high-risk disease. This will be the first time such an approach is conducted in Morocco and will provide Moroccan-specific variant patterns that can be applied to novel prognostic and predictive assays in PCa management for Moroccan patients and help to choose the best possible treatment for each patient while avoiding unnecessary action, pain and cost for the patients.

Key words: Prostate cancer (PCa), genomics, whole exome sequencing, GSTP1 gene, TMPRSS2-ERG gene fusion, Moroccan population

RESUME

Le cancer de prostate (PCa) est la deuxième cause de décès par cancer chez les hommes au Maroc, avec une augmentation du nombre de patients au cours de la dernière décennie. Ces dernières années, de grands progrès dans le domaine de la génomique et la bioinformatique à côté du développement des technologies du séquençage de nouvelle génération (NGS) ont servi positivement à la caractérisation moléculaire du cancer de prostate. Ces connaissances ont contribué de manière décisive au développement de traitements ciblés dans les stades avancés du cancer de prostate. C'est dans cet objectif que s'inscrit les travaux de notre thèse : Dans une première partie, nous avons mené une étude épidémiologique sur le PCa des hommes marocains dans le but de déterminer les spécificités et les besoins de cette population car ce type de cancer reste un problème de santé public important et alarmant. Le type histologique dominant est l'adénocarcinome avec plus de 70% de patients qui arrivent à un stade clinique avancé avec des tumeurs à risque intermédiaire a élevé. Dans la deuxième partie, Nous nous sommes intéressés à l'étude de méthylation du promoteur de la glutathion S-transférase pi 1 (*GSTP1*) dans l'ADN isolé du plasma de patients atteints de PCa. L'altération épigénétique la plus fréquente rencontrée dans le cancer de prostate est le changement du statut de méthylation de l'ADN. Ainsi, l'hyperméthylation du promoteur de gène *GSTP1*, qui est un gène favorisant la réparation de l'ADN, se retrouve de manière quasi systématique dans le cancer de prostate. Dans notre étude, 94% des patients présentent un gène *GSTP1* méthylé. L'identification de ce dernier, sur sérum, serait le meilleur marqueur indépendant de pronostique pour les cancers localisés de prostate. Dans la troisième partie, nous nous sommes intéressés à l'identification du gène de fusion transmembranaires de la protéase sérine 2:V-Ets virus de l'érythroblastose E26 oncogène like (*TMPRSS2-ERG*) en utilisant la méthylation MSP puis à l'étude de la corrélation potentielle de ces gènes de fusions avec les résultats cliniques et les réponses au traitement du PCa. Notre étude représente la première étude de ce genre au Maroc et la quatrième étude sur ce gène de fusion chez des patients atteint du cancer de prostate dans le continent African. La présence des gènes de fusion TMPRSS2-ERG pourrait servir comme un marqueur pronostique pour les cancers localisés traités par prostatectomie totale. Cette découverte devrait pouvoir conduire, dans un avenir proche, à des traitements ciblés. Dans notre étude, le gène de fusion *TMPRSS2-ERG* était présent dans 60% des cas de patients atteints de cancer de prostate. Le mauvais pronostic et l'agressivité tumoral ont été toujours associés à la présence de *TMPRSS2-ERG* sans pourtant tirer une conclusion significative et définitive due aux effectifs limités des patients atteints de PCa dans la plupart des études y compris la nôtre. La quatrième partie était consacrée à l'identification du profil génomique des individus de la population marocaine atteints du cancer de prostate. Ceci a été réalisé suite au séquençage de l'exome entier (WES) à partir de tissus tumoraux d'individus avec PCa. Grâce au séquençage et à l'analyse bioinformatique on a pu identifier le type de mutation le plus répandu dans notre cohorte qui est la mutation faux-sens qui induit le changement de l'acide aminé associé. En plus, on a pu révéler les voies de signalisation les plus touchées comme le RTK-RAS et TP53. Nous avons, aussi, tirer d'autres informations génétiques pertinentes, telles que des variants germinales et CNV, sur le cancer de prostate dans la population Marocain. Nous avons pu identifier de nouveaux variants somatiques, d'importance inconnue qui pourraient contribuer à des voies de signalisation et de traitement spécifique à la population. A notre connaissance, c'est la première fois qu'on réalise le séquençage, de l'exome complet, dans une population au Maroc et même en Afrique du nord, atteinte de cancer de prostate.

Mots clés : **Cancer de prostate (PCa), gène GSTP1, gène de fusion TMPRSS2-ERG, Séquençage exome (WES), population Marocaine.**

ملخص

يعتبر سرطان البروستاتا هو ثاني سبب رئيسي للوفاة بالسرطان بين الرجال في المغرب، مع تزايد عدد المرضى خلال العقد الماضي. و تعزى الزيادة إلى تغييرات مختلفة في نمط الحياة. تم تقسيم هذه الأطروحة إلى أربعة أجزاء. في الجزء الأول، أجرينا دراسة وبائية عن سرطان البروستاتا لدى الرجال المغاربة بهدف تحديد خصوصيات و احتياجات هذه الفئة السكانية، لأن هذا السرطان لا يزال يمثل مشكلة صحية متنامية و مثيرة للقلق. تم تخصيص الجزء الثاني، لدراسة حالة الميثيل لجين GSTP1 في الحمض النووي المستخرج من البلازما لمرضى سرطان البروستاتا. في الجزء الثالث، تم تحديد المتغيرات و العلاقة المحتملة لهذه الاندماجات مع النتائج السريرية و الاستجابة للعلاج. في الجزء TMPRSS2-ERG جين الاندماج الرابع، اقترحنا استخدام علم الجينوم و المعلوماتية الحيوية لتحديد الخصائص الجينية الفريدة و المميزة لسرطان البروستاتا في المغرب

GENERAL INTRODUCTION

PCa is one of the most common cancers diagnosed among men in most western countries (1). Approximately 75% of patients are diagnosed after 65 years. There is considerable variation in PCa incidence between countries and between different ethnic groups. Most prostate tumors are adenocarcinomas (1). A distinguishing feature of PCa which is effectively the risk factor of its intimate association with aging and genetic alterations has been studied extensively for decades. Furthermore, PCa has a high rate of genomic instability, such as deletions, and chromosomal rearrangements (1).

Compared to other cancer types, the genomics of prostate cancer have been difficult to study due to several reasons including the anatomic location of the prostate gland, which makes it challenging to obtain samples for research. Hence, recent genetic and genomic studies of prostate cancer help clarify the genetic basis of this common but complex disease. A better understanding of cancer genomics in general has occurred due to the use of high performing next-generation sequencing (NGS) technologies and the development of computational biology, new bioinformatic tools and statistical approaches that helped to address the large datasets generated from several genomic studies around the world (2).

This thesis reflects the work completed throughout my doctoral project. We have chosen to work on this pathology, which has become a significant alarming and growing public health burden due to its severity among the Moroccan population. In the first chapter, we provide a literature review with the background context for the genomics of PCa. In addition to the current diagnostic, prognostic and treatment decisions for this disease. In the second chapter, we conducted an epidemiological study on PCa in Moroccan men to determine the specificities and needs of this population. We believe this study will highlight the need for further epidemiological studies and the urgent need to establish a National Cancer registry. The third part of this thesis focuses on exploring the methylation status of glutathione promoter S-transferase pi 1 (*GSTP1*) in DNA isolated from the plasma of PCa patients. GSTP1 promoter hypermethylation is found systematically in PCa. Furthermore, this gene plays a role in the oncogenic process, and has been proven to be a potential diagnostic and prognostic marker for PCa. Hence, the importance of adopting this type of methylation test which is now fast, reliable, specific and easy to perform. In particular, recent studies on screening and prognostic analysis have shown the limits of PSA dosage. Moreover, *GSTP1* methylation in serum would be the

best independent prognostic marker of biological recurrence for localized prostate cancers. The aim of the fourth chapter was to identify the most common gene fusion in prostate cancer that has been observed in about half of tumors studied and plays a causal role in cancer development, which is the transmembrane fusion gene of the 2:V-Ets serum protease erythroblastosis virus E26 oncogene like (*TMPRSS2-ERG*) fusion gene. We explored the transcript variants of this fusion gene, using specific methylation PCR (Polymerase Chain Reaction), and the potential correlation of such fusions to the clinical outcome and treatment responses. The presence of *TMPRSS2-ERG* fusion genes could serve as a prognostic marker for localized cancers treated with total prostatectomy. This discovery should lead to targeted treatments in the near future. The data generated from our study aside with other data, if validated by extensive studies, could have diagnostic, prognostic and therapeutic implications in the future. Our study represents the first study of its kind in Morocco and the fourth study on this fusion gene in prostate cancer patients in the African continent. The purpose of the fifth part, was to identify the genomic profile of prostate cancer in the Moroccan population by analyzing the whole exome sequencing (WES) data of FFPE and fresh tissue. To the best of our knowledge, this is the first study of its kind in Morocco and North Africa. It presents the first and most complete characterization of the whole exome of prostate cancer in the Moroccan population. We believe this study will contribute to the development of an effective population-specific treatment and help to improve the clinical outcomes. This work will contribute to the establishment of Moroccan genomics of PCa for personalized treatment.

STATEMENT OF PROBLEM

This thesis has four main purposes:

- The first objective is to conduct an epidemiological study on PCa in Moroccan patients to determine the specificities and needs of this population, as this cancer remains an important and alarming growing health problem. The following aims were addressed in part one:
  - **Aim 1**: Collect clinical information from PCa patients' medical records
  - **Aim 2:** Review each patient's medical record: the clinical characteristics, laboratory results and pathological reports
  - **Aim 3**: Analyze variables including: age, symptoms, tumor stage, Gleason score, histological type and type of treatment.
- The second objective is to investigate *GSTP1* promoter hypermethylation in circulating DNA isolated from the plasma of PCa patients. *GSTP1* is the most promising biomarker candidate for PCa diagnosis, and its methylation thus, plays a role in the oncogenic process. Additionally, the prognostic potential of DNA methylation markers for PCa has already been demonstrated. Analysis of *GSTP1* promoter hypermethylation by methylation-specific polymerase chain reaction (MSP) thus provides a specific tool for molecular diagnosis and has prognostic potential as a DNA methylation marker for this disease. The following aims were addressed in part two:
  - **Aim 1: Cohort constitution:** Recruit patients from consenting persons in PCa at university hospital center Hassan II in Fez. Blood samples were collected only from the patients who signed the consent form.
  - **Aim 2: DNA isolation and methylation- specific PCR (MSP):** Extract circulating DNA from the plasma of PCa patients. Then, the epigenetic DNA alteration which is the hypermethylation of the *GSTP1* promoter was detected using methylation-specific PCR (MSP)**.**
- The third objective is identifying the *TMPRSS2-ERG* fusion transcript variants and the potential correlation of such fusions to the clinical outcome and treatment responses. The following aims were addressed in part three:

- **Aim 1: Cohort constitution:** Recruit patients from consenting persons in PCa at University Hospital of Oujda that cover patients from a large area in Morocco. Then, formalin-fixed paraffin-embedded tissues were collected from the tumor regions in the prostate.
- **Aim 2: RNA extraction, RT-PCR and DNA sequencing**: RNA was extracted from the formalin-fixed paraffin-embedded tissue collected from PCa patients for RT-PCR and DNA was sequenced. Then, evaluated *TMPRSS2-ERG* fusion transcript variants by RT-PCR.

❖ The fourth objective is to study the prostate tumor exome variants in the Moroccan population, to identify novel and unique variants (CNVs, SVs…) of the Moroccan population. The proposed work will allow the identification of different pathways of tumor progression and lead to prognostic genomic information to use in stratified treatment strategies and personalized follow up. The following aims were addressed in part four:

- **Aim 1: Cohort constitution:** Recruit patients from consenting persons in PCa from Oujda and Casablanca University hospitals. For PCa patients, tissue samples were collected from the tumor and adjacent non-tumor regions in the prostate and blood.
- **Aim 2: Exome sequencing, sequence alignment and variant calling:** As part of the genomic characterization, first, DNA from tumor and non-tumor biopsy and blood samples collected from Moroccan PCa patients with intermediate to high-risk disease. Then, the whole exome was sequenced before performing genomic analysis to characterize the abnormal exome regions and identify several variations including single nucleotide variants (SNVs), INsertions/DELetions (INDEL), copy number variations (CNV), translocations, and other chromosomal rearrangements (CRs) using bioinformatics tools.

This work will pave the way to a larger Precision Medicine Initiative in Morocco and Africa for cancer research and therapy. This study will lead to the embracing of new molecular diagnosis and prognostic tests of *TMPRSS2-ERG* gene fusion and *GSTP1* markers in Moroccan clinical routine. However, the genetic information generated from the WES data can be applied to novel prognostic and predictive assays in PCa management. This study will contribute to the understanding of the genomic landscape of prostate cancer in African populations.

Chapter 1: Literature Review

Chapter 1: Literature Review

I. Prostate Anatomy and Function

The prostate is a male organ (gland). It is located between the neck of the bladder and the urogenital diaphragm and surrounds the urethra. It is a fibro-muscular glandular structure measuring approximately 3 cm long. The prostate is regulated by dihydrotestosterone which is synthesized from testosterone. The function of the prostate is to produce a thin fluid containing citric acid and acid phosphatase, which is added to seminal fluid during ejaculation. The prostate is classified into four different zones or lobes that are used more in pathology: the peripheral zone (PZ), the central zone (CZ), the transition zone (TZ) and the anterior fibro-muscular zone (Figure 1.1). The PZ forms approximately 70% of the prostate and surrounds the urethra. Almost 80% of prostatic cancers develop in the PZ. The CZ forms 25% of the prostate and surrounds the ejaculatory ducts (only 2.5% of prostatic cancers develop in this region, however they are the most aggressive) (3). The TZ surrounds the proximal urethra and accounts for 20% of prostatic cancers. Finally, the anterior fibro-muscular zone consists of muscle and fibrous tissue only and accounts for approximately 20% of all prostate malignancies (4).

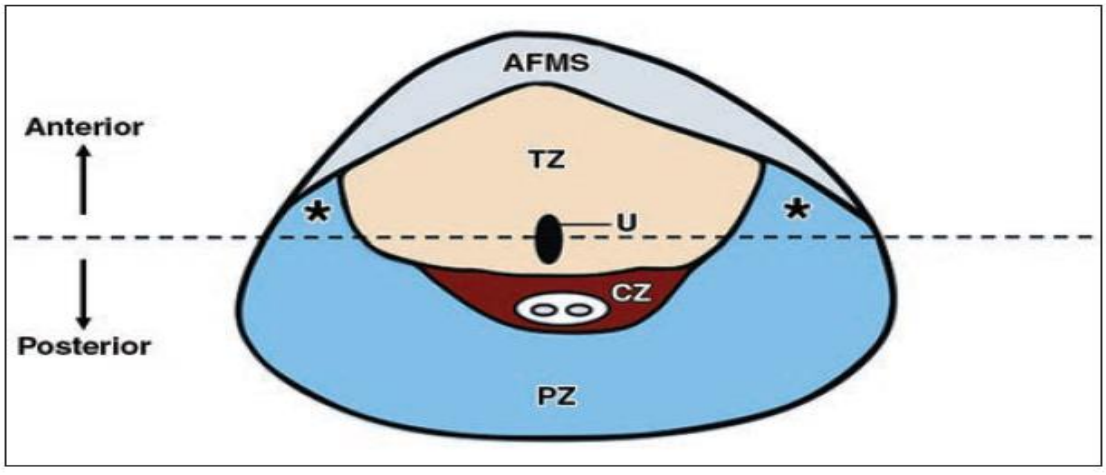


Figure 1. 1: Drawing shows anterior and posterior prostate glands. The anterior prostate gland was defined as area anterior to prostatic urethra. Asterisks indicate anterior horn of peripheral zone. AFMS = anterior fibromuscular stroma, TZ = transition zone, U = prostatic urethra, CZ = central zone, PZ = peripheral zone. (credit-Drawing by Shinmoto H) (5)**.**

II. Prostate Cancer Epidemiology worldwide

PCa is the second most commonly diagnosed cancer and the sixth leading cause of cancer death among men worldwide. Based on GLOBOCAN 2020 estimates, 1, 414, 259 new cases of PCa and 375 304 deaths were reported globally. By 2040, the worldwide PCa burden is expected to grow to almost 2.3 million new cases and 740 000 deaths due to the growth and aging of the population. The root to 70% of prostate cancers is from the peripheral zone, 20% of the transition zone and 10% of the central zone of the prostate. The average age of diagnosis of PCa is 72 years. There is a direct relationship between age as a risk factor and the incidence of PCa. Nearly three-quarters of the cases of PCa diagnosed worldwide occur in men over 65 years of age (1,6–8) (Figure 1.2).

More attention will be required to the lifestyle of individuals as a significant risk factor due to the seriousness of this type of cancer and its potential threats (9).

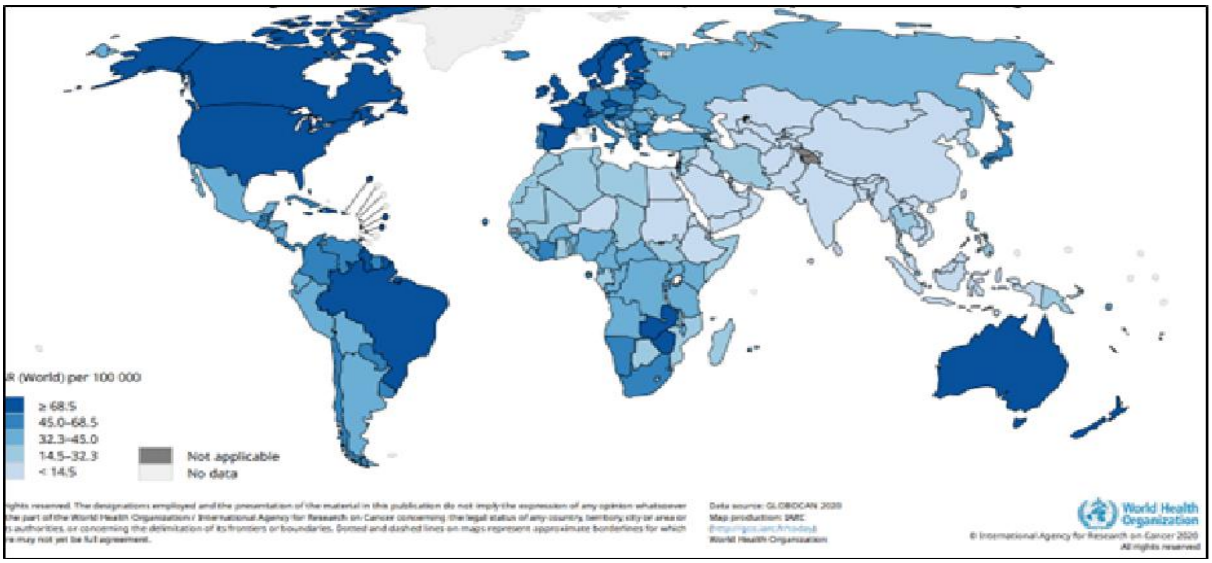


Figure 1. 2: Estimated age-standardized incidence rates (world) in 2020, prostate cancer, males, and all ages (Data source: GLOBOCAN 2020/ Map production: IARC (http://gco.iarc.fr/today)/ World Health Organization)

II.1 Prostate cancer in Africa

It is still quite difficult to accurately describe the burden of PCa in Africa even though this disease presents a significant public health problem on the African continent due to poor cancer registration systems. There is lack of a functional cancer registration system in several African countries according to the World Health Organization (WHO) (10). The few cancer registries in these regions were from population- and hospital-based cancer registries, largely due to the

difficulty in establishing high-quality population-based cancer registries in Africa (11–13). This information is so scarce that only five countries from Africa contributed to the very first volume of the IARC's "Cancer Incidence in Five Continents" (14). As a result, an accurate representation of the cancer burden is yet to be determined, due to the lack of data. In addition, as urbanization, westernization, population growth and life expectancies continue to increase in Africa, the cancer burden will increase (15).

In 2019, in all Northern African countries the average rate of incidence per 100 000 of the population was 11 800, and there were 5 100 deaths. Regarding the incidence rate for every population: in Morocco 3 990, Algeria 2 578, Libya 317, Egypt 3 109, Sudan 938, and Tunisia 819. The mortality rates were: 1861 in Morocco, 1033 in Algeria, 93 in Libya, 1237 in Egypt, 498 in Sudan, and 421 in Tunisia and the highest numbers of new cases were found in Morocco and Egypt (7,16,17) (Figure 1.3). In a statistical study by Boushbika et al., (2013) published in 2013, PCa was the second most common cancer among men in Morocco, and all of the incidences were reported after 45 years of age. Reports showed a higher incidence of PCa in Morocco compared to the one registered in Algeria or Tunisia. This is mainly attributable to the difference in screening practices. Between 2004 and 2007, there was a 62% incidence increase in Casablanca which shows that the recent screening, and therefore diagnosis practices are working. However, the incidence of PCa reported in Western countries is higher than that observed in African countries. This is attributable to major differences in screening, but the difference in diet may also play a role. It has been suggested that the Mediterranean diet may protect against PCa and is the reason behind the low incidence observed in North Africa (18).
Reports show that lower PCa incidence and mortality rates were observed in Northern Africa at 10.6 and 7.0 per 100,000, compared to the average rates in sub-Saharan African countries with 34.3 and 22.1 per 100,000, respectively. Such differences suggest that North African men do not share the high susceptibility to PCa observed in the population of sub-Saharan African origin (figure 1.4) (18,19).

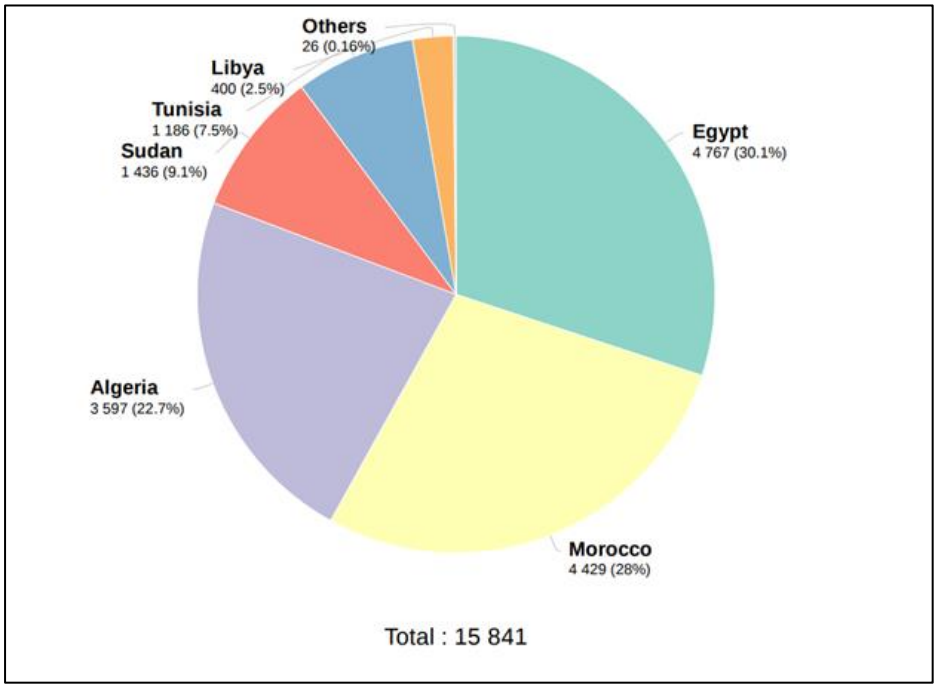


Figure 1. 3: Estimated new cases of prostate cancer in North African countries for all ages in 2020, (Data source: Globocan 2020/ Graph production: Global Cancer/ Observatory (http://gco.iarc.fr))

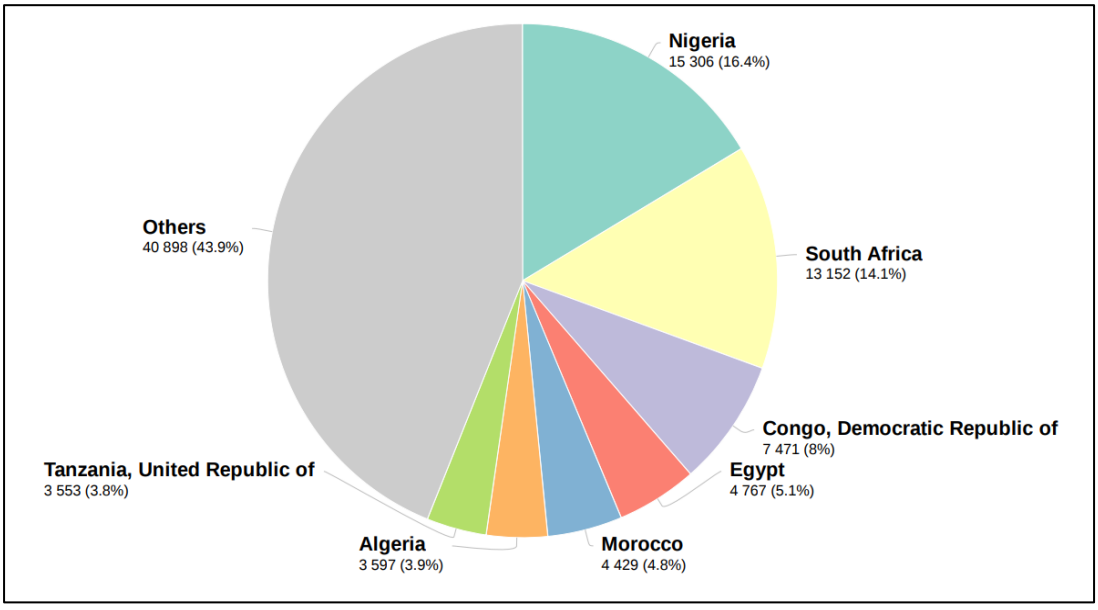


Figure 1. 4: New cases of prostate cancer in all the African countries, for all ages in 2020 ((Data source: Globocan 2020/ Graph production: Global Cancer/ Observatory (http://gco.iarc.fr))

II.2 Prostate Cancer in Morocco

In Morocco, PCa is the second leading cause of cancer related deaths among cancers that affect

men after lung cancer according to the Cancer Registry of Rabat (2012) and the Cancer Registry of Casablanca (2016) (20,21). However, PCa was classified as the most common cancer among men according to the Cancer Registry of Fez in 2017 (Figure 1.5).

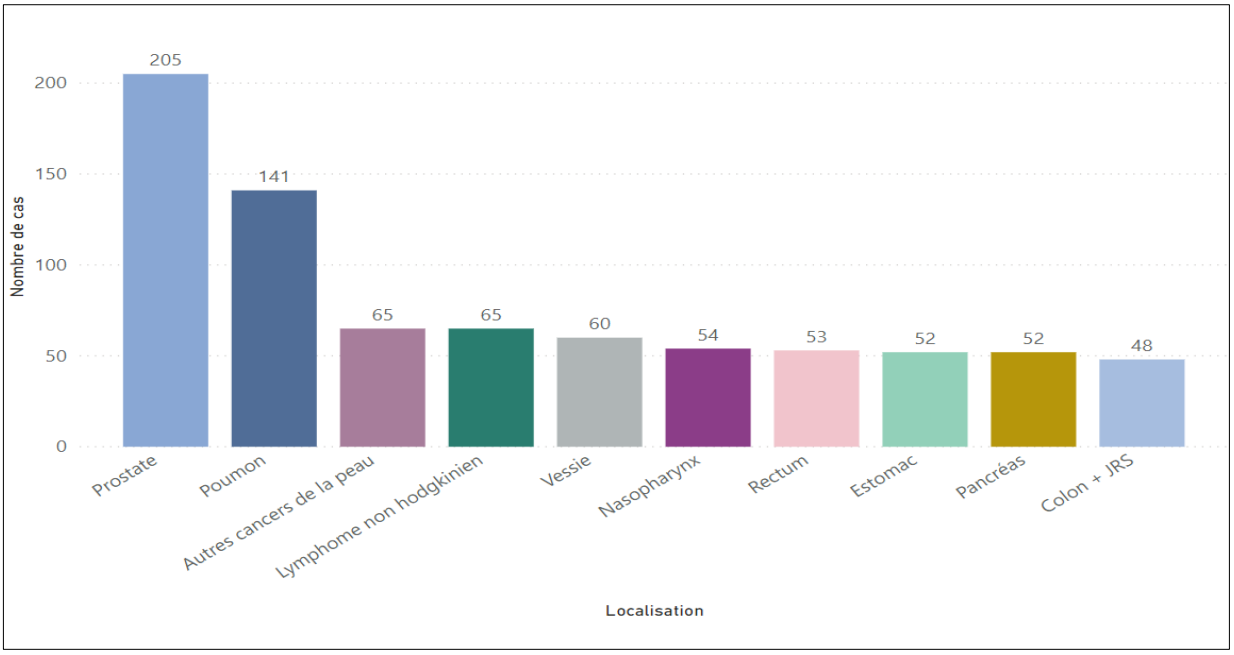


Figure 1. 5: The 10 frequent cancers in males, Hospital registry of cancers-CHU Hassan II. Fez (2017). Source : https://www.irc.ma/registres-observatoires/registre-des-cancers/registre-hospitalier

More than 99% of PCa cases are adenocarcinomas. Its incidence rate in Morocco is quite similar with slight differences to other North African countries but remains much lower than the rates recorded in European and North American countries (20).

Epidemiologically, the well-established PCa risk factors are: genetic factors, advanced age, family history and ethnicity. However, numerous factors have also been implicated in the development of PCa including : diet, alcohol consumption, ultraviolet radiation, male hormones (androgens), sexually transmitted infections and behavioral factors (1,22).

III. Prostate Cancer Diagnosis

In the early stages of this disease, when the tumor is confined to the prostate, symptoms are rare. As PCa progresses, patients may experience problems urinating, blood in semen or urine, erectile dysfunction, pain in hips, spine, or ribs, weakness or numbness in legs or feet, and/or loss of bladder control. The main secondary sites for PCa metastasis are bone, lymph nodes, and lungs. It often appears as skeletal pain in the back, hips or pelvis (23,24).

PCa is detected via a digital rectal exam (DRE) and sometimes by clinical symptoms (for

example, difficulties emptying the bladder) when no symptoms are present. Various guidelines for further evaluation exist (23,25).

The physician uses DRE to examine the prostate gland for any irregularities in size, shape, and texture by feeling the area by inserting a gloved, lubricated finger into the rectum. It is not a very precise technique and does not usually detect early PCa.

III.1 Prostate Specific Antigen (PSA)

Today, PCa is most commonly detected via the prostate specific antigen (PSA) blood test, a biomarker currently used to detect this disease, however it is not specific to PCa but rather to the general biology of the prostate gland itself, because PSA levels naturally increase as men age and this protein is basically produced by both cancerous and noncancerous tissue in the prostate. However, high levels of PSA may indicate the presence of PCa or prostatitis. The use of PSA screening has permitted the detection of this disease at much earlier stages than was previously possible, often when the tumor is still localized to the prostate (26).

PSA is a glandular serine protease that is produced and secreted by the epithelial cells of the prostate. The level of this protein is encoded by the Kallikrein-Related Peptidase 3 (*KLK3*) gene (also known as prostate-specific antigen: PSA), which is measured during a PSA test and used as a prostate cancer marker. The function of this protease is to cleave the semenogelins, which is the main protein of the semen coagulum that regulates sperm function, in the seminal fluid coagulum and plays an important role in fertility. PSA levels less than 4 ng/mL are considered “normal,” and results over 10 ng/mL are considered “high” and indicates the risk of cancer as greater than 50%. The results between 4 and 10 ng/mL are usually considered “moderate” and predict the risk of PCa to be approximately 25%. However, an elevated PSA value can also indicate benign conditions, such as benign prostatic hyperplasia (BPH) or prostatitis. Therefore, the total serum PSA level gives only an estimate of the likelihood of cancer. This is why both PSA and digital rectal examination (DRE) together are used to detect the presence of this disease (27,28).

If PCa is suspected, and the results of an initial scan that include a suspicious digital rectal exam and/or elevated PSA levels (often is combined with a family history of PCa), then a prostate biopsy is needed to determine histological grading using the Gleason scoring method which is a relevant predictor of prognosis and behavior of PCa to confirm (or exclude) the diagnosis of

cancer (29).

III.2 Histopathology

The Gleason score is obtained by the sum of the two most prevalent growth patterns (Gleason grades) in the gland. The Gleason grade is used by pathologists to evaluate the level of cancer cell differentiation and aggressiveness based on histological patterns that take into account levels of gland architecture and cell differentiation (30).

Two of the most predominant tissue patterns are graded from 1 (for well differentiated prostate glandular cells) to 5 (for poorly differentiated glandular cells) and are summed to calculate the Gleason score. Tissue from needle core biopsies or radical prostatectomies are observed at low magnification and areas of gland architecture are divided into five patterns (or grades) based on decreasing differentiation. The most prevalent (primary) and second most prevalent (secondary) grades are summed to obtain the final Gleason score and are then used as an indicator of tumor aggressiveness. Grades 2-4 (representing benign hyperplasia or very early PCa) are rarely seen. Grades 5-6 indicate low grade disease or low rates of progression post therapy. Grades 7-10 indicate a biologically aggressive cancer with poor prognosis (28). Primary tumors are also classified according to the TNM (tumor, node, metastasis) staging system (31).

The following main classification categories are defined by the American Joint Committee on Cancer (AJCC) (32).

T stage (primary tumor)

TX: Primary tumor cannot be assessed.

T0: No evidence of primary tumor.

T1: Clinically inapparent tumor neither palpable nor visible by imaging.

T2: Tumor confined within the prostate.

T3: The tumor extends through the prostate capsule.

T4: The Tumor is fixed or invades adjacent structures other than seminal vesicles, such as external sphincter, rectum, bladder, levator muscles, and/or pelvic wall. Some sub-categories exist, for example, T1c are tumors identified by needle biopsy (e.g. because of elevated PSA).

N stage (regional lymph nodes)

NX: Regional lymph nodes were not assessed.

N0: No regional lymph node metastasis.

N1: Metastasis in regional lymph node(s).

M stage (distant metastasis)

M0: No distant metastasis.

M1: Distant metastasis.

Today's diagnostic tools (PSA, Gleason score and TNM staging) perform sub-optimally in this task. The following prognostic risk groups are commonly defined (23):

- Low risk: T1-T2a, Gleason score ≤6 and PSA 20 µg/l.
- Intermediate risk: T2b or Gleason score 7 or PSA 10-20 µg/l.
- High risk: T2c-T3 or Gleason score 8-10 or PSA>20 µg/l. The majority of PCa cases are adenocarcinomas that originate in the luminal or basal gland epithelium of the prostate (33,34) (Figure 1.6).

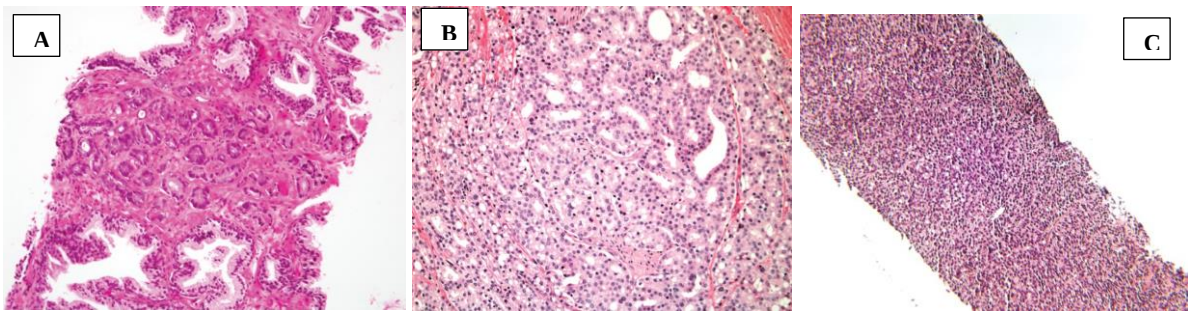


Figure 1. 6: (A) Adenocarcinoma of the prostate, Gleason grade 3 + 3 = score of 6, with single, separate, well-formed glands in needle core biopsy from prostate. (B) Adenocarcinoma of the prostate, Gleason grade 4 + 4 = score of 8, with cribriform and fused glands in needle core biopsy from the prostate. (C) Adenocarcinoma of the prostate, Gleason grade 5 + 5 = score of 10, with solid sheet-like growth in needle core biopsy from prostate (35).

III.3 Molecular Genetic Tests

PCa is considered clinically and molecularly heterogeneous between patients and within tumors. To complement the currently used screening and detection strategies, several clinical testing laboratories have introduced genetic tests aimed at identifying mutations in prostate-cancer-associated genes. According to the Orphanet (www.orpha.net/) and GeneTests (www.genetests.org/) websites, more than 20 molecular genetic tests for PCa are now commercially available in several European and Northern American countries. Approximately half of these tests are multigene panels, containing 13 to 94 genes, and are designed to assess the genetic predisposition for up to nine hereditary cancers (36) (Table 1.1) (Table 1.2).

Table 1. 1: molecular genetic tests for prostate cancer (36)

Test	Method	Gene(s)*	Laboratory
Prostate cancer sequencing panel	Sequencing	*BRCA1*, *BRCA2*, *CHEK2*, *NBN*, *TP53*	CEN4GEN (Edmonton, Canada)
Molecular diagnosis of familial prostate cancer	Full gene sequencing, Deletion/Duplication testing	*BRCA2*, *ELAC2*, *RNASEL*, *SRD5A2*, *STAG1*, *ZNF783*	Centogene AG (Rostock, Germany)
Prostate cancer test	Sequencing	*BRCA2*, *ELAC2*, *RNASEL*, *SRD5A2*	Diagenom GmbH (Rostock, Germany)
Molecular diagnosis of familial prostate cancer	NA	*BRCA2*	Institut für Klinische Genetik (Stuttgart, Germany)
PCA3 for prostate cancer	Mutation scanning of select exons	*PCA3*	Parseh Pathobiology & Genetics Laboratory (Tehran, Iran)
Molecular diagnosis of familial prostate cancer	NA	*HOXB13*	Azienda Ospedaliera Istituti Ospitalieri di Cremona (Cremona, Italy)
HOXB13 gene analysis	Sequencing	*HOXB13*	Academic Medical Centre (Amsterdam, Netherlands)
Prostate cancer test (genetic predisposition)	Sequencing, Deletion/ Duplication testing	*CHEK2*, *NBN*	GENESIS Center for Medical Genetics (Poznan, Poland)
Prostate cancer 1	Sequencing	*RNASEL*	CGC Genetics (Porto, Portugal)
Molecular diagnosis of familial prostate cancer	Sequencing	*BRCA2*, *CHEK2*	CIALAB (Alicante, Spain)
Molecular diagnosis of familial prostate cancer	Sequencing	*BRCA1*, *BRCA2*	Lorgen G.P. (Armilla, Spain)
Molecular diagnosis of familial prostate cancer	Sequencing	*BRCA1*, *BRCA2*, *CHEK2*	Laboratorio de Genética Clinica S.L. (Madrid, Spain)
Molecular diagnosis of susceptibility to familial prostate cancer	Sequencing, MLPA	*BRCA1*, *BRCA2*	IMOMA (Oviedo, Spain)
Molecular diagnosis of predisposition to breast and prostate cancer	NA	*BRCA1*, *BRCA2*	Genetiks – Genetic diagnosis and research center (Istanbul, Turkey)
Molecular diagnosis of HNF1B-gene-related diseases	Sequencing, MLPA	*HNF1B*	Centre Hospitalier Universitaire Vaudois (Lausanne, Switzerland)
HOXB13 mutation analysis (G84E)	Sequencing	*HOXB13*	Mayo Clinic (Rochester, USA)

* BRCA1/2 = Breast Cancer 1/2 Early Onset, CHEK2 = Checkpoint Kinase 2, ELAC2 = ElaC Ribonuclease Z 2, HNF1B = Hepatocyte Nuclear Factor 1-Beta, HOXB13 = Homeobox B13, NBN = Nibrin, PCA3 = Prostate Cancer Associated 3, RNASEL = Ribonuclease L, SRD5A2 = Steroid-5-Alpha-Reductase 2, STAG1 = Stromal Antigen 1, TP53 = Tumor Protein 53, G84E = germline mutation , ZNF783 = Zinc Finger Family Member 783. NA = not available (36).

Table 1. 2: Genomic diagnostic tests for prostate cancer (37)

Test	Test material	Methodology	Sample type	Distinguishing features
Decipher	Tumor RNA expression	Whole transcriptome microarray of 22 coding and noncoding RNAs	Tissue	Predicts metastasis
Oncotype DX	Tumor RNA expression	RT-PCR of 12 cancer-related and 5 reference genes	Tissue	Predicts BCR
Prolaris	Tumor RNA expression	RT-PCR of 31 cell cycle and 15 reference genes	Tissue	Predicts BCR and metastasis
PORTOS	Tumor RNA expression	RT-PCR of 24 DNA damage, immune and radiation response genes	Tissue	Predicts response to postoperative radiation therapy
FoundationACT	Somatic mutations in cell-free DNA	ctDNA of 62 genes and 6 gene fusions	Peripheral whole blood	Advantageous if tissue is not available
Modified and adapted from Falzarano et al. (2015) BCR, biochemical recurrence; ctDNA, circulating tumor DNA; RT-PCR, reverse transcriptase polymerase chain reaction				

IV. Prostate Cancer Treatment

Many treatment options for prostate cancer exist, including the following:

- Surgery
- Radical prostatectomy (RP)
- Various forms of radiotherapy that are used for more aggressive cancers that have not spread into nearby tissues or lymph nodes and have not metastasized (T1-2, N0, and M0) hormone therapy in the case of recurrent or advanced disease; and surveillance active or ‘watchful waiting’ that is usually recommended to patients with a localized low risk tumor, especially if the patient is older than 70 years and has additional diseases or if the tumor is small in size and grows slowly.

Watchful waiting is an alternative strategy where the disease is monitored (without treatment) until clinical symptoms appear, followed by hormonal treatment (androgen deprivation therapy; ADT). The choice of strategy depends on the severity of the symptoms clinicopathological

assessment (PSA serum concentrations and evolution, digital rectal exam, transrectal ultrasonography and Gleason score from biopsy), as well as patient considerations (family history, age, comorbidities and personal preferences) (23,24,38–40). (Table 1.3)

Unfortunately, curative treatment for advanced, metastatic PCa (T1-4, N0-1, and M1) is not yet available. The patient is usually treated with hormones (ADT) to reduce testosterone levels, in combination with radiation and chemotherapy to delay disease progression. However, eventually most advanced prostate cancers progress to castration resistance. At this stage, patients with metastatic disease are offered palliative treatment (24,41). Fortunately, there are some promising future therapies for metastatic castration-resistant PCa (Table 1.4).

Table 1. 3: Prostate cancer treatments (41)

Therapeutic Options	**Drug name (Generic Name)**	**Mechanism of Action**	**Cancer Stage**
Active surveillance			Early/localized PCa
Radiation therapy		Use of charged particles or waves directed toward cancer cells to induce tumor cell death.	Early/localized
Immunotherapy	Sipuleucel-T	Enhance the ability of the patient's immune system to attack cancer cells.	Early/localized
Androgen biosynthesis inhibitors	Abiraterone acetate	Inhibit the biosynthesis of testosterone by blocking an enzyme that is necessary to produce a testosterone precursor	
Bilateral orchiectomy		Surgically remove both testes to reduce testosterone levels	
Luteinizing hormone-releasing hormone (LHRH) agonists	Goserelin	Overstimulate the pituitary gland to downregulate the GnRH receptor, decrease LH production in the testes.	High-risk/advanced
	Histrelin		High-risk/advanced
	Leuprolide		High-risk/advanced
	Triptorelin		High-risk/advanced
LHRH antagonists	Degarelix	Block GnRH receptor, decrease LH production, which lowers testosterone production in the testes.	Advanced
CYP17 inhibitor	Abiraterone		Advanced (high risk/CRPC)
Anti-androgens	Apalutamide	Bind to androgen receptor and block activation of the androgen receptor	mCRPC
	Bicalutamide		mCRPC
	Darolutamide		mCRPC
	Enzalutamide		mCRPC
	Flutamide		mCRPC
	Ketoconazole		mCRPC

	Nilutamide		mCRPC
Chemotherapy	Cabazitaxel	Damage DNA and inhibt cellular machinery involved in cell division to impair mitosis of malignant cells	mCRPC
	Carboplatin		mCRPC
	Docetaxel		mCRPC
	Estramustine		mCRPC
	Mitoxantrone		mCRPC
	Temozolo-mide		CRPC

Treatments for PCa, generic and trade names of drugs, mechanism of action and stage of PCa approved for treatment. Abbreviations: LHRH=luteinizing hormone-releasing hormone; CRPC= castrate-resistant prostate cancer; PCa=prostate cancer; mCRPC: metastatic castration-resistant prostate cancer (42)

Table 1. 4 : Future Therapies (43)

Treatment/ target	**Cancer Stage**
PTEN/AKT Pathway	mCRPC
DNA Repair Pathway	mCRPC
PARP Inhibition	mCRPC
Platinum-Based Chemotherapy	mCRPC
Mismatch Repair Defects	mCRPC

V. Driver genes of Prostate Cancer

Driver genes are necessary genes for the creation of a malignant cancer with frequent mutations that usually contribute to cancer development and can be inherited (44).

The molecular etiology of PCa is still not yet clear; however, the progression of the disease includes androgen, PI3K/Akt, and *PTEN* signaling, as cycle and apoptotic pathways. Alterations in oncogenes, which are defined as driver genes in which driver mutations are activating or result in new functions, and tumor suppressor genes such as *PIK3CA*, *BRAF*, *KRAS* and *TP53* are not very common. A total of 54 oncogenes and 71 tumor suppressor genes are considered mutation driver genes (Table 1.5) (45,46).

A high percentage of PCa is considered sporadic and is caused mainly by somatic mutations in which there are some candidate genes such as *AR, TP53, KLF6, EPHB2, CHEK2, ZFHX3* (formerly known as *ATBF1*), *NCOA2, PTEN, MYC, PIK3CA, FOXA1, KIT*, and various histone-modifying genes. In general, the most commonly altered genes in PCa are: *AR*; *ETS* family gene fusions; *SPOP, TP53, PTEN* (point mutations); *CHD1* (deletion); *SPINK1*

(overexpression); and *FOXA1* (mutation and overexpression). Mitochondrial DNA (mtDNA) somatic mutations seem to be more common and frequent than autosome mutations in PCa (47–51), although only a minority of cases are considered familial and hereditary PCa (52). The major alterations seen in PCa are summarized below:

Table 1. 5: Role of the main somatic mutations in the principal target genes in prostate cancer (46)

Somatic mutation	Type mutation	Clinical relevance	References
TP53	Copy no. loss	Invasive metastasis	(53)
KRAS	Gene fusions	Metastatic progression	(54)
APC	Hypermethylation of the *APC* gene promoter	Metastatic tumors	(55)
PTEN	Deletion	Initiation, development and progression of prostate malignancies	(56)
EGFR	Overexpression of mutation of the receptor	Development of cancer	(57)
EGFR and *HER2*	Point mutations	Mitogenic signaling pathways implicated in the progression	(58)
AR	Silencing of one of the somatic AR mutations (missense mutation) can be associated simultaneously with both ‘‘gain-of-function’’ phenotype and a ‘‘loss-of-function’’ phenotype.	Advanced metastatic disease	(59,60)
SPOP	Somatic point mutations and indels	Localized and advanced prostate tumors	(61)
FOXA1	Non-silent mutations	Carcinogenesis and progression	(61)

Here are the most studied genes with strong association with prostate tumors:

The potential association between ***BRCA1*** and *BRCA2* mutations and the progression of PCa after diagnosis has been discussed for decades (62). Germline mutations in *BRCA* genes are associated with an increased risk for PCa or a more aggressive phenotype and worse outcomes (43,63,64). The families with deleterious mutations in *BRCA1* and *BRCA2* had an increased number of prostate cancers compared with families without known inherited predisposition. The risk of developing PCa is increased if a *BRCA1* or *BRCA2* mutations are present and *BRCA1* mutations might be associated with aggressive prostate disease (65,66).

BRCA2 proteins are responsible for repairing DNA. Researchers have identified hundreds of mutations in the *BRCA2* gene that are usually insertions or deletions of a small number of DNA

base pairs in the gene. *BRCA2* is a high-risk PCa susceptibility gene. There is an evidence for loss of heterozygosity of the *BRCA2* region in PCa, particularly in those at an advanced stage (66,67).

CHEK2 (Human gene Checkpoint kinase 2) is a tumor suppressor gene that encodes for the protein CHK2, a serine/threonine kinase. *CHK2* plays a critical role in DNA replication by responding to double-stranded breaks. Three deletion variants in the kinase domain in exon 10, and a missense mutation in the FHA domain in exon 3, were linked to inherited susceptibility to some tumors including PCa. Certainly, the I157T missense mutation and *CHEK2* 1100delC mutation are associated with an increased risk of PCa and should be considered in men with a familial history of this disease (68–70).

ELAC2 was identified as a candidate gene for hereditary prostate cancer (HPC) susceptibility and has been hypothesized to alter the risk of this disease. Two *ELAC2* missense variants (Ser217Leu and Ala541Thr) have been reported as risk variants of developing PCa with different risk profiles, and were also observed when cases were stratified by disease aggressiveness. These variants have been estimated to be responsible for ≤5% of PC in the general population (71–73).

SPINK1 serine peptidase inhibitor, Kazal type 1 is located on chromosome 5q32, and encodes a 56-amino acid peptide. The protein encoded by this gene is a trypsin inhibitor. Some reports suggested that *SPINK1* overexpression may represent an increased risk of PCa progression. Previous studies have shown that *SPINK1* is overexpressed in PCa tissues, especially high-grade tumors (Gleason score >7) (74,75). Furthermore, *SPINK1* does not appear to be a useful biomarker for prognostication of early-stage PCa. The correlation between such mutations in *SPINK1* and PCa, still needs more clarification at the epidemiological and molecular levels (74–76).

Mutations in the ***PTEN*** (phosphatase and tensin homolog) tumor suppressor gene are among the most frequent in PCa. Up to 70% of men with this cancer are estimated to have lost a copy of the *PTEN* gene at the time of diagnosis. The inactivation of *PTEN* either by deletion or mutation has been detected in approximately 20% of primary PCa samples at radical prostatectomy and in as many as 50% of castration-resistant tumors. Furthermore, loss of *PTEN*

function is common in prostate adenocarcinoma and leads to activation of the PI3K-AKT (phosphoinositide 3-kinase-RAC-alpha serine/threonine-protein kinase) pathway and is strongly associated with adverse oncological outcomes, making *PTEN* a potentially useful genomic marker to distinguish indolent from aggressive disease in patients with clinically localized tumors (77–79).

Several studies have examined the role of ***HOXB13*** in cancerous prostate biology, and showed that *HOXB13* has been implicated as both a tumor suppressor and an oncogene in prostate. However, it is thought that the G84E variant likely modulates the interaction between the *HOXB13* protein and the androgen receptor and affects *FOXA1*. Furthermore, the expression of *HOX* genes likely up or downregulated in human tumors including PCa (80–82).

Germline mutation G84E of ***HOXB13*** was reported in hereditary PCa. This mutation associated with an elevated risk of PCa in men of European descent and European-American men (83). One percent of the population carries the variant (rs138213197: G84E of *HOXB13*). In other studies, rare missense HOXB13 variants (Y88D, L144P, G216C, and R229G) appear to be implicated in prostate carcinogenesis (84–86).

SPOP the gene encoding Speckle-type POZ protein (*SPOP*) promotes prostate tumorigenesis. The *SPOP* gene is the most commonly mutated tumor suppressor gene in human primary prostate cancer and many mutations in this gene lead to a type of hereditary prostate tumor involving 10-14.4% of all prostate tumors. The results from The Cancer Genome Atlas (TCGA) studies show that *SPOP*-mutated prostate cancers possess the highest transcriptional activity of androgen receptor (87).

According to accumulated evidence, the expression and mutational status of *SPOP*, defined a new biotype of PCa associated with a worse prognosis (88).

VI. Molecular Pathways

Each stage of the PCa progression from high-grade prostatic intraepithelial neoplasia (HGPIN) to castration-resistant prostate cancer (CRPC) and subsequent organ-confined PCa is driven by genomic mutations. The presence of germline risk alleles in *BRCA1/2* and other gene polymorphisms increase the chance of developing this cancer. One of the first genomic events observed in cancer cells is the fusion between *TMPRSS2* and *ERG* in approximately 50% of all tumors. Subsequent mutations include the loss of one *PTEN* allele and amplification of c-MYC

on chromosome 8q24. In late-stage tumors, the second *PTEN* allele is lost and *TP53* mutations are observed. After hormone therapy, androgen receptor (AR) amplification and mutations are observed in castration-resistant prostate cancer (CRPC). Recently, sequencing efforts identified SPOP as a mutated gene in late stage cancers (89). (Figure 1.7).

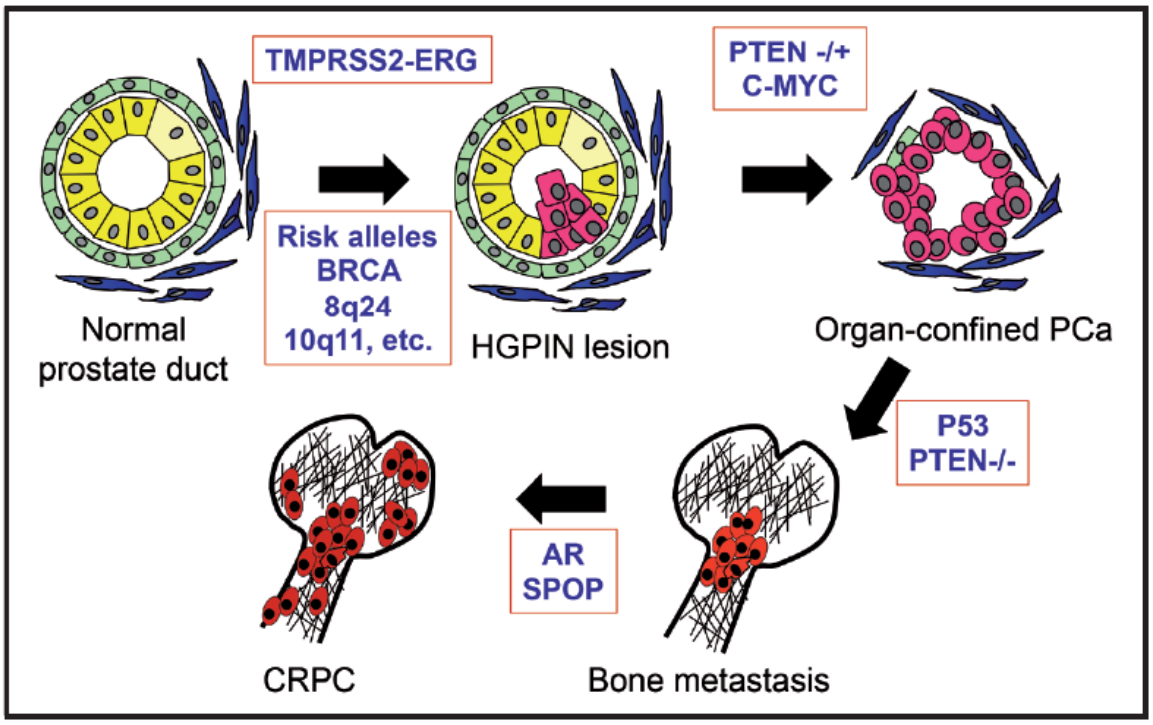


Figure 1. 7: Evolution of prostate cancer. prostatic intraepithelial neoplasia (HGPIN); androgen receptor (AR); castration-resistant prostate cancer (CRPC) (89)**.**

VII. TMPRSS2-ERG gene fusion in prostate cancer and health disparities

In 2005, Tomlin et al., reported the first gene fusion of the *TMPRSS2* (transmembrane protease, serine 2) gene and the oncogene *ERG*, which is a member of the *ETS* (erythroblast transformation-specific)-related gene) family of transcription factors that is considered an oncogenic event, as one of the commonly fusions found in prostate tumors (90).

The *TMPRSS2-ERG* gene fusion is present in about half of all prostate cancers and one of the most frequently-occurring genetic aberrations in both primary and recurrent prostate tumors and is the predominant molecular subtype of PCa (91,92).

The fusion gene may result in high expression of *ERG* by driving the androgen reaction element (ARE) of the *TMPRSS2* gene (93), which plays a critical role in the regulation of cell growth,

differentiation, and apoptosis (94). This phenomenon indicates that *TMPRSS2-ERG* fusion may be a driver of PCa progression by affecting a series of downstream oncogenic effects. Thus, the *TMPRSS2–ERG* gene is considered to be an early event in the development of PCa. Furthermore, *TMPRSS2-ERG* gene fusion can be generated by chromosomal translocation or interstitial deletion (95,96). The frequencies of translocation, deletion, and concurrence were 61.9%, 38.1%, and 0% in CA patients, and 20%, 60%, and 20% in AA cases, respectively (97). The frequency of *TMPRSS2-ERG* fusions in PCa shows a notable racial disparity (98). A fusion between the prostate-specific, androgen-regulated *TMPRSS2* gene, and the *ETS* genes *ERG*, *ETV1*, or *ETV4*; has been described in clinical prostate tumors, and occurs at a high frequency in PCa, approximately 50% of prostate cancers. The fusion frequency of *TMPRSS2-ERG* was approximately 50% in Caucasians, 31.3% in African-Americans, and 18.5% in Asians. Other studies distinguish between *TMPRSS2/ERG* and *TMPRSS2*/*ETV1* fusion genes with different subtypes existing in the tissues of PCa. *TMPRSS2*/*ERG* and *TMPRSS2*/*ETV1* fusion genes, may be used as diagnostic tools for this disease (91,97,99,100).

The *TMPRSS2-ERG* fusion is the most common genomic alteration of PCa men from a European ancestry. It is present with more advanced disease in African American men but yet their tumors are less likely to acquire *TMPRSS2-ERG*. However, data for Africa are scarce (98). African ancestry is significantly associated with high-grade (Gleason score ≥ 8) PCa (101), which questions the role of *TMPRSS2-ERG* fusions in tumor development in African populations. A 2.1-fold increase in high-grade PCa presentation was reported in Black South African men compared to African Americans (102).

VIII. Genomics of Prostate Cancer

VIII.1 Whole genome sequencing and whole exome sequencing

WGS was used on the DNA of an acute myeloid leukemia (AML) patient who led to the catalog of somatic variations and the report of the first cancer genome in 2008 (103).

Since then, WGS studies on a variety of cancers have been published. Recently whole-genome sequencing (WGS) and whole-exome sequencing (WES) have become primary strategies for NGS technologies in many genomic alterations. In most cases, CNVs and SVs, especially complex rearrangement events (such as mixed events of insertions, deletions, inversions, etc.), are identified from WGS data in cancer research (104,105). Deep WGS of cancers, using a

DNA consensus sequence from paired adjacent normal tissue as control, has the advantage of discovering a variety of somatic alterations across the whole genome including complex SV landscape in a variety of solid tumors (106–108), highlighting the heterogeneity and complexity of genome instabilities during cancer evolution.

However, WGS is considered too expensive for research involving large cohorts, and WES, can offer lower cost, higher coverage, and less complex data analysis, which are endearing for identifying clinically relevant aberrations in cancer and have become an alternative, cost-effective strategy (104,109,110).

Whole transcriptome sequencing (RNA-seq) is an appealing approach for understanding abnormal gene expression and regulation in cancer. RNA-seq data can provide valuable information to detect novel expressed transcripts, expressed somatic mutations, alternative splicing forms of transcripts, RNA editing sites that may alter the function of proteins and genome wide gene-gene fusions in the tumor tissue. Recently, a variety of in-frame fusion events that can lead to oncogene activation have been discovered by RNA-seq (111–114).

The technologies have advanced tremendously and there is an increasing knowledge of the molecular biology of PCa, so it is expected that within the next decade, whole genome sequencing can be performed in a few days for a thousand pounds. If so, the implementation of NGS in the clinic for mutation analysis of PCa will likely become reality. Using NGS, individual PCa samples can be screened for diagnostic and prognostic markers and therapy targets to determine on a personal basis, and more chromosome SVs or gene fusions that may drive cancer development or could be used as potential drug targets that can be discovered (115,116).

VIII.2 Bioinformatics Tools on Cancer Genomics

Cancer is a disease determined by several genetic and epigenetic alterations. Due to technological advances in the omics disciplines, cancer research is going through a revolution. More recently, long read sequencing technology and single molecular sequencing have enabled the detection of large and complex SVs more accurately (117,118).

In cancer genome studies, detecting different kinds of somatic variations accurately and efficiently is relevant especially with the development of several bioinformatics tools. Bias can be introduced during the library preparation and sequencing process. This is where mathematic

and probability models will be introduced to reduce the influence of the sequencing and mapping error or bias. One of the commonly used tools is GATK. The latter is used to detect somatic and germline mutations (119). Another major challenge for cancer genome analysis is detecting large SVs and complex events. In cancer, CNVs is one of the most studied epigenetic events. However, there is limited performance of the current CNV detection tools for WES data in cancer which indicates the need for developing tools with user-friendly interfaces and visualization features (120).

Currently, several powerful bioinformatics webservers/tools and approaches, are used either to identify hub genes that could serve as prognostic markers and potential treatment targets for some cancers or to analyze public transcriptomic datasets along with clinical information for oncology research. Some of these tools are KM plotter, GEPIA (Gene Expression Profiling Interactive Analysis), Oncomine and TIMER (Tumor Immune Estimation Resource) (121–125).

Recently in PCa, machine learning methods have become used to identify genes for which expression specificity predicts outcomes of this disease and could be used for screening, developing diagnostic tools, determining a personalized treatment and producing targeted drug regimens. It started a new era of artificial intelligence and machine learning in PCa (126).

VIII.3 Metagenomics of Prostate Cancer

It is known that the gut microbiome varies between different ethnicities, and it has been shown that African men have an increased risk for both aggressive prostate disease and infection. Additionally, inflammation is a hallmark of PCa and pathogenic microbes could contribute to PCa since bacteria have been shown to be involved in the development of several cancer types, hence the pathogenic agent is still unknown especially in African populations.

Feng et al., (2019), conducted a study on the prostate microbiome on African vs non-African men, together with elevated tumor mutational burden, suggesting the possibility that bacterially-driven oncogenic transformation within the prostate microbiome may contribute to aggressive disease presentation in Africa. They found that African prostate tissue samples show a 1.6-fold increased bacterial burden, with an abundance of anaerobic bacteria, increased bacterial content and enriched for Escherichia and Acidovorax, with an abundance of Eubacterium associated with host tumor hypermutation (127).

However, Pseudomonas infection may impede metastasis Escherichia and Propionibacterium have previously been shown to stimulate the progression of PCa in in vitro experiments (6, 8, 9),

Feng et al., (2019), conducted the first study on the prostate microbiome of PCa patients employing an integrated metagenomics and meta-transcriptomics approach. The composition of the bacteria detected between the matched tumor and adjacent benign tissues was not significantly different nor between different tumor grades. They have identified a strongly correlated expression profile between Pseudomonas genes and human small RNAs that may be related to metastasis (128). Further studies are needed to explore the contribution of the microbes in the PCa microenvironment to adverse disease presentation within Africa and worldwide and the accurate mechanism of this host-pathogen interaction in the progression of PCa.

IX. Precision medicine applications in prostate cancer through genomics

NGS technologies have accelerated the molecular classification of PCa based on genomic sequencing. However, because every type and every molecular subclass of PCa fingerprint is different, each cancer needs a custom treatment (Table 1.6). This is a promising approach for the development of precision medicine for making more informed treatment decisions (129).

One of these potential therapies is Olaparib which is the first ribose polymerase inhibitor (PARPi) oral drug that was approved by the US Food and Drug Administration for men with mCRPC with HRR gene mutations and has been incorporated into the National Comprehensive Cancer Network (NCCN) guidelines for prostate cancer as a Category 1 indication for men with mCRPC with HRR (Homologous Recombination Repair) gene mutations including the BRCA2 gene (130).

Table 1. 6: Genetic abnormalities in prostate cancer and potential therapies (37)

Gene	Type of aberration	Function and traits	Potential therapy
BRCA2	Somatic mutations	Repairs DNA strand breaks Tumor suppressor	Olaparib
CYP17A1	Differential expression in the tumor microenvironment	Involved in sterol synthesis	Ketoconazole, abiraterone, seviteronel

SOX2	Differential expression in tumor cells undergoing lineage plasticity	Transcription factor important for stem cells	EGFR inhibitors
OATP1B3	Differential expression in tumor cells, and inherited polymorphisms	Transporter of drugs, hormones and small molecules	–
AR-V7	Splice variants	Androgen receptor variant. Marker of resistance to abiraterone and enzalutamide	Miclosamide, onalespib, aurora A kinase inhibitors
TMPRSS2- ERG	Somatic mutations	Fusion of an enzyme and a transcription factor. Found in 40%–80% of prostate cancers[13]	–
MSI	Somatic mutations and hypermutability resulting from Lynch syndrome	Frequent mutational status	Pembrolizumab
CYP17A1, cytochrome P450 17A1; EGFR, epidermal growth factor receptor; MSI, microsatellite instability; OATP1B3, organic anion transporting polypeptide 1B3; SOX2, sex determining region Y box2.			

Precision medicine is the management of treatment profiles using genomic profiling of patient tumors across cancer types to determine the diagnosis of disease subcategories, the prognosis of patients and to predict therapies for individual cancer patients. Precision medicine for PCa has the potential to impose changes in clinical practice, and recent advances in PCa therapy have involved markers of prostate tumorigenesis, progression, and acquired therapeutic resistance (131). Different strategies have been developed for precision medicine, despite the major logistical challenge toward implementing precision medicine for PCa in Africa is the difficulty to establish an infrastructure for genomic sequencing in cancer, to collect the tissue samples and the skills to analyze the large amount of genomic data generated (129,132) (Figure 1.8).

Genomics is an emerging field that can be used for PCa trials as well, especially with discovery of a number of novel genomic alterations for this disease (129) (Figure 1.9).

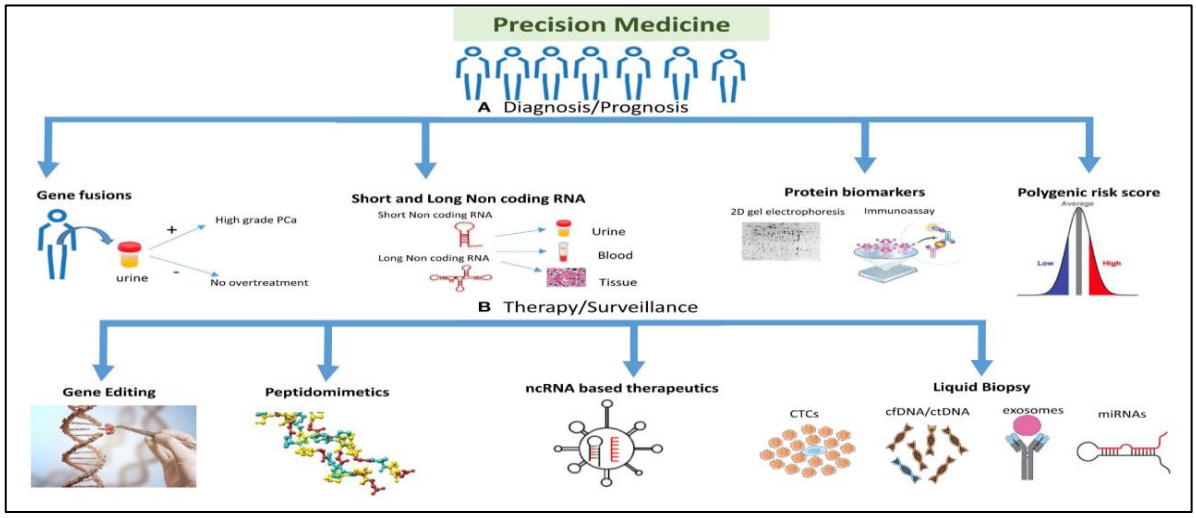


Figure 1. 8 : Highlighting the different strategies used for precision medicine. The precision medicine approach could be divided into different strategies and technologies which are being used to target the disease. (A) Diagnosis/prognosis: polygenic risk profiling could help differentiate a population or individual into high/intermediate/low risk patient, whereas molecular markers like gene fusions, protein biomarkers (e.g., 2D gel electrophoresis, MS-based proteomics and immunoassays) and non-coding RNA (short and long) could help detect prostate cancer (PCa) at different stages of the disease including primary tumor stage or treatment response. Gene fusions could help in detecting PCa at different stages and also in reducing overtreatment for patients. (B) Therapy/surveillance: clinical utility of circulating tumor cells (CTCs), circulating tumor DNA (ctDNA), and cell free DNA (cfDNA), microRNAs (miRNAs), and exosomes represents an evolution in cancer diagnosis, prognosis, and treatment. New viral/nanoparticle-based non-coding RNA (ncRNA) therapeutics have evolved in the twenty-first century with many siRNA and miRNA-based therapies in clinical trials. Antisense oligonucleotides and peptidomimetics offer an out-of-the-box approach to target genes and proteins at transcriptional and translational levels repressing their activities. Gene editing is a fascinating approach being improved on a daily basis, which could target the disease at DNA level to repair mutations or inhibit fusion genes. Gene editing image credit: Getty images (https://bit.ly/2ql54Gk) (131).

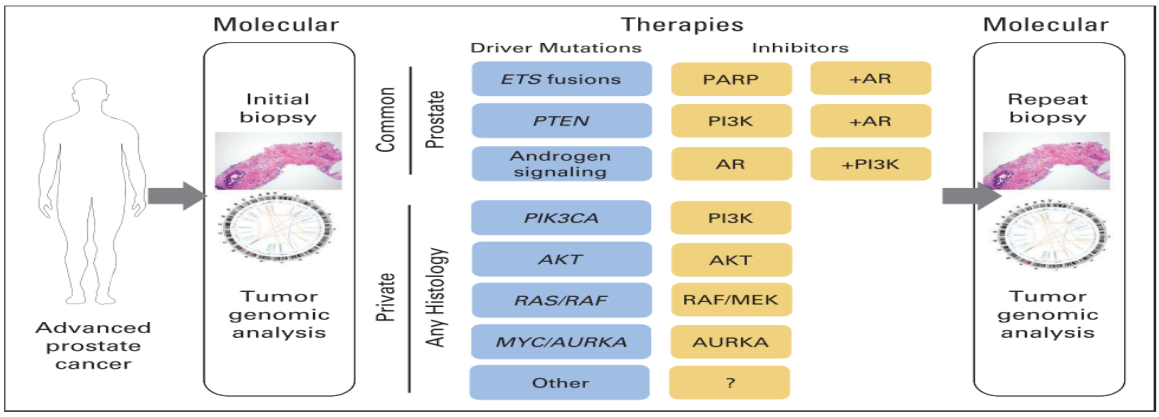


Figure 1. 9: Translating genomics for prostate cancer trials. This diagram outlines a path to genomics-driven trials in prostate cancer. Patients with advanced prostate cancer would undergo a fresh tumor biopsy for assessment of their current disease and molecular stratification to trials, including random assignment within those groups. Because of the prevalence of phosphatidylinositide 3-kinase (PI3K) pathway activation, ETS rearrangements, and androgen signalling, genomic enrichment of patients for these common disease subsets may follow traditional trial structures

including combination treatments. However, for rare or private molecular disease subsets, these patients may be better suited to studies based on molecular aberrations and that include multiple histologies. For all studies, repeat biopsy for genomic assessment will be valuable for evaluating mechanisms of resistance, including tumor subclone selection. AR, androgen receptor; PARP, poly (ADP- ribose) polymerase (129).

Reference

1. Prashanth Rawla. Épidémiologie du cancer de la prostate. Article de revue. Oncology [Internet]. 2019;32(1):2–4. Available from: doi: https://doi.org/10.14740/wjon1191
2. Rubin MA, Demichelis F. The genomics of prostate cancer: A historic perspective. Cold Spring Harb Perspect Med. 2019;9(3):1–28.
3. Cohen RJ, Shannon BA, Phillips M, Moorin RE, Wheeler TM, Garrett KL. Central Zone Carcinoma of the Prostate Gland: A Distinct Tumor Type With Poor Prognostic Features. J Urol. 2008;179(5):1762–7.
4. Bott SRJ, Young MPA, Kellett MJ, Parkinson MC. Anterior prostate cancer: Is it more difficult to diagnose? BJU Int. 2002;89(9):886–9.
5. Shinmoto H, Tamura C, Soga S, Okamura T, Horiguchi A, Asano T, et al. Anterior prostate cancer: Diagnostic performance of T2-weighted MRI and an apparent diffusion coefficient map. Am J Roentgenol. 2015;205(2):W185–92.
6. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin [Internet]. 2021;0(0):1–41. Available from: doi: 10.3322/caac.21660
7. Culp MBB, Soerjomataram I, Efstathiou JA, Bray F, Jemal A. Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates. Eur Urol [Internet]. 2020;77(1):38–52. Available from: https://doi.org/10.1016/j.eururo.2019.08.005
8. Reissigl A, Pointner J, Strasser H, Ennemoser O, Klocker H, Bartsch G. Frequency and clinical significance of transition zone cancer in prostate cancer screening. Prostate. 1997;30(2):130–5.
9. Suzuki K. Epidemiology of prostate cancer and benign prostatic hyperplasia. Japan Med Assoc J. 2009;52(6):478–83.
10. WHO. World Health statistics 2013. 2013.
11. Awadelkarim KD, Mariani-Costantini R, Elwali NE. Cancer in the Sudan: An overview of the current status of knowledge on tumor patterns and risk factors. Sci Total Environ [Internet]. 2012;423:214–28. Available from: http://dx.doi.org/10.1016/j.scitotenv.2010.09.010
12. Bowa K, Wood C, Chao A, Chintu C, Mudenda V, Chikwenya M. A review of the epidemiology of cancers at the University Teaching Hospital, Lusaka, Zambia. Trop Doct. 2009;39(1):5–7.
13. Odedina FT, Akinremi TO, Chinegwundoh F, Roberts R, Yu D, Reams RR, et al. Prostate cancer disparities in Black men of African descent: A comparative literature review of prostate cancer burden among Black men in the United States, Caribbean, United Kingdom, and West Africa. Infect Agent Cancer. 2009;4(SUPPL. 1):1–8.
14. Valsecchi MG, Steliarova-Foucher E. Cancer registration in developing countries:

luxury or necessity? Lancet Oncol. 2008;9(2):159–67.
15. Adeloye D, David RA, Aderemi AV, Iseolorunkanmi A, Oyedokun A, Iweala EEJ, et al. An estimate of the incidence of prostate cancer in Africa: A systematic review and meta-analysis. PLoS One. 2016;11(4):1–18.
16. Cherif A. Cherif, S Dhaouadi, M Osman, M. Hsairi. 2019. "12th European Public Health Conference 2019–01: Poster Walks 311." 22:2019. Eur J Public Heal Vol 29. 2019;22:2019.
17. (IARC) IA for R on C. GLOBOCAN 2018: Latest global cancer data. CA Cancer J Clin. 2018;(September):13–5.
18. Bouchbika Z, Haddad H, Benchakroun N, Eddakaoui H, Kotbi S, Megrini A, et al. Cancer incidence in Morocco: Report from Casablanca registry 2005-2007. Pan Afr Med J. 2013;16:1–10.
19. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359–86.
20. Register-Rabat. Incidence of Cancers in Rabat (2006-2008). Cancer Regist Rabat. 2012;
21. Register-Casablanca. Register of Cancers of Casablanca (2005-2006-2007. 2016;
22. Leitzmann MF, Rohrmann S. Risk factors for the onset of prostatic cancer: Age, location, and behavioral correlates. Clin Epidemiol. 2012;4(1):1–11.
23. Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, Van Der Kwast T, et al. EAU guidelines on prostate cancer. Part 1: Screening, diagnosis, and local treatment with curative intent - Update 2013. Eur Urol [Internet]. 2014;65(1):124–37. Available from: http://dx.doi.org/10.1016/j.eururo.2013.09.046
24. Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, Van Der Kwast T, et al. EAU guidelines on prostate cancer. Part II: Treatment of advanced, relapsing, and castration-resistant prostate cancer. Eur Urol [Internet]. 2014;65(2):467–79. Available from: http://dx.doi.org/10.1016/j.eururo.2013.11.002
25. Freedland SJ, Partin AW. Prostate-Specific Antigen: Update 2006. Urology. 2006;67(3):458–60.
26. Albertsen PC. Treatment of localized prostate cancer: When is active surveillance appropriate? Nat Rev Clin Oncol [Internet]. 2010;7(7):394–400. Available from: http://dx.doi.org/10.1038/nrclinonc.2010.63
27. Stamey TA, Yang N, Hay AR, McNeal JE, Freiha FS, Redwine E. Prostate-Specific Antigen as a Serum Marker for Adenocarcinoma of the Prostate. N Engl J Med. 1987;317(15):909–16.
28. Greene KL, Albertsen PC, Babaian RJ, Carter HB, Gann PH, Han M, et al. Prostate specific antigen best practice statement: 2009 update. J Urol [Internet]. 2013;189(1 SUPPL):S2–11. Available from: http://dx.doi.org/10.1016/j.juro.2012.11.014
29. Schoots IG, Padhani AR. Personalizing prostate cancer diagnosis with multivariate risk prediction tools: how should prostate MRI be incorporated? World J Urol [Internet]. 2020;38(3):531–45. Available from: https://doi.org/10.1007/s00345-019-02899-0
30. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. The 2014 international society of urological pathology (ISUP) consensus conference on gleason grading of prostatic carcinoma definition of grading patterns and proposal for a new grading system. Am J Surg Pathol. 2016;40(2):244–52.
31. Cheng L, Montironi R, Bostwick DG, Lopez-Beltran A, Berney DM. Staging of

prostate cancer. Histopathology. 2012;60(1):87–117.
32. Edge SB, Compton CC. The american joint committee on cancer: The 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol. 2010;17(6):1471–4.
33. Stoyanova T, Cooper AR, Drake JM, Liu X, Armstrong AJ, Pienta KJ, et al. Prostate cancer originating in basal cells progresses to adenocarcinoma propagated by luminal-like cells. Proc Natl Acad Sci U S A. 2013;110(50):20111–6.
34. Wang X, Julio MK De, Economides KD, Walker D, Yu H, Halili MV, et al. A luminal epithelial stem cell that is a cell of origin for prostate cancer. Nature [Internet]. 2009;461(7263):495–500. Available from: http://dx.doi.org/10.1038/nature08361
35. Humphrey PA. Histopathology of prostate cancer. Cold Spring Harb Perspect Med. 2017;7(10):1–21.
36. Laitinen V. Genetic Risk Factors for Hereditary Prostate Cancer in Finland.
37. McCrea EM, Lee DK, Sissung TM, Figg WD. Precision medicine applications in prostate cancer. Ther Adv Med Oncol. 2018;10:1–13.
38. Klotz L. Active surveillance for favorable-risk prostate cancer: Background, patient selection, triggers for intervention, and outcomes. Curr Urol Rep. 2012;13(2):153–9.
39. Autran-Gomez AM, Scarpa RM, Chin J. High-Intensity Focused Ultrasound and Cryotherapy as Salvage Treatment in Local Radio-Recurrent Prostate Cancer. Urol Int. 2012;89(4):373–9.
40. Gosselaar C, Roobol MJ, Roemeling S, Schröder FH. The Role of the Digital Rectal Examination in Subsequent Screening Visits in the European Randomized Study of Screening for Prostate Cancer (ERSPC), Rotterdam. Eur Urol. 2008;54(3):581–8.
41. Attard G, Parker C, Eeles RA, Schröder F, Tomlins SA, Tannock I, et al. Prostate cancer. Lancet. 2016;387(10013):70–82.
42. Prettyman J, Engel L, Boldt-Houle DM, Atkinson S, Wilt W. Personalizing Treatment in the Delivery of Care by Nurses To Patients with Prostate Cancer. Urol Nurs. 2019;39(2):83.
43. Sumanasuriya S, De Bono J. Treatment of advanced prostate cancer—A review of current therapies and future promise. Cold Spring Harb Perspect Med. 2018;8(6):1–13.
44. Torkamani A, Verkhivker G, Schork NJ. Cancer driver mutations in protein kinase genes. Cancer Lett [Internet]. 2009;281(2):117–27. Available from: http://dx.doi.org/10.1016/j.canlet.2008.11.008
45. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. Science (80-). 2013;340(6127):1546–58.
46. Alvarez-Cubero MJ, Martinez-Gonzalez LJ, Robles-Fernandez I, Martinez-Herrera J, Garcia-Rodriguez G, Pascual-Geler M, et al. Somatic Mutations in Prostate Cancer: Closer to Personalized Medicine. Mol Diagnosis Ther. 2017;21(2):167–78.
47. Lai CH, Huang SF, Liao CT, Chen IH, Wang HM, Hsieh LL. Clinical Significance in Oral Cavity Squamous Cell Carcinoma of Pathogenic Somatic Mitochondrial Mutations. PLoS One. 2013;8(6).
48. Lindberg J, Mills IG, Klevebring D, Liu W, Neiman M, Xu J, et al. The mitochondrial and autosomal mutation landscapes of prostate cancer. Eur Urol. 2013;63(4):702–8.
49. Tapia-Laliena MA, Korzeniewski N, Hohenfellner M, Duensing S. High-risk prostate cancer: A disease of genomic instability. Urol Oncol Semin Orig Investig [Internet]. 2014;32(8):1101–7. Available from: http://dx.doi.org/10.1016/j.urolonc.2014.02.005
50. Barbieri CE, Tomlins SA. The prostate cancer genome: Perspectives and potential. Urol

Oncol Semin Orig Investig. 2014;32(1):53.e15-53.e22.
51. Rubin MA, Demichelis F. The Genomics of Prostate Cancer: emerging understanding with technologic advances. Nat Publ Gr [Internet]. 2018 [cited 2020 Nov 21];31:1–11. Available from: www.modernpathology.org
52. Kral M, Rosinska V, Student V, Grepl M, Hrabec M, Bouchal J. Genetic determinants of prostate cancer: A review. Biomed Pap. 2011;155(1):3–10.
53. Morris E V., Cerundolo L, Lu M, Verrill C, Fritzsche F, White MJ, et al. Nuclear iASPP may facilitate prostate cancer progression. Cell Death Dis. 2014;5(10):1–14.
54. Xiao-Song Wang, Sunita Shankar, Saravana M. Dhanasekaran, Bushra Ateeq, Atsuo T. Sasaki, Xiaojun Jing, Daniel Robinson, Qi Cao, John R. Prensner, Anastasia K. Yocum, Rui Wang, Daniel F. Fries, Bo Han, Irfan A. Asangani, Xuhong Cao, Yong Li, Gilbert S. Ome and AMC. Characterization of KRAS Rearrangements in Metastatic Prostate Cancer. Cancer Discov [Internet]. 2011;(1):35–43. Available from: 763/pdf/nihms412728.pdf
55. Kenneth C. Valkenburg, Xiuping Yu, Angelo M. De Marzo, Tyler Spiering, Robert J. Matusik and BOW. Activation of Wnt/β-catenin signaling in a subpopulation of murine prostate luminal epithelial cells induces high grade prostate intraepithelial neoplasia. Prostate. 2014;1506–1520.
56. Tian L, Fang Y xiang, Xue J lun, Chen J zhong. Four microRNAs promote prostate cell proliferation with regulation of PTEN and its downstream signals in vitro. PLoS One. 2013;8(9):1–19.
57. Hour TC, Chung SD, Kang WY, Lin YC, Chuang SJ, Huang AM, et al. EGFR mediates docetaxel resistance in human castration-resistant prostate cancer through the Akt-dependent expression of ABCB1 (MDR1). Arch Toxicol. 2015;89(4):591–605.
58. Muñoz-Moreno L, Schally A V., Prieto JC, Carmena MJ, Bajo AM. Growth hormone-releasing hormone receptor antagonists modify molecular machinery in the progression of prostate cancer. Prostate. 2018;78(12):915–26.
59. Koochekpour S. Androgen receptor signaling and mutations in prostate cancer. Asian J Androl [Internet]. 2010;12(5):639–57. Available from: http://dx.doi.org/10.1038/aja.2010.89
60. Li W, Cavasotto CN, Cardozo T, Ha S, Dang T, Taneja SS, et al. Androgen receptor mutations identified in prostate cancer and androgen insensitivity syndrome display aberrant ARTt-27 coactivator function. Mol Endocrinol. 2005;19(9):2273–82.
61. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet [Internet]. 2012 Jun [cited 2020 Nov 21];44(6):685–9. Available from: https://pubmed.ncbi.nlm.nih.gov/22610119/
62. Messina C, Cattrini C, Soldato D, Vallome G, Caffo O, Castro E, et al. BRCA Mutations in Prostate Cancer: Prognostic and Predictive Implications. J Oncol. 2020;2020.
63. Pritchard CC, Mateo J, Walsh MF, De Sarkar N, Abida W, Beltran H, et al. Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. N Engl J Med. 2016;375(5):443–53.
64. Barbieri CE, Chinnaiyan AM, Lerner SP, Swanton C, Rubin MA. The Emergence of Precision Urologic Oncology: A Collaborative Review on Biomarker-driven Therapeutics. Eur Urol [Internet]. 2017;71(2):237–46. Available from: http://dx.doi.org/10.1016/j.eururo.2016.08.024

65. Kirchhoff T, Kauff ND, Mitra N, Nafa K, Huang H, Palmer C, et al. BRCA Mutations and Risk of Prostate Cancer in Ashkenazi Jews. Clin Cancer Res. 2004;10(9):2918–21.
66. Mitra A, Fisher C, Foster CS, Jameson C, Barbachanno Y, Bartlett J, et al. Prostate cancer in male BRCA1 and BRCA2 mutation carriers has a more aggressive phenotype. Br J Cancer. 2008;98(2):502–7.
67. Shaughnessy J, Tian E, Sawyer J, Bumm K, Landes R, Badros A, et al. High incidence of chromosome 13 deletion in multiple myeloma detected by multiprobe interphase FISH. Blood. 2000;96(4):1505–11.
68. Cybulski C, Huzarski T, Górski B, Masojć B, Mierzejewski M, Dębniak T, et al. A Novel Founder CHEK2 Mutation is Associated with Increased Prostate Cancer Risk. Cancer Res. 2004;64(8):2677–9.
69. Teodorczyk U, Cybulski C, Wokołorczyk D, Jakubowska A, Starzyńska T, Ławniczak M, et al. The risk of gastric cancer in carriers of CHEK2 mutations. Fam Cancer. 2013;12(3):473–8.
70. Liu C, Wang QS, Wang YJ. The CHEK2 I157T variant and colorectal cancer susceptibility: A systematic review and meta-analysis. Asian Pacific J Cancer Prev. 2012;13(5):2051–5.
71. Brzezniak LK, Bijata M, Szczesny RJ, Stepien PP. Involvement of human ELAC2 gene product in 3′ end processing of mitochondrial tRNAs. RNA Biol. 2011;8(4).
72. Rökman A, Ikonen T, Mononen N, Matikainen MP, Koivisto PA, Schleutker J, et al. ELAC2/HPC2 involvement in hereditary and sporadic prostate cancer. Cancer Res. 2001;61(16):6038–41.
73. Stanford JL, Sabacan LP, Noonan EA, Iwasaki L, Shu J, Feng Z, et al. Association of HPC2/ELAC2 polymorphisms with risk of prostate cancer in a population-based study. Cancer Epidemiol Biomarkers Prev. 2003;12(9):876–81.
74. Paju A, Hotakainen K, Cao Y, Laurila T, Gadaleanu V, Hemminki A, et al. Increased Expression of Tumor-Associated Trypsin Inhibitor, TATI, in Prostate Cancer and in Androgen-Independent 22Rv1 Cells. Eur Urol. 2007;52(6):1670–81.
75. Lippolis G, Edsjö A, Stenman UH, Bjartell A. A high-density tissue microarray from patients with clinically localized prostate cancer reveals ERG and TATI exclusivity in tumor cells. Prostate Cancer Prostatic Dis. 2013;16(2):145–50.
76. Ateeq B, Tomlins SA, Laxman B, Asangani IA, Cao Q, Cao X, et al. Therapeutic targeting of SPINK1-positive prostate cancer. Sci Transl Med. 2011;3(72).
77. Jamaspishvili T, Berman DM, Ross AE, Scher HI, De Marzo AM, Squire JA, et al. Clinical implications of PTEN loss in prostate cancer. Nat Rev Urol [Internet]. 2018;15(4):222–34. Available from: http://dx.doi.org/10.1038/nrurol.2018.9
78. Wang S, Gao J, Lei Q, Rozengurt N, Pritchard C, Jiao J, et al. Prostate-specific deletion of the murine Pten tumor suppressor gene leads to metastatic prostate cancer. Cancer Cell. 2003;4(3):209–21.
79. Mulholland DJ, Tran LM, Li Y, Cai H, Morim A, Wang S, et al. Cell autonomous role of PTEN in regulating castration-resistant prostate cancer growth. Cancer Cell [Internet]. 2011;19(6):792–804. Available from: http://dx.doi.org/10.1016/j.ccr.2011.05.006
80. Bhatlekar S, Fields JZ, Boman BM. HOX genes and their role in the development of human cancers. J Mol Med. 2014;92(8):811–23.
81. Decker B, Ostrander EA. Dysregulation of the homeobox transcription factor gene HOXB13: Role in prostate cancer. Pharmgenomics Pers Med. 2014;7(1):193–201.

82. Ewing CM. Germline Mutations in HOXB13 and Prostate-Cancer Risk. 2012;141–9.
83. Breyer JP, Avritt TG, McReynolds KM, Dupont WD, Smith JR. Confirmation of the HOXB13 G84E germline mutation in familial prostate cancer. Cancer Epidemiol Biomarkers Prev. 2012;21(8):1348–53.
84. Karlsson R, Aly M, Clements M, Zheng L, Adolfsson J, Xu J, et al. A population-based assessment of germline HOXB13 G84E mutation and prostate cancer risk. Eur Urol. 2014;65(1):169–76.
85. Haiman CA, Chen GK, Blot WJ, Strom SS, Berndt SI, Kittles RA, et al. Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. Nat Genet [Internet]. 2011;43(6):570–3. Available from: http://dx.doi.org/10.1038/ng.839
86. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet. 2007;39(8):977–83.
87. Zhuang M, Calabrese MF, Liu J, Waddell MB, Nourse A, Hammel M, et al. Structures of SPOP-Substrate Complexes: Insights into Molecular Architectures of BTB-Cul3 Ubiquitin Ligases. Mol Cell [Internet]. 2009;36(1):39–50. Available from: http://dx.doi.org/10.1016/j.molcel.2009.09.022
88. Blattner M, Lee DJ, O'Reilly C, Park K, MacDonald TY, Khani F, et al. SPOP mutations in prostate cancer across demographically diverse patient cohorts. Neoplasia (United States). 2014;16(1):14–20.
89. Alves IT. Characterization of Novel Genetic Alterations in Prostate Cancer [Internet]. 2015. Available from: http://repub.eur.nl/pub/78528/
90. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science (80-). 2005;310(5748):644–8.
91. Tomlins SA, Laxman B, Varambally S, Cao X, Yu J, Helgeson BE, et al. Role of the TMPRSS2-ERG gene fusion in prostate cancer. Neoplasia. 2008;10(2):177–88.
92. Salagierski M, Schalken JA. Molecular diagnosis of prostate cancer: PCA3 and TMPRSS2:ERG gene fusion. Vol. 187, Journal of Urology. No longer published by Elsevier; 2012. p. 795–801.
93. Clinckemalie L, Spans L, Dubois V, Laurent M, Helsen C, Joniau S, et al. Androgen regulation of the TMPRSS2 gene and the effect of a SNP in an androgen response element. Mol Endocrinol. 2013;27(12):2028–40.
94. Clark JP, Cooper CS. ETS gene fusions in prostate cancer. Nat Rev Urol [Internet]. 2009;6(8):429–39. Available from: http://dx.doi.org/10.1038/nrurol.2009.127
95. Hermans KG, Van Marion R, Van Dekken H, Jenster G, Van Weerden WM, Trapman J. TMPRSS2:ERG fusion by translocation or interstitial deletion is highly relevant in androgen-dependent prostate cancer, but is bypassed in late-stage androgen receptor-negative prostate cancer. Cancer Res [Internet]. 2006 Nov 15 [cited 2020 Nov 21];66(22):10658–63. Available from: https://pubmed.ncbi.nlm.nih.gov/17108102/
96. Perner S, Demichelis F, Beroukhim R, Schmidt FH, Mosquera JM, Setlur S, et al. TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. Cancer Res. 2006;66(17):8337–41.
97. Magi-Galluzzi C, Tsusuki T, Elson P, Simmerman K, Lafargue C, Esgueva R, et al. TMPRSS2-ERG gene fusion prevalence and class are significantly different in prostate cancer of Caucasian, African-American and Japanese patients. Prostate.

2011;71(5):489–97.
98. Blackburn J, Patrick SM, Vecchiarelli S, Heyer EE, Lyons RJ, Mercer TR, et al. TMPRSS2 - ERG fusions linked to prostate cancer racial health disparities : A focus on Africa. 2019;(March):1191–6.
99. Panagopoulos I, Soller MJ, Isaksson M, Elfving P, Soller W, Lundgren R. Confirmation of the high frequency of the TMPRSS2/ERG fusion gene in prostate cancer [Internet]. Vol. 45, Genes Chromosomes and Cancer. Genes Chromosomes Cancer; 2006 [cited 2020 Nov 21]. p. 717–9. Available from: https://pubmed.ncbi.nlm.nih.gov/16575875/
100. Ren S, Peng Z, Mao JH, Yu Y, Yin C, Gao X, et al. RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings. Cell Res. 2012;22(5):806–21.
101. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA Cancer J Clin. 2014;64(1):9–29.
102. Tindall EA, Monare LR, Petersen DC, van Zyl S, Hardie R-A, Segone AM, et al. Clinical presentation of prostate cancer in Black South Africans. Prostate [Internet]. 2014;74(8):880–91. Available from: http://doi.wiley.com/10.1002/pros.22806
103. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature. 2008;456(7218):66–72.
104. Amarasinghe KC, Li J, Hunter SM, Ryland GL, Cowin PA, Campbell IG, et al. Inferring copy number and genotype in tumour exome data. BMC Genomics. 2014;15(1):1–12.
105. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet. 2006;7(2):85–97.
106. Yang L, Luquette LJ, Gehlenborg N, Xi R, Haseley PS, Hsieh CH, et al. Diverse mechanisms of somatic structural variations in human cancer genomes. Cell [Internet]. 2013;153(4):919–29. Available from: http://dx.doi.org/10.1016/j.cell.2013.04.010
107. Moncunill V, Gonzalez S, Beà S, Andrieux LO, Salaverria I, Royo C, et al. Comprehensive characterization of complex structural variations in cancer by directly comparing genome sequence reads. Nat Biotechnol. 2014;32(11):1106–12.
108. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet [Internet]. 2010;11(10):685–96. Available from: http://dx.doi.org/10.1038/nrg2841
109. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature [Internet]. 2010;464(7289):704–12. Available from: http://dx.doi.org/10.1038/nature08516
110. Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. Annu Rev Med. 2012;63:35–61.
111. Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, et al. Transcriptome sequencing to detect gene fusions in cancer. Nature [Internet]. 2009;458(7234):97–101. Available from: http://dx.doi.org/10.1038/nature07638
112. Maher CA, Palanisamy N, Brenner JC, Cao X, Kalyana-Sundaram S, Luo S, et al. Chimeric transcript discovery by paired-end transcriptome sequencing. Proc Natl Acad Sci U S A. 2009;106(30):12353–8.
113. Berger MF, Levin JZ, Vijayendran K, Sivachenko A, Adiconis X, Maguire J, et al. Integrative analysis of the melanoma transcriptome. Genome Res. 2010;20(4):413–27.
114. Palanisamy N, Ateeq B, Kalyana-sundaram S, Pflueger D, Ramnarayanan K, Shankar

S, et al. NIH Public Access. 2011;16(7):793–8.
115. Desai AN, Jere A. Next-generation sequencing: Ready for the clinics? Clin Genet. 2012;81(6):503–10.
116. Roychowdhury S, Iyer MK, Robinson DR, Lonigro RJ, Wu YM, Cao X, et al. Personalized oncology through integrative high-throughput sequencing: A pilot study. Sci Transl Med. 2011;3(111).
117. Wang M, Beck CR, English AC, Meng Q, Buhay C, Han Y, et al. PacBio-LITS: A large-insert targeted sequencing method for characterization of human disease-associated chromosomal structural variations. BMC Genomics. 2015;16(1):1–12.
118. Zheng, G. X. BTL et al. Haplotyping germline and cancer genomes using high-throughput linked-read sequencing. Nat Biotechnol. 2016;34(3):303–11.
119. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.
120. Zare F, Dow M, Monteleone N, Hosny A, Nabavi S. An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. BMC Bioinformatics. 2017;18(1):1–13.
121. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Res. 2017;77(21):e108–10.
122. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform. Neoplasia. 2004;6(1):1–6.
123. Gyorffy B, Surowiak P, Budczies J, Lánczky A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. PLoS One. 2013;8(12).
124. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017;45(W1):W98–102.
125. Huang H, Zhang Q, Ye C, Lv JM, Liu X, Chen L, et al. Identification of prognostic markers of high grade prostate cancer through an integrated bioinformatics approach. J Cancer Res Clin Oncol. 2017;(0123456789):1–9.
126. Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. Nat Rev Urol [Internet]. 2019;16(7):391–403. Available from: http://dx.doi.org/10.1038/s41585-019-0193-3
127. Feng Y, Jaratlerdsiri W, Patrick SM, Lyons RJ, Haynes AM, Collins CC, et al. Metagenomic analysis reveals a rich bacterial content in high-risk prostate tumors from African men. Prostate. 2019;79(15):1731–8.
128. Feng Y, Ramnarine VR, Bell R, Volik S, Davicioni E, Hayes VM, et al. Metagenomic and metatranscriptomic analysis of human prostate microbiota from patients with prostate cancer. BMC Genomics. 2019;20(1):1–8.
129. Roychowdhury S, Chinnaiyan AM. Advancing precision medicine for prostate cancer through genomics. J Clin Oncol. 2013;31(15):1866–73.
130. Levee A, Lin CY, Posadas E, Figlin R, Bhowmick NA, Di Vizio D, et al. Clinical utility of olaparib in the treatment of metastatic castration-resistant prostate cancer: A review of current evidence and patient selection. Onco Targets Ther. 2021;14(July):4819–32.
131. Malik A, Srinivasan S, Batra J. A New Era of Prostate Cancer Precision Medicine.

Front Oncol. 2019;9(November).
132. Meric-Bernstam F, Farhangfar C, Mendelsohn J, Mills GB. Building a personalized medicine infrastructure at a major cancer center. J Clin Oncol. 2013;31(15):1849–57.

Chapter 2: Update on Prostate Cancer Epidemiology in Morocco

Chapter 2: Update on Prostate Cancer Epidemiology in Morocco

With more than 600,000 new cases diagnosed each year, and approximately 200,000 deaths, prostate cancer is the second most common noncutaneous cancer diagnosed in men worldwide (1). Epidemiological studies on prostate cancer are crucial in order to determine the specificities of a given population. The present study aims to document the epidemiological status of prostate cancer and to determine the extent of risk for under-diagnosis in the Moroccan population. A retrospective cross-sectional study of 141 prostate cancer cases was conducted. Prostate cancer confirming biopsies were performed at the University Hospital Hassan II in Fez and at the University Hospital Mohammed VI in Oujda during the period from 2015 to 2021. The statistical analysis was performed using SPSS version 21 software. The mean age of the enrolled patients was 72 years.

I. Background

The 2020 GLOBOCAN data (1) indicate that prostate cancer (PCa) incidence rates (IRs) in Africa are highly variable. IR presents a low rate of 16.6 per 100 000 in Northern Africa, with an increase of 6 per 100 000 compared to IR reported in GLOBOCAN 2012, while in Southern Africa, a higher rate of 65.9 per 100 000 was observed (1,2).

PCa is the most commonly diagnosed cancer and the second leading cause of cancer deaths in men in Morocco, after lung cancer (2,3). The number of PCa patients has been increasing over the last decade (3), and its IR is the highest among all cancers in patients over 65 years (more than 350 per 100 000 habitants per year) (3). According to the cancer registry of Casablanca, PCa represents 12,4% of all cancers (3). More than 99% of cases are adenocarcinomas according to the Cancer Registry of Rabat (2).

Despite the increased PCa incidence, there is a lack of information about the factors that may influence it in the Moroccan population. Age, family history and obesity are among the identified risk factors (4). Several studies have revealed that genetic alterations are also a strong risk factor in all types of PCa (5).

Accurate and extensive data on PCa patients along with effective genetic studies are relevant to

better understand the etiology of this disease and improve diagnostics and decision making in the management and treatment thereof. The aim of this study is to document the PCa occurrence in Morocco and to describe the epidemiological and histological aspects of this cancer type focusing on the two main cities of Central and Northeast Morocco, Fez and Oujda. Moreover, the study aimed to determine whether the target population presented with a high-risk disease and to evaluate the impact of the under-diagnosis of PCa.

II. Methods

In 2020, the Moroccan population was estimated to be 36.87 million; the region of Fez-Meknes (Central Morocco) contains a population of 4 377 214 while the Oujda-Angad province (Northeast Morocco) population is estimated to be 2 427 547 (6). The patients from these two major regions are mainly hosted by the Hassan II University Hospital in Fez (CHU Fez) and the Mohammed VI University Hospital in Oujda (CHU Oujda), two of five university hospitals in the country.

A. Cohort constitution

A total of 141 PCa biopsies with histopathology results confirming the PCa diagnosis were collected, 90 from CHU Fez and 51 from CHU Oujda from 2015 to 2021. All benign prostate hyperplasia (BPH) was excluded from the study. Medical records of each patient were reviewed in order to collect clinical characteristics, laboratory results and pathological reports. Several features were studied: age, symptoms, histological type, type of treatment, and stage of tumor that was allocated according to the American Joint Committee on Staging (AJCC) system as [T1 (clinically unapparent or nonpalpable tumor), T2 (tumor confined within the prostate), T3 (tumor extension through the prostatic capsule) and T4 (tumor invading into peri-prostatic tissue other than the seminal vesicles)] (7). The Gleason score was also recorded (Gleason 6 (3+3), Gleason 7 (4+3 and 3+4), Gleason 8 (4+4), Gleason 9 (4+5 and 5+4) or Gleason 10 (5+5)) (7). Patients were categorized into the following clinical categories: metastatic disease (evidence of metastases or PSA ≥ 100 ng/ml), locally advanced disease (clinical stage above T3a, no evidence of metastases and PSA < 100 ng/ml) and organ confined disease (clinical stage T3a or below, no evidence of metastases and PSA < 100 ng/ml).

B. Statistical analysis

The data were analyzed using the Statistical Package for the Social Sciences (SPSS) software version 21 (IBM SPSS statistics 21). Descriptive analyses of patient clinical characteristics and inferential statistics were generated. Two subgroup analyses were performed: one subgroup of patients younger and older than 70 years to determine if delayed presentation is related to advanced disease at the time of diagnosis, and the other subgroup was composed of patients with Mutual healthcare insurance (which is one of the healthcare systems adopted in Morocco aside with RAMED) (8) and those with RAMED healthcare insurance (which is the healthcare system designed to support the poor and vulnerable population) (9) to establish whether limited medical coverage was a possible contributing factor to advanced disease.

The Wilcoxon rank-sum test was used to compare differences for continuous variables by dichotomous age classification for the first subgroup and by type of medical coverage classification for the second subgroup. The Pearson chi-square ($\chi 2$) test and Fisher's exact test were used to assess the association between categorized variables, dichotomous age and dichotomous medical coverage classification.

III. Results

A total of 141 Moroccan PCa patients were enrolled in the current study. The mean age at presentation was 72 (standard deviation [SD] = 8, range: 52-91) (Table 2.1). The median PSA level at presentation was 34 ng/ml (interquartile range [IQR] = 12-160). Clinically staged (T2, T3 and T4) disease was present in 71.1% of patients, with 54.4% of patients at T2c stage (tumor involves both sides). The percentage of PCa patients presenting each clinical stage is reported in Table 2.1. Gleason grades of 106 PCa patients were available. Poorly differentiated tumors (Gleason grades 8, 9 and 10) were found in 29.2% of patients (Table 2.1). A total of 98.3% (n = 56) of patients were active smokers. A total of 87.5% of patients possess the RAMED healthcare insurance, while only 12.5% of patients presented with Mutual healthcare insurance.

Table 2. 1: Clinical and pathological characteristics of Moroccan PCa patients

	Non-missing observations	Mean (SD)	Median (IQR)	Range	Frequency (%)
Age	141	72(8)	72(67-78)	52-91	--
Age group		--	--	--	
< 70 years	141	--	--	--	52(36.9)
≥ 70 years		--	--	--	89(63.1)
PSA (ng/ml)	113	414 (1432)	34(12-160)	1-12006	--
T stage (detailed)					
cT1a		--	--	--	--
cT1c		--	--	--	30(29.1)
cT2a		--	--	--	3(2.9)
cT2b	103	--	--	--	8(7.8)
cT2c		--	--	--	56(54.4)
cT3a		--	--	--	3(2.9)
cT3b		--	--	--	1(1.0)
cT4b		--	--	--	2(1.9)
T stage (broad)					
T1		--	--	--	30(28.8)
T2	104	--	--	--	68(65.4)
T3		--	--	--	4(3.8)
T4		--	--	--	2(1.9)
Gleason score					
5		--	--	--	4(3.8)
6		--	--	--	27(25.5)
7	106	--	--	--	44(41.5)
8		--	--	--	21(19.8)
9		--	--	--	8(7.5)
10		--	--	--	2(1.9)
Smoking					
Yes	57	--	--	--	56(98.2)
No					1(1.8)
Type of Healthcare Insurance					
Ramedist	72	--	--	--	63(87.5)
Mutualist					9(12.5)

Legend: IQR: Interquartile range, PSA; Prostate Specific Antigen; T1a: Tumor incidental histologic finding in 5% or less of tissue resected; T1b: Tumor incidental histologic finding in more than 5% of tissue resected; T1c: Tumor identified by needle biopsy found in one or both sides, but not palpable; T2a: Tumor involving one-half of one side or less; T2b: Tumor involving more than one-half of one side but not both sides; T2c: Tumor involving both sides; T3a: Extraprostatic extension (unilateral or bilateral); T3b: Tumor invading seminal vesicle(s).

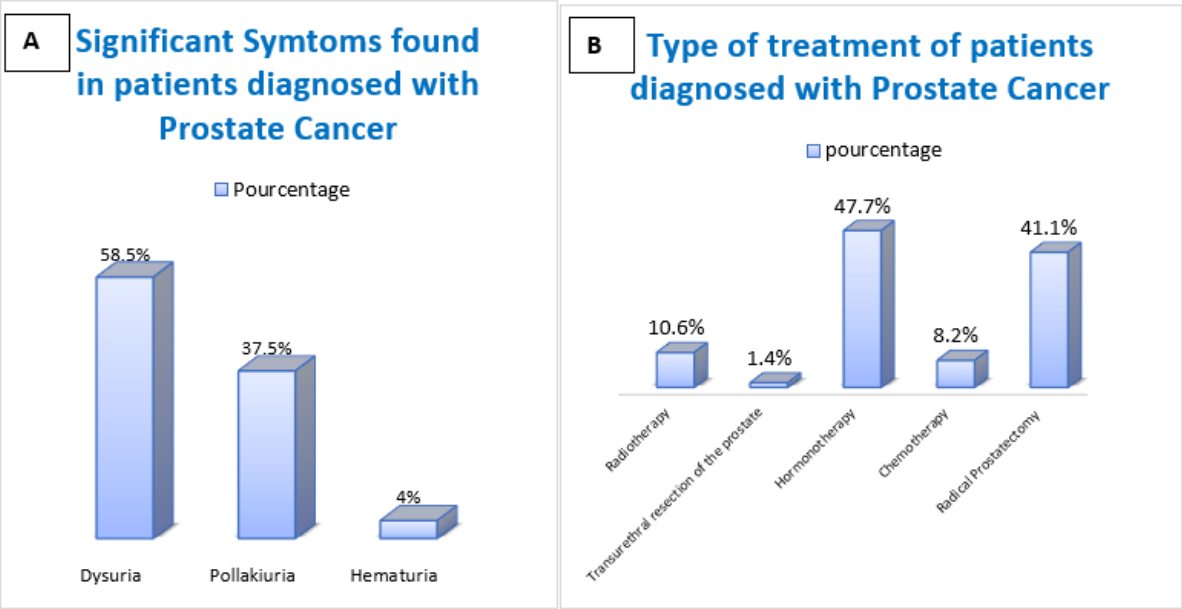


Figure 2. 1: **Clinical characteristics of PCa patients**. (A) Symptoms found in patients diagnosed with PCa. (B) Histological types found in the cohort. (C) Different types of treatments used to treat the patients in the cohort.

Descriptive analysis demonstrated that all PCa cases were adenocarcinoma (either slightly differentiated adenocarcinoma or moderately differentiated adenocarcinoma or carcinoma). The frequent symptoms were dysuria (58.5%), pollakiuria (37.5%) and hematuria (4%) (Figure 2.1 A). The majority of cases were treated by hormonal therapy (47.7% of cases), followed by radical prostatectomy (41.1%), then radiotherapy and chemotherapy were performed in 10.6% and 8.2% of cases, respectively. Transurethral resection of the prostate was less frequent (Figure 2.1 B).

The first subgroup analysis of PCa patients younger than 70 years compared to those older than 70 years suggests that patients in both groups presented an advanced disease state (Table 2.2). The median PSA level was lower in the younger group [21.98 ng/ml; IQR = 11-186] than in the older group [39.35 ng/ml; IQR = 15-152].

Table 2. 2: Comparison of clinical and pathological characteristics by age group and by healthcare insurance type

Variable	Age group		p-value	Type of healthcare Insurance		p-value
	< 70	≥ 70		RAMED	Mutual	
PSA (ng/ml):						
Non-missing observations	41	72	0.393 ‡	55	7	0.446 ‡
Median (IQR)	21.98 (11-186)	39.35 (15-152)		21 (13-155)	39.40 (10-100)	
PSA category: n (%)						
< 100	33(27.5%)	49(40.8%)	0.335*	39(60.9%)	6(9.4%)	0.756†
≥ 100	13(10.8%)	25(20.8%)		17(26.6%)	2(3.1%)	
T stage (broad): n (%)						
T1	11(10.6%)	19(18.3%)	0.945†	16(27.6%)	1(1.7%)	0.793†
T2	27(26%)	41(39.4%)		33(56.9%)	5(8.6%)	
T3	2(1.9%)	2(1,9%)		2(3.4%)	0(0.0%)	
T4	1(1%)	1(1%)		1(1.7%)	0(0.0%)	
Metastatic: n (%)						
Yes	13(9.8%)	26(19.7%)	0.558*	17(24.6%)	2(2.9%)	0.864†
No	37(28%)	56(42.4%)		44(63.8%)	6(8.7%)	
High risk n (%)						
Yes	11(8.3%)	24(18.2%)	0.286*	14(20.3%)	3(4.3%)	0.397*
No	39(29.5%)	58(43.9%)		47(68.1%)	5(7.2%)	
Risk stratification						
Metastatic/PSA ≥ 100 ng/ml	13(9.9%)	26(19.8%)	0.370†	17(24.6%)	2(2.9%)	0.896†
Locally Advanced	1(0.8%)	2(1.5%)		2(2.9%)	0(0.0%)	
Organ confined, high risk	11(8.4%)	24(18.3%)		14(20.3%)	3(4.3%)	
Organ confined, intermediate risk	14(10.7%)	21(16%)		21(30.4%)	2(2.9%)	
Organ confined, low risk	11(8.4%)	8(6.1%)		7(10.1%)	1(1.4%)	

*** Fisher's exact test**
† Pearson chi-square ($\chi 2$) test
‡ Wilcoxon rank-sum test

Individuals above 70 years old had a higher frequency of "PSA ≥ 100 ng/ml" than those under 70 years (20.8% versus 10.8%)). Most patients in the older group presented with clinically determined T2 disease. PCa patients from the 70-year-old and older group were more likely to present radiographic evidence of metastatic and high-risk disease than those younger than 70 years. However, these different comparisons of the statistical analysis performed on this subgroup of younger versus older PCa patients were not statistically significant (Table 2.2). This might be due to either the size of the overall cohort or because the number of PCa patients older than 70 years is larger, hence the need for a larger cohort.

The second subgroup analysis of patients presented with the RAMED type-insurance (Ramedists) compared to those presenting conventional Mutual Insurance (Mutualists) suggests that patients in the RAMED group were highly likely to present with advanced disease (Table 2.2). The median PSA was lower in the "Ramedist" group [21 ng/ml (IQR = 13-155)] than in the "Mutualist" group [39.40 ng/ml (IQR = 10-100)], but this difference was not significant (p-value = 0.446).

RAMEDists had a higher frequency of "PSA ≥ 100 ng/ml" than Mutualists (26.6% versus 3.1%). This difference was also not statistically significant (p-value = 0.756). Most patients in the RAMED group presented with clinically determined T2 disease but again, this was not statistically significant (p-value = 0.793). Patients with RAMED insurance were more likely to present with radiographic evidence of metastatic and high-risk disease than Mutualists; however, this was not significant (p-value = 0.864 and p-value =0.397). Similarly, there was no significant difference in the clinical stratification of Ramedist versus Mutualist patients (p-value = 0.896) (Table 2.2).

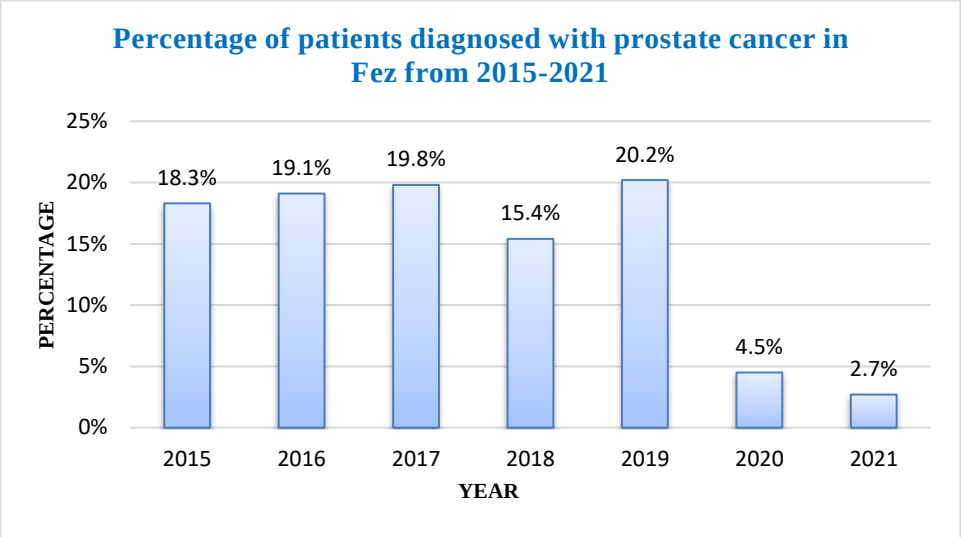


Figure 2. 2 : Percentage of diagnosed PCa patients in Fez area from 2015 to 2021

Although these findings were not statistically significant, they are indeed clinically relevant because patients were presented with advanced disease regardless of the earlier age or the type of healthcare insurance. The percentage of diagnosed PCa patients decreased significantly in 2020 and 2021, by 4.5% and 2.7%, respectively, compared to the previous years due to the COVID pandemic (Figure 2.2).

IV. Discussion

Similar to other parts of the world, North African countries have been witnessing a steady rise in PCa incidence in recent decades. In 2019, the average rate of PCa incidence per 100,000 population in North Africa was 11 800 cases, with 5100 deaths. The incidence in Morocco, specifically. was 3990 while the mortality rate was 1861 (10,11). PCa has been the object of an extensive epidemiological study for the purpose of a better management of the disease. The country has two cancer registries, the registry of Rabat and the registry of Casablanca. More recently, in 2017, a regional cancer hospital registry of the region of Fez-Meknes (Central Morocco) was also established to cover a region with a population of 4 million.

Our study aims to update national epidemiological studies on PCa and concerns a fair dataset collected from the regions of Fez and Oujda cities that, until recently, have not had a dedicated cancer registry (6). PCa risk factors have been explored extensively; they include various parameters such as diet, age, smoking, ethnicity and genetics (12,13). In our present study,

diagnosed PCa patients ranged in age between the fifties and nineties (46.3% were diagnosed under 70 and 81.3% above 70 years), which is consistent with the Casablanca registry report that states the majority of men diagnosed with PCa occur after the age of 50 years. Interestingly, no case was detected under the age of 40 years, however the risk sharply increases with age (3). More than 70% of PCa histological types are adenocarcinomas; they are the most common histological type in PCa (14,15). The mean age of PCa presentation in our study was higher than the mean age of presentation in other studies conducted in African populations (16). There are several reasons that may be a cause for this presentation with high-risk disease among Moroccan patients including delayed diagnosis until patients suffer from symptoms and the poor socioeconomic status seen in a major portion of patients with PCa. In addition, there is a lack of PCa screening programs in Morocco.

Similar to other studies (14), the PCa histological type presented in our patients was prostatic adenocarcinoma. In comparison to South Africa, which shows 66% radiographic evidence of metastatic disease and Nigeria which presents 88.9% of locally advanced or metastatic disease, the prevalence within our study population was much lower (17,18).

Several studies have shown that the type of health insurance coverage may be an important determinant of cancer stage at diagnosis and appears to have a unique effect on the quality life of men after treatment for PCa (19). These findings are crucial for clinicians and policy makers who need to evaluate how health insurance status might affect outcomes in PCa incidence and management (20). This is particularly relevant as many African populations, especially those in poor rural communities, deal with poor access to quality and affordable healthcare (18).

In order to exclude delayed diagnosis as the only cause of the incidence results, we performed, a subgroup analysis of patients younger vs older than 70 years. The older patients presented with high PSA levels, clinically advanced disease and high rates of intermediate and high-risk disease, and these findings were not significantly different from those of the analysis conducted on the younger group. This might be due to either the size of the overall cohort or because the number of PCa patients older than 70 years is larger, hence the need for a larger cohort. Another subgroup analysis of patients with two different healthcare systems was performed. The analysis results of patients who presented with RAMED insurance and advanced clinical characteristics were not significantly different from the results of the group with Mutual

healthcare insurance. Younger patients and those with RAMED insurance presented intermediate and high-risk disease despite their earlier age of diagnosis or their health insurance coverage type. All patients with RAMED insurance are from a low socioeconomic level which explains the advanced stage of the diagnosis patients presented with this type of insurance.

Additional analyses were performed on patients from the region of Fez city because it has the largest sample size compared to the patients from the region of Oujda city. We compared the percentage of patients diagnosed with PCa from 2015 to 2021. We identified a significant decrease in the percentage of patients diagnosed with PCa in 2020 and 2021 versus 2015 to 2019. This finding shows the effects of the COVID-19 pandemic on the diagnosis of patients with PCa (21,22). Several studies reported the impact that efforts to control the pandemic have had on routine cancer care encounters, either against the COVID-19 infection risks to patients and medical staff or concerning the number of daily hospital trips and admissions that have been minimized (21,22).

This epidemiological study had some limitations including the absence of many patients' health information and the relatively small sample size. More complete data collection must be performed and made available to clinicians and decision makers in order to obtain an accurate and complete overview of the Moroccan PCa epidemiology profile. Furthermore, having collected data from two institutions may not reflect the exact figures about PCa in the whole country. However, it gives an idea about the low incidence of PCa in Morocco, as well as the late presentation of this tumor in a considerable number of patients.

V. Conclusions

In this study, we aimed to describe the epidemiological profile of Moroccan PCa. It was performed using data from the last six years, relative to 141 patients who were diagnosed with PCa from two university hospitals. Prostate cancer is being grossly underdiagnosed in these populations. An extensive epidemiological study of this cancer is required in order to provide detailed data on the presentation of PCa in the Moroccan population. This study showed that Moroccan men presenting PCa tend to have intermediate and high-grade disease at diagnosis and delayed diagnosis. This study represents a step forward in addressing the need for further research and routine screening for the early detection of this disease. To to improve PCa

outcome and survival, screening remains the cornerstone of PCa control in the population. Public awareness of PCa will enable men to recognize early symptoms and seek medical help before the disease advances. Given its specificities and increasing importance, we suggest establishing a dedicated National Prostate Cancer Registry to better understand the characteristics of this disease in Moroccan men and further improve the health outcome of PCa patients in the country.

Furthermore, there are some specificities of epidemiologic studies of prostate cancer that should be taken into consideration when conducting such studies, especially in African countries. There are several methodological challenges including a lack of metadata on the hospital registries. Hence, future studies of prostate cancer epidemiology should focus more on high-grade and advanced-stage prostate cancer and their clinical relevance, since most of the cases in Africa are diagnosed at an advanced stage. Moreover, epidemiologic studies of prostate cancer should be designed with consideration of potential biases specific to this disease. One of these biases is the ascertainment of information on PSA screening. The medical record of each patient should be complete and updated with all the clinical and molecular information to have useful metadata for future epidemiological studies.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin [Internet]. 2021;0(0):1–41. Available from: doi: 10.3322/caac.21660
2. Register-Rabat. Incidence of Cancers in Rabat (2006-2008). Cancer Regist Rabat. 2012;
3. Register-Casablanca. Register of Cancers of Casablanca (2005-2006-2007. 2016;
4. Pernar CH, Ebot EM, Wilson KM, Mucci LA. The Epidemiology of Prostate Cancer. Cold Spring Harb Perspect Med. 2018;8(12):1–19.
5. Bostwick DG, Burke HB, Djakiew D, Euling S, Ho SM, Landolph J, et al. Human prostate cancer risk factors. Cancer. 2004;101(10 SUPPL.):2371–490.
6. high commissariat of planning. Regions_Morocco, High commissariat of planning [Internet]. Available from: https://www.hcp.ma/downloads/
7. Mark K. Buyyounouski, Peter L. Choyke, FACR, Jesse K. McKenney, Oliver Sartor, Howard M. Sandler, FASTRO, Mahul B. Amin, Michael W. Kattan and DWL. Prostate Cancer – Major Changes in the American Joint Committee on Cancer Eighth Edition Cancer Staging Manual. CA Cancer J Clin. 2017;67(3):245–253.
8. Manager of Mandatory Health Insurance (2016). Manager of Mandatory Health Insurance (2016). “National fund of social welfare organisations”. Archived from the original on 2016-10-15. Retrieved.
9. Access Health International “The RAMED project”2016. Cit 2. Access Health International. “The RAMED project”. Archivedfrom the original on 2016-10-28. Retrieved 2016-10-28.
10. Culp MBB, Soerjomataram I, Efstathiou JA, Bray F, Jemal A. Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates. Eur Urol [Internet]. 2020;77(1):38–52. Available from: https://doi.org/10.1016/j.eururo.2019.08.005
11. Haimer A, Habib F, Soulaymani A, Mokhtari A, Hami H. Prostate Cancer in Morocco: The Epidemiological Evidence. J Glob Oncol [Internet]. 2018 Sep 28;4(Supplement 2):43s-43s. Available from: https://doi.org/10.1200/jgo.18.41500
12. Kevin H. Kensler TRR. Cancer progress and priorities: Prostate cancer. Cancer Epidemiol Biomarkers Prev. 2020;29(2):267–277.
13. Patel AR, Klein EA. Risk factors for prostate cancer. Nat Clin Pract Urol. 2009;6(2):87–95.
14. Haas GP, Delongchamps N, Brawley OW, Wang CY, de la Roza G. The worldwide epidemiology of prostate cancer: perspectives from autopsy studies. Can J Urol. 2008;15(1):3866–71.
15. Oluwole OP, Rafindadi AH, Shehu MS, Samaila MOA. A ten-year study of prostate cancer specimens at Ahmadu Bello University Teaching Hospital (A.B.U.T.H), Zaria, Nigeria. African J Urol [Internet]. 2015;21(1):15–8. Available from: http://dx.doi.org/10.1016/j.afju.2014.09.004
16. Ali M. Clinical Presentation, Pathological Pattern and Treatment Options of Prostate Cancer at Al-Azhar University Hospitals Over the Last 30 Years. African J Urol. 2011;17(4):135–40.
17. Ajape A, Ibrahim KOO, Fakeye JA, Abiola OO. An overview of cancer of the prostate diagnosis and management in Nigeria: The experience in a Nigerian tertiary hospital. Ann Afr Med. 2010;9(3):113–7.

18. H A Le Roux , R J Urry , B Sartorius CA. Prostate Cancer at a regional hospital in South Africa: we are only seeing the tip of the iceberg. SAJS [Internet]. 2015;53(3&4):1–6. Available from: https://doi.org/10.1093/jnci/91.16.1409
19. Richard G. Roetzheim, Naazneen Pal, Colleen Tennant, Lydia Voti, John Z. Ayanian, Annette Schwabe JPK. Effects of health insurance and race on early detection of cancer. J Natl Cancer Inst [Internet]. 1999;91(16):1409–1415. Available from: https://doi.org/10.1093/jnci/91.16.1409
20. Penson DF, Stoddard ML, Pasta DJ, Lubeck DP, Flanders SC, Litwin MS. The association between socioeconomic status, health insurance coverage, and quality of life in men with prostate cancer. J Clin Epidemiol. 2001;54(4):350–8.
21. De Vincentiis L, Carr RA, Mariani MP, Ferrara G. Cancer diagnostic rates during the 2020 “lockdown”, due to COVID-19 pandemic, compared with the 2018-2019: An audit study from cellular pathology. J Clin Pathol. 2021;74(3):187–9.
22. Obek C, Doganca T, Argun OB, Kural AR. Management of prostate cancer patients during COVID-19 pandemic. Prostate Cancer Prostatic Dis [Internet]. 2020;23(3):398–406. Available from: http://dx.doi.org/10.1038/s41391-020-0258-7

Chapter 3: Epigenetic Analysis of Circulating Tumor DNA in Prostate Cancer: Evaluation of GSTP1 Methylation

Chapter 3: Epigenetic Analysis of Circulating Tumor DNA in Prostate Cancer: Evaluation of GSTP1 Methylation

I. Introduction

Prostate cancer (PCa) is the most frequently diagnosed cancer among men in 112 countries, and it is the leading cause of cancer-related deaths among men in more than 40 countries (1).

Cancer presents many genetic alterations such as DNA copy number variants and epigenetic alterations such DNA methylation, which have become a major focus of investigation in prostate cancer research (2,3) (Figure 3.1).

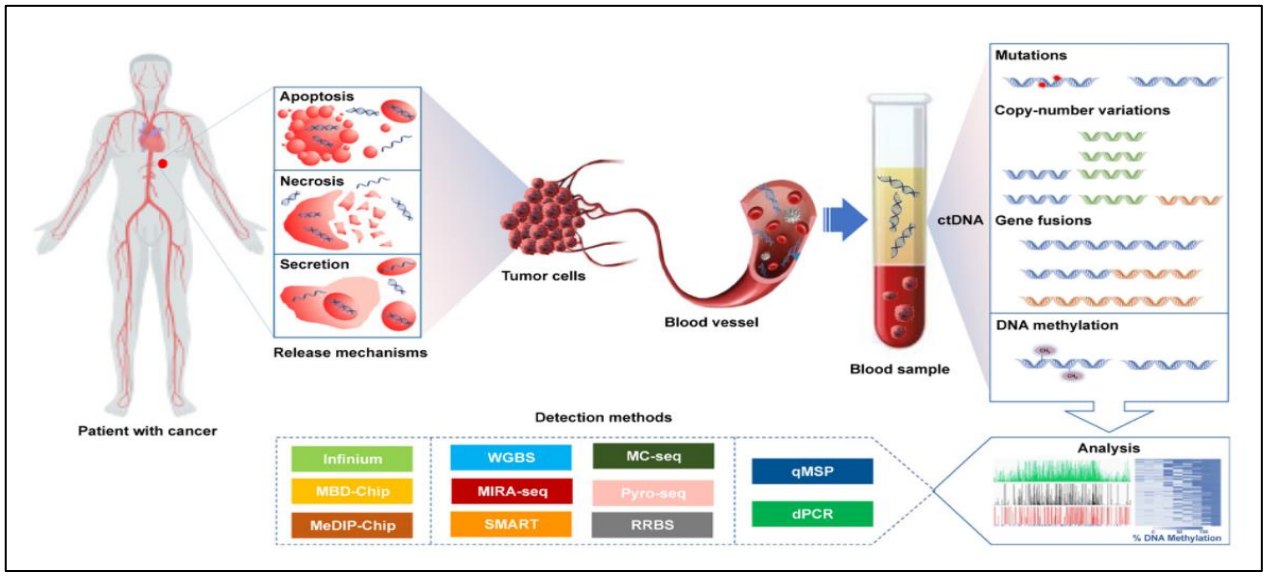


Figure 3. 1: Origins of Cell Free Tumor DNA (ctDNA) and Various Genomic Alterations Utilized in Liquid Biopsies. As illustrated, peripheral blood is the most common source of liquid biopsy (other body fluids can also be used). ctDNA released by apoptotic or necrotic tumor cells can be isolated from a simple blood draw and the ctDNA can be analyzed to identify multiple common DNA-based alterations observed in tumors, including mutations, copy-number variations, gene fusions, and DNA methylation changes. As shown at the bottom, specialized methods for ctDNA methylation analysis have been developed. Abbreviations: dPCR, digital PCR; Infinium, Infinium Human Methylation BeadChips; MBD, methyl-CpG-binding domain; MC-seq, methyl-capture sequencing; MeDIP, methylated DNA immunoprecipitation; MIRA-seq, methylated-CpG island recovery assay combined with sequencing; Pyro-seq, pyrosequencing; qMSP, quantitative methylation-specific PCR; RRBS, reduced-representation bisulfite-sequencing; SMART, single molecule, real-time sequencing; WGBS, whole-genome bisulfite sequencing (4)

In prostate cancer, DNA methylation alterations have been known and widely studied for decades and seem to be among the earliest and most common events in tumorigenesis and the most recurrent events in both locally advanced and metastatic prostate tumors (2,5–7). Moreover, these events impact gene expression potential (7,8). Many gene promoters contain

CpG islands (GC-rich regions of DNA) including the *GSTP1* gene. Methylation of CpG islands may result in downregulation of transcriptional activity via local effects on DNA-binding proteins and alteration in chromatin structure. Abnormal methylation of genes, including tumor suppressor genes has been reported in a large type of human cancers (9). *GSTP1* is expressed in normal prostate epithelial tissue and it has been found that this gene could be somatically inactivated by hypermethylation of its promoter region. The methylation of this gene is the most common molecular alteration described in prostate cancer (2,7).

Similar to most cancers, the detection of PCa at an early stage is key for successful treatment and better outcomes. Therefore, the alterations in DNA methylation in cancer have been revealed to be promising targets for the development of powerful diagnostic, prognostic, and predictive biomarkers and tools for various cancer types notably, prostate cancer because there is an inverse correlation between survival rates and prostate tumor aggression, despite advances in diagnostic, prognostic and treatment strategies (10–14).

Currently, it is easier to integrate DNA methylation assays into routine clinical practice due to their similarity with DNA-sequence-based biomarker approaches (15). ConfirmMDx (MDxHealth, Irvine, CA, USA) is a validated epigenetic test designed for diagnostics that is commercially available, and uses the methylation profile of three genes (*APC, RASSF1*, *GSTP1*) to detect cancer in histologically negative biopsies. In addition to 14 methylation-based biomarker assays that are also currently commercially available and are designed to measure the methylation of only 13 genes in total: bone morphogenetic protein 3 (*BMP3*); septin-9 (*SEPT9*); glutathione *S*- transferase P (*GSTP1)*; adenomatous polyposis coli protein (*APC*); ras association domain- containing protein 1 (*RASSF1*); N- myc downregulated gene 4 (*NDRG4*); short stature homeobox protein 2 (*SHOX2*); twistrelated protein 1 (*TWIST1*); branched- chain-amino- acid aminotransferase, cytosolic (*BCAT1*); homeobox protein OTX1 (*OTX1*); one cut domain family member 2 (*ONECUT2*); methylated DNA-protein-cysteine methyltransferase (*MGMT*); and DNA- binding protein Ikaros (*IKZF1*). Only nine of these 14 assays have been included in clinical guidelines (*GSTP1*, *APC, RASSF1*, *NDRG4*, *BMP3*, two *SEPT9*, *SHOX2*, and *MGMT*) (11,16–18).

PCa-derived aberrant DNA methylation patterns have also been detected in bodily fluids such as blood and urine samples, which paved the way for the development of urine-based assays that integrate GSTP1 biomarker (19–23).

Evaluation of the methylation status of the *GSTP1* gene may be cost-effective and could be

used extensively for cancer diagnosis and prevention. Recent progress has provided extraordinary opportunities to identify prognostic and predictive markers of efficacy to chemotherapy. Genetic markers can be used to identify patients who will benefit from therapy, exclude patients at high risk of developing severe toxicity, and adjust dosing (24).

We investigated *GSTP1* promoter hypermethylation in DNA isolated from plasma, from 15 patients with prostate cancer. Via this pilot study, we assessed the feasibility of a plasma test for prostate cancer detection in a high-risk patient cohort based on methylation-specific PCR analysis of the *GSTP1* gene promoter methylation status in the Moroccan clinical routine.

II. Background

1. Cell-free DNA as biomarkers

Currently, circulating cell-free DNA represents a promising predictive tool to monitor cancer progression and treatment outcome, especially for diseases that are hard to cure, such as the metastatic castrate-resistant prostate cancer (25).

Cell-free circulating DNA was demonstrated to be increased and at higher levels in cancer patients than in normal controls and subjects with benign prostatic hyperplasia (BPH), so it may be useful as a prognostic, diagnostic and/or predictive tool in prostate cancer (25–28).

Such markers could be used instead of invasive detection of prostate cancer and could minimize the monitoring of disease pathogenesis which is attractive to clinicians and scientists. However, reports vary widely in their estimations of the proportion of circulating DNA in cancer patient blood that is derived from tumors, since DNA levels in plasma reflect a number of many cellular processes, including circulating tumor cells (19,29,30).

The discovery of cell-free circulating DNA in blood, plasma and serum was reported by Mandel and Métais in 1948 (31). Moreover, cell-free circulating DNA has been correlated with the type of tumor stage, Gleason score, metastatic status and treatment regimens, among many other factors (28,30,32,33).

In contrast to mRNA, circulating DNA-based tumor markers exhibit greater stability and enhanced tumor specificity, and alterations in cell-free circulating DNA have been shown to match mutations present in primary tumors. Hence, circulating DNA-based is a potential tool warranting tumor grading/staging, prognostic estimation and aiding therapeutic decision-making (29).

Circulating cell-free DNA remains expressing the same behaviors as mutations and represents a low invasive technique with high specificity, sensitivity and clinical usability. In this respect, circulating cell-free DNA has emerged as an interesting marker to stratify the disease. Due to high-throughput technologies, liquid biopsy on circulating cell-free DNA will soon be utilized in the clinical management and therapy response monitoring of prostate cancer patients (i.e metastatic castrate-resistant prostate cancer) (25).

2. GSTP1

The glutathione S-transferase pi gene (*GSTP1*) is the most well-studied DNA methylation biomarker of PCa, particularly in diagnosis (34). It encodes a detoxifying enzyme and tumor suppressor, glutathione S-transferase, which is involved in drug metabolism and DNA protection from oxidative damage (Figure 3.2) (35). Hypermethylation of *GSTP1* is observed frequently in PCa tissue but rarely in normal prostate tissue (36). Several studies have investigated the prognostic value of *GSTP1* methylation. Importantly, some of these studies did not find any prognostic value in *GSTP1* methylation in the prediction of biochemical recurrence BCR (as the clinical endpoint) low vs. high Gleason score (GS) cancers or PCa death (37–47).

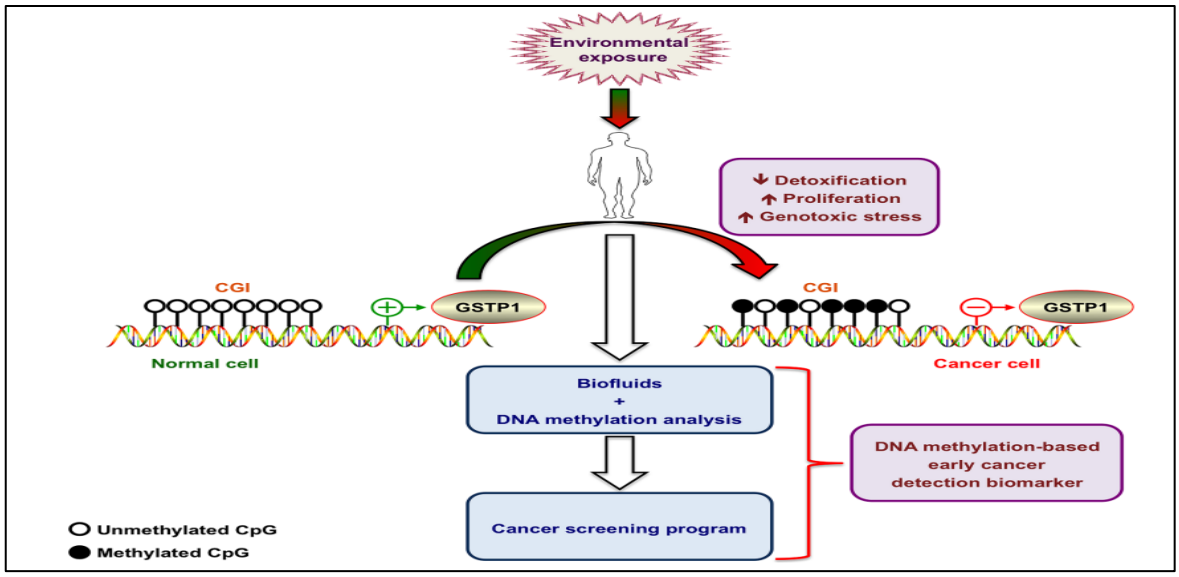


Figure 3. 2: Hypermethylation of the glutathione S-transferase P1 promoter as an early cancer biomarker. Environmental exposure leads to progressive methylation of CpG islands (CGI) at the GSTP1 promoter region. The resulting decreased cell detoxification capacity leads to increased genotoxic stress and progressive accumulation of additional genetic alterations accompanied by an increased proliferation rate. DNA methylation analysis of the GSTP1 promoter in body fluids allows detecting progressive hypermethylation and turns this molecular feature into a valid early cancer biomarker. Figure was generated by using ScienceSlides (48).

Rosenbaum et al. (2005) assessed *GSTP1* methylation using qMSP in a cohort of GS 7 (3 + 4) patients and reported that *GSTP1* hypermethylation was significantly associated with time to progression in univariate analysis, and as an independent predictor in multivariate analysis with other candidate genes (37). Moreover, Maldonado et al. (2014) showed that *GSTP1* methylation was a significant independent prognostic factor of progression (38). Litovkin et al. (2015) found trichotomised *GSTP1* methylation to be an independent prognostic predictor of high-risk PCa patients (40).

3. Hypermethylation events

Several studies have shown the potential for non-invasive monitoring of DNA methylation, which is a key feature observed during early tumorigenesis and occurs at a much higher frequency and more consistently than genetic monitoring, as well as in its progression and metastatic development, in cell-free DNA to detect prostate cancer. Notably, methylation of *GSTP1* is present at all stages of prostate tumorigenesis (44,49–55).

DNA methylation occurs when a methyl group is added to cytosine, which is followed by guanine. This is known as a CpG dinucleotide. It is associated with gene regulation (i.e., gene silencing) and function (56).

To date, DNA methylation at CpG dinucleotides is the most widely studied epigenetic alteration; 5-cytosine DNA methylation is highly concentrated in CpG islands within the promoter region of genes, and promoter hypermethylation in cancers has been linked to the silencing of tumor suppressor genes and subsequent oncogenesis (57,58) (figure 3.3).

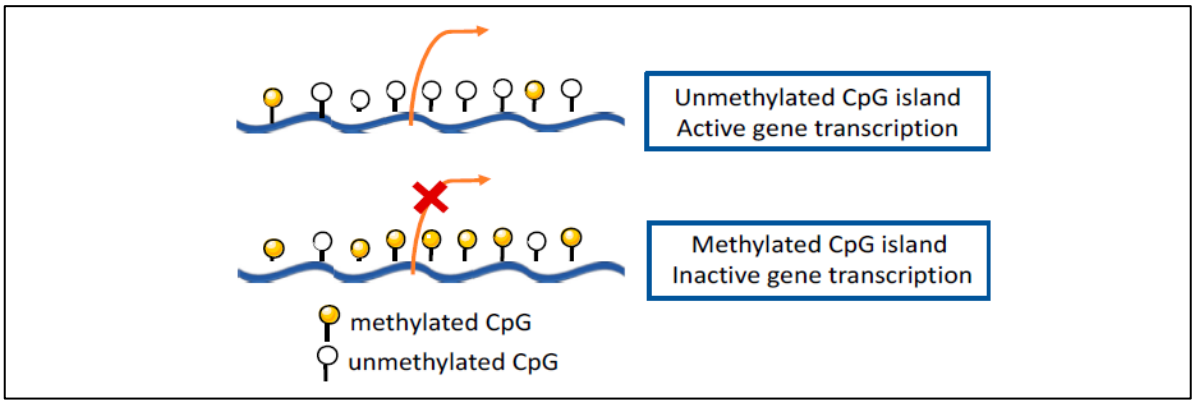


Figure 3. 3: DNA methylation within a gene promoter region. Unmethylated CpG islands enable gene transcription. When CpG islands are methylated, gene transcription is repressed (59)

In prostate cancer, gene promoter hypermethylation is one of the three types of DNA alterations that have been investigated as plasma/serum biomarkers (28,32,60–67). These represent

potentially attractive biomarkers, and methylation-specific PCR (MSP) is used for the detection of gene promoter hypermethylation since it is a highly sensitive application, requiring only 0.1–0.001% of serum DNA to be of tumor origin (49).

III. Materials and Methods

In this section, a detailed description of the methods used to perform the different experiments will be given starting from the cohort constitution to the circulating DNA extraction, bisulfite treatment, MSP, and interpretation of the results obtained.

1. Sample collection

Cohort constitution: We recruited patients from consenting persons in prostate cancer from University Hospital Center Hassan II of Fez. Blood samples were obtained prior to surgery, radiation therapy or chemotherapy and at least 6 weeks after transrectal prostate biopsies. Only patients who signed the consent form were part of this study, which was approved by the Ethics Committees of the University Hospital Hassan II.

2. Nucleic Acid Isolation

After informed consent was obtained, 15 patients with histologically confirmed adenocarcinoma of prostate cancer were enrolled in this study. Blood samples were collected. Plasma was separated from the cellular fraction by centrifugation twice at 2000 rpm for 10 min at 4°C. Circulating DNA was isolated using the Magmax Cell Free DNA Isolation kit (Thermo Fisher Scientific), eluted in 15 µl of MagMAX Cell Free DNA Elution Solution and stored at −20°C.

3. Bisulfite treatment

Circulating DNA was modified by bisulfite treatment (Thermo Scientific™ EpiJET™ Bisulfite Conversion Kit) (Appendix 3), resulting in the generation of altered DNA sequences, because in the bisulfite reaction all unmethylated cytosines are deaminated and converted to uracils, while methylated cytosines remain unchanged. Bisulfite converted unmethylated cytosines are detected as thymines in the following PCR. Basically, 120 µl of prepared Modification Reagent

solution (which was prepared by adding 0.9 ml of molecular biology grade water to 200 µl of Modification Solution I and 60 µl of Modification Solution II to one vial of dry Modification Reagent) was added to 20 µl of DNA sample in a PCR tube. PCR reaction (98°C/ 10 min; 60°C/ 150 min) was performed. The converted DNA was loaded into 400 µl of Binding Buffer in the column, centrifuged at 12,000 rpm for 30 sec, and then washed once with 200 µl of Wash buffer. Desulfonation Buffer (10 ml of 96-100% ethanol in 3.5 mL of Desulfonation Buffer) was added to the column and incubated for 20 min at room temperature, followed by two washes with 200 µl Wash Buffer and 700 µl RPE, with 30 sec 12,000 rpm spin between each wash. DNA was eluted in 10 µl of Elution Buffer in a fresh 1.5 mL Eppendorf tube (Figure 3.4).

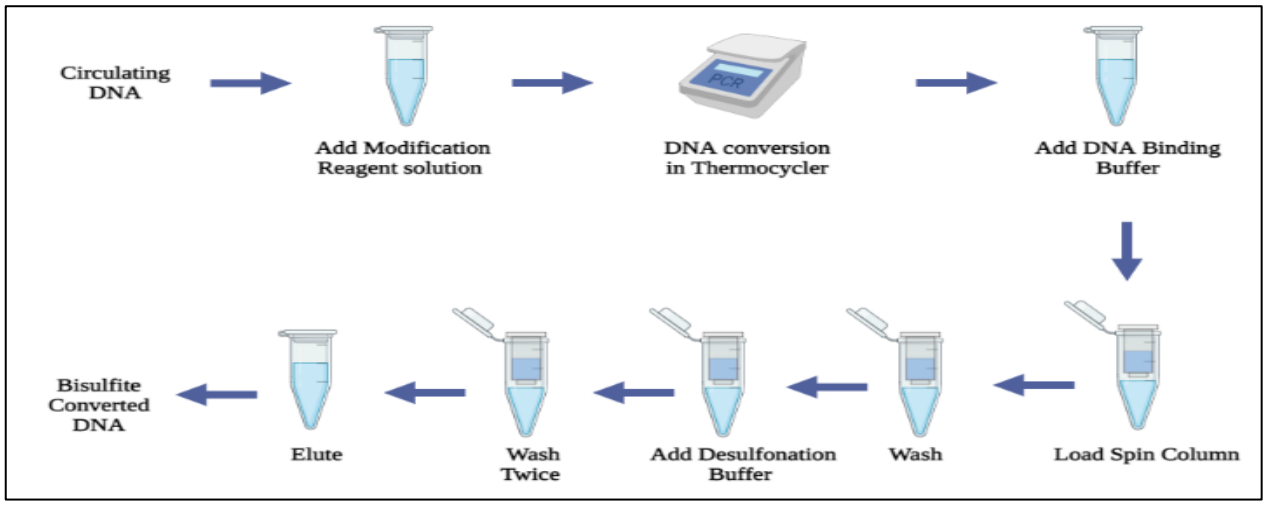


Figure 3. 4: Flow chart of the Bisulfite Conversion Reaction

4. Methylation Specific PCR (MSP)

Methylation-specific PCR (MS-PCR or MSP) is one of the most commonly used methods for the gene/sequence-specific detection of DNA methylation. The DNA undergoes bisulfite conversion of cytosine to uracil and then the methylated sequences are selectively amplified with primers specific for methylation.

In our study, the modified DNA was subjected to MSP targeting a short (methylated reaction, 92 bases; unmethylated reaction, 99 bases) promoter sequence of the human *GSTP1* gene primer set using two primer pairs: one that detected unmethylated alleles and one that detected methylated alleles:

- GSTP1 unmethylated specific primer:

5'-GAT GTT TGG GGT GTA GTG GTT GTT-3' _ (forward);

5’-CCA CCC CAA TAC TAA ATC ACA ACA-3’ _ (reverse)

- GSTP1 methylated-specific primer:

  5′-TTC GGG GTG TAG CGG TCG T-3′_ (forward);

  5′-GCC CCA ATA CTA AAT CAC GAC G-3′_ (reverse)

Primers for MSP were designed using PrimerBlast (National Center for Biotechnology Information (NCBI), Bethesda, MD). Primer pairs were selected for a length of 15-21 base pairs, PCR product length of 100-150 bp, melting temperature 58-62ºC, G/C content 20-70%, and zero 3’ self-complementarity. Human specificity for each sequence was verified using BLAST (NCBI). Primers were produced by Biolegio. Bisulfite PCR primer design is critical for successful implementation of subsequent bisulfite sequencing analysis. Therefore, we have carefully optimized the PCR.

For PCR amplification, 1 µl of bisulfite modified DNA was added to a final volume of 25 µl of PCR mix containing 2.5 µl of 10X PCR buffer -Mg, 0.5 µl of dNTP Mix (10 mM), 0.2 µl of 5 unit Recombinant Taq DNA polymerase (Life Technologies, Invitrogen) and 2µl of MgCl2 (50 mM), 9.8 µl pure water and 2 µl of each primer (forward and reverse) (Table 3.1). Primer sequences for either methylated or unmethylated *GSTP1* have been described previously. Pure water served as a negative control. PCR was carried out using the following conditions: 60°C was the annealing T° for the methylated alleles and 58°C was the annealing T° for the unmethylated alleles (Table 3.2).

Table3. 1: PCR mix reaction of the Methylated and Unmethylated alleles

10X PCR buffer	2.5 µl
MgCl2 solution (50 mM)	2 µl
dNTP mixture (10 nM)	0.5 µl
Taq DNA polymerase (5 unit)	0.2 µl
Forward primer	2 µl
Reverse primer	2 µl
1µl Bisulfite modified DNA	
Nuclease/RNase free water	Up to 25 µl

Table3. 2: PCR conditions of the Methylated and Unmethylated alleles

	Methylated		Unmethylated		
	95°C	3 min	95°C	3 min	
X55 cycle	95°C	30 sec	95°C	30 sec	X55 cycle
	60°C	30 sec	58°C	30 sec	
	72°C	1 min	72°C	1 min	
	72°C	9 min	72°C	9 min	
	4°C	∞	4°C	∞	

PCR products were directly loaded into a 2% agarose gel (Appendix 2), stained with ethidium bromide, and run at 120V for 20-30 min. The samples were visualized under UV illumination.

5. Clinical Characteristics

The following table presents the principal clinical and pathological characteristics of our 15 prostate cancer patients:

Table3. 3: Clinical Characteristics of Prostate Cancer Patients

Patient num	Age	Gleason score	PSA (ng/ml)	Histological Type
P01	62	4+5=9	>100	Adenocarcinoma
P02	62	4+4=8	80	Adenocarcinoma
P03	72	4+3=7	418	Adenocarcinoma
P04	64	4+4=8	637	Adenocarcinoma
P05	65	4+4=8	>100	Adenocarcinoma
P06	73	4+3=7	300	Adenocarcinoma
P07	71	4+4=8	>100	Adenocarcinoma
P08	74	4+3=7	64	Adenocarcinoma
P09	67	4+4=8	>100	Adenocarcinoma
P10	77	4+4=8	>100	Adenocarcinoma
P11	79	4+4=8	50	Adenocarcinoma
P12	66	4+4=8	90	Adenocarcinoma
P13	65	4+4=8	>100	Adenocarcinoma
P14	69	4+3=7	200	Adenocarcinoma
P15	75	4+3=7	70	Adenocarcinoma

IV. Results and Discussion

Figure 3 represents the agarose gel profile of the MSP screening of *GSTP1* methylation. We have 4 gel profiles, each one of them represents the results of *GSTP* methylation in 4 patients. The first lane in the first gel profile represents the ladder. The last lane in each gel represents the negative control. We ran two lanes on the gel for each patient: unmethylated *GSTP1* and methylated *GSTP1*. Analysis of the PCa plasma samples showed that *GSTP1* is methylated in 14 out of 15 samples (94%). In all patients with a Gleason Score >=7 plasma samples were positive for *GSTP1* promoter hypermethylation. However, the only patient who was negative for *GSTP1* promoter hypermethylation had a Gleason score of 9 (4+5).

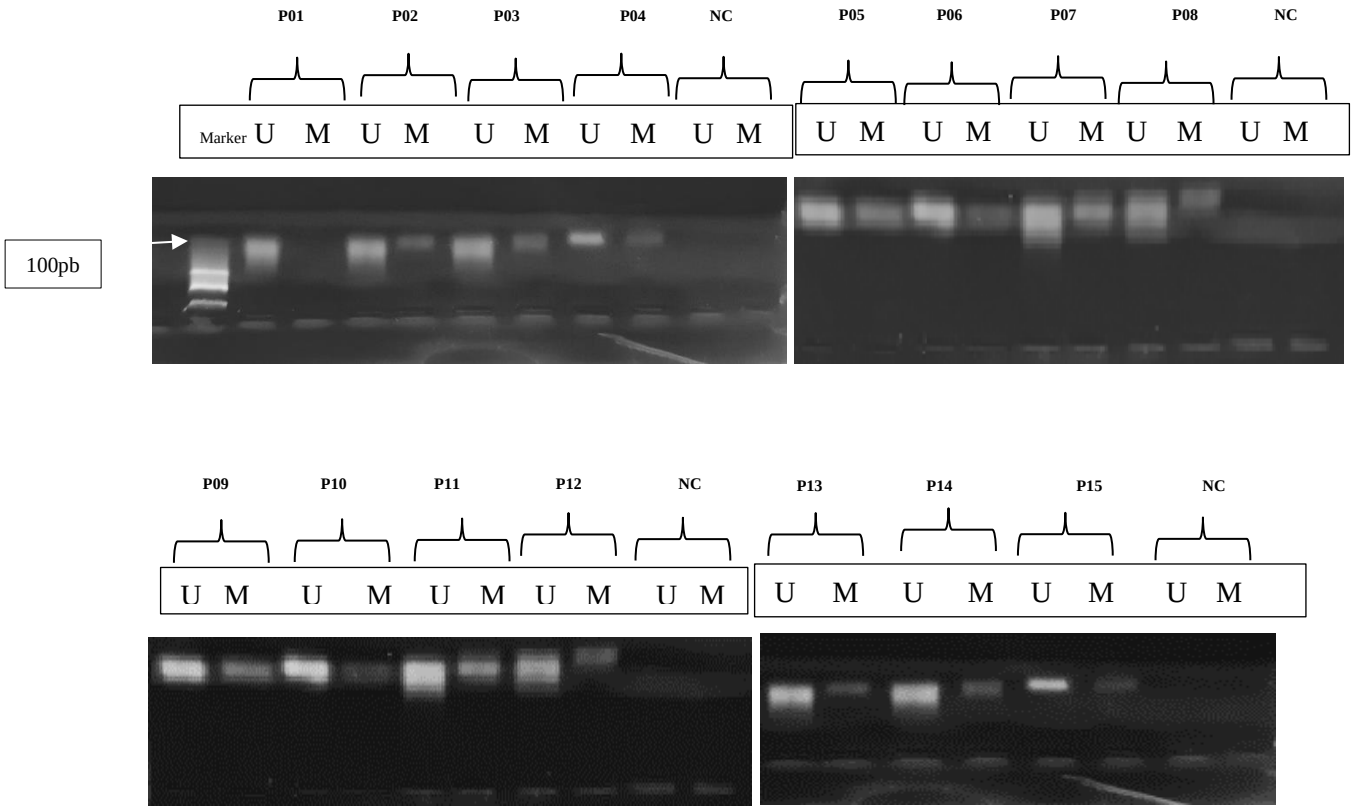


Figure 3. 5: Agarose gels showing MSP screening of GSTP1 methylation: analysis of GSTP1 in the Generation of PCR product indicates presence of unmethylated (U) or methylated (M) GSTP1 alleles for our 15 patients. NC is the Negative Control (blank _ water contamination control).

Since cell-free tumor DNA can be detected in bodily fluids of various cancer patients we investigated *GSTP1* promoter hypermethylation in circulating DNA isolated from the plasma of 15 patients with prostate cancer by MSP. This technique has been successfully applied for molecular tumor detection in different types of cancers (68–72). In most studies on the occurrence of tumor DNA in the plasma or serum, the number of healthy controls or patients

with benign diseases was zero (69) or small (72) as in our study. We found 14 (94%) of 15 adenocarcinomas positive for *GSTP1* methylation (Figure 5). All of these patients had either a Gleason score of 7 (4+3) or 8 (4+4). The only patient identified as negative for *GSTP1* methylation had a Gleason score of 9(4+5).

This result is in accordance with literature that confirmed that the finding of circulating prostate cancer DNA in plasma/serum was stage-dependent (73) and found that *GSTP1* is epigenetically silenced in approximately 90 to 95% of adenocarcinomas by somatic hypermethylation of its regulatory CpG island (7,74,75). Our findings confirm the role of *GSTP1* as a biomarker of early-stage prostate cancer and can be used in the diagnosis of this disease. Our results confirm that *GSTP1* promoter hypermethylation is a highly prevalent event in prostate cancer and its analysis provides a highly specific tool for the DNA-based diagnosis of prostate cancer in plasma (76). However, *GSTP1* hypermethylation was higher than that observed in some studies; it was 94% in plasma; however, this percentage was found in the tissue in previous studies and only approximately 72% was found in plasma (49,73). This could be attributed to our small sample size.

GSTP1 methylation has been proposed by many research papers as an epigenetic marker for early prostate cancer diagnosis (36) and it is certainly the most widely studied epigenetic marker in prostate cancer in the last decade (77). *GSTP1* is associated with different tumors, as it is involved in the regulation of the cell cycle, and overexpression is linked to drug resistance (78,79). Moreover, with regard to prostate cancer, extensive literature data have demonstrated its hypermethylation (36,77,80–82). In the following years a number of papers and reviews have been published on the role of *GSTP1* methylation as a potential diagnostic marker (34,52). Even though the methylation of *GSTP1* is a good marker for PCa, it requires validation and therefore is not used in current diagnosis and screening techniques. The present study aimed to validate the role of *GSTP1* methylation. Goessl et al. (2001) (73) investigated GSTP1 promoter hypermethylation in the plasma and/or serum of prostate cancer patients and controls with BPH. They found that the detection of circulating prostate cancer DNA in plasma/serum was stage-dependent and the rate of this detection was only 56% in localized T2–T3N0M0 prostate cancer (73).

Notably, *GSTP1* appears to be methylated and silenced in patients with low grade (≤6) and low stage (T2a), confirming its involvement in early carcinogenesis and suggesting that *GSTP1* alterations (methylation or expression) may be considered a useful early diagnostic markers, so as to avoid unnecessary biopsies, as recently demonstrated by Zelic et al. (2016) (83). *GSTP1* behaviour in the various stages of tumor development is interesting and should be better investigated. However, because no association between the *GSTP1* genotype and cancer stage in prostate cancer patients has been found (84), this potential limitation is obviated. The small cohort and the lack of long-term follow-up are the potential limitations of our study. Notably, *GSTP1* promoter hypermethylation is a specific feature of cell-free tumor DNA derived from prostatic adenocarcinoma and might become a valuable DNA-based tumor marker for molecular staging in men with prostate cancer (76). Currently, the detection of cancer in its early stages represents the most effective way to improve the clinical management of this disease (85).

V. Conclusion

To the best of our knowledge, this is the first study to report the presence/absence of *GSTP1* hypermethylation in Moroccan prostate cancer subjects. Our study investigated hypermethylation of the *GSTP1* promoter region in Moroccan prostate cancer patients. Our results suggest that blood *GSTP1* methylation is a promising source for a non-invasive biomarker that may therefore be used to stratify prostate cancer patients with negative initial biopsy results. The efficacy of this noninvasive molecular test for diagnosis and prognosis of prostate cancer could be included in the clinical management routine of this disease. For many years, DNA methylation has attracted attention in basic cancer research, as being readily detectable in body fluids by MSP, by promising new putative cancer markers for the diagnosis, prognosis and prediction of therapy outcome in PCa.

In conclusion, we assert that *GSTP1* represents a typical example of an ideal epigenetic “liquid biopsy biomarker”, as its methylation is involved in prostate cancer and could be detected with good results in circulating cell-free DNA by MSP.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209–49.
2. Nelson G, Isaacs B, Steven G. CG Island Methylation Changes Near the GSTP1 Gene in Prostatic Carcinoma Cells Detected Using the Polymerase Chain Reaction: A New Prostate Cancer Biomarker. 1997;6(June):443–50.
3. Rennie PS, Nelson CC. Epigenetic mechanisms for progression of prostate cancer. Cancer Metastasis Rev. 1998;17(4):401–9.
4. Roy D, Tiirikainen M. Diagnostic Power of DNA Methylation Classifiers for Early Detection of Cancer. Trends in Cancer [Internet]. 2020;6(2):78–81. Available from: https://doi.org/10.1016/j.trecan.2019.12.006
5. Friedlander TW, Roy R, Tomlins SA, Ngo VT, Kobayashi Y, Azameera A, et al. Common structural and epigenetic changes in the genome of castration-resistant prostate cancer. Cancer Res. 2012;72(3):616–25.
6. Bedford MT, van Helden PD. Hypomethylation of dna in pathological conditions of the human prostate. Cancer Res. 1987;47(20):5274–6.
7. Lee WH, Morton RA, Epstein JI, Brooks JD, Campbell PA, Bova GS, et al. Cytidine methylation of regulatory sequences near the π-class glutathione S-transferase gene accompanies human prostatic carcinogenesis. Proc Natl Acad Sci U S A. 1994;91(24):11733–7.
8. Herman JG, Lapidus RG, Issa JPJ, Davidson NE, Sidransky D, Baylin SB, et al. Inactivation of the CDKN2/pl6/MTS1 Gene Is Frequently Associated with Aberrant DNA Methylation in All Common Human Cancers. Cancer Res. 1995;55(20):4525–30.
9. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. Nat Rev Genet. 2002;3(6):415–28.
10. Wallace TJ, Torre T, Grob M, Yu J, Avital I, Brücher BLDM, et al. Current approaches, challenges and future directions for monitoring treatment response in prostate cancer. J Cancer. 2014;5(1):3–24.
11. Koch A, Joosten SC, Feng Z, De Ruijter TC, Draht MX, Melotte V, et al. Analysis of DNA methylation in cancer: Location revisited. Nat Rev Clin Oncol [Internet]. 2018;15(7):459–66. Available from: http://dx.doi.org/10.1038/s41571-018-0004-4
12. Hao X, Luo H, Krawczyk M, Wei W, Wang W, Wang J, et al. DNA methylation markers for diagnosis and prognosis of common cancers. Proc Natl Acad Sci U S A. 2017;114(28):7414–9.
13. Board RE, Knight L, Greystoke A, Blackhall FH, Hughes A, Dive C, et al. DNA Methylation in Circulating Tumour DNA as a Biomarker for Cancer. Biomark Insights. 2007;2:117727190700200.
14. Gai W, Sun K. Epigenetic biomarkers in cell-free DNA and applications in liquid biopsy. Genes (Basel). 2019;10(1).
15. Bock C, Halbritter F, Carmona FJ, Tierling S, Datlinger P, Assenov Y, et al. Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. Nat Biotechnol. 2016;34(7):726–37.
16. Stewart GD, Van Neste L, Delvenne P, Delrée P, Delga A, McNeill SA, et al. Clinical utility of an epigenetic assay to detect occult prostate cancer in histopathologically negative biopsies: Results of the MATLOC study. J Urol [Internet]. 2013;189(3):1110–6. Available from: http://dx.doi.org/10.1016/j.juro.2012.08.219

17. Carroll PR, Parsons JK, Andriole G, Bahnson RR, Castle EP, Catalona WJ, et al. Prostate cancer early detection, Version 2.2016: Featured updates to the NCCN guidelines. JNCCN J Natl Compr Cancer Netw. 2016;14(5):509–19.
18. Bibbins-Domingo K, Grossman DC, Curry SJ, Davidson KW, Epling JW, García FAR, et al. Screening for colorectal cancer: US preventive services task force recommendation statement. JAMA - J Am Med Assoc. 2016;315(23):2564–75.
19. Ellinger J, Müller SC, Stadler TC, Jung A, von Ruecker A, Bastian PJ. The role of cell-free circulating DNA in the diagnosis and prognosis of prostate cancer. Urol Oncol Semin Orig Investig [Internet]. 2011;29(2):124–9. Available from: http://dx.doi.org/10.1016/j.urolonc.2009.05.010
20. Okegawa T, Nutahara K, Higashihara E. Association of circulating tumor cells with tumor-related methylated DNA in patients with hormone-refractory prostate cancer. Int J Urol. 2010;17(5):466–75.
21. Zhao F, Vesprini D, Liu RSC, Olkhov-Mitsel E, Klotz LH, Loblaw A, et al. Combining urinary DNA methylation and cell-free microRNA biomarkers for improved monitoring of prostate cancer patients on active surveillance. Urol Oncol Semin Orig Investig [Internet]. 2019;37(5):297.e9-297.e17. Available from: https://doi.org/10.1016/j.urolonc.2019.01.031
22. Zhao F, Olkhov-Mitsel E, Kamdar S, Jeyapala R, Garcia J, Hurst R, et al. A urine-based DNA methylation assay, ProCUrE, to identify clinically significant prostate cancer. Clin Epigenetics. 2018;10(1):1–12.
23. Bakavicius A, Daniunaite K, Zukauskaite K, Barisiene M, Jarmalaite S, Jankevicius F. Urinary DNA methylation biomarkers for prediction of prostate cancer upgrading and upstaging. Clin Epigenetics. 2019;11(1):1–10.
24. Di Francia R, Siesto RS, Valente D, Spartà D, Berretta M. Pharmacogenomics panel test for prevention toxicity in patient who receive Fluoropirimidine/ Oxaliplatin-Based therapy. Eur Rev Med Pharmacol Sci. 2012;16(9):1211–7.
25. Sobhani N, Sirico M, Generali D, Zanconati F, Scaggiante B. Circulating cell-free nucleic acids as prognostic and therapy predictive tools for metastatic castrate-resistant prostate cancer. World J Clin Oncol. 2020;11(7):450–63.
26. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the Serum of Cancer Patients and the Effect of Therapy. Cancer Res. 1977;37(3):646–50.
27. Delgado PO, Alves BCA, De Sousa Gehrke F, Kuniyoshi RK, Wroclavski ML, Del Giglio A, et al. Characterization of cell-free circulating DNA in plasma in patients with prostate cancer. Tumor Biol. 2013;34(2):983–6.
28. Altimari A, Grigioni ADE, Benedettini E, Gabusi E, Schiavina R, Martinelli A, et al. Diagnostic role of circulating free plasma DNA detection in patients with localized prostate cancer. Am J Clin Pathol. 2008;129(5):756–62.
29. Sita-Lumsden A, Fletcher CE, Dart DA, Brooke GN, Waxman J, Bevan CL. Circulating nucleic acids as biomarkers of prostate cancer. Biomark Med. 2013;7(6):867–77.
30. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, et al. DNA fragments in the blood plasma of cancer patients: Quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res. 2001;61(4):1659–65.
31. MANDEL P, METAIS P. Nuclear Acids In Human Blood Plasma. C R Seances Soc Biol Fil. 1948 Feb;142(3–4):241–3.
32. Bastian PJ, Palapattu GS, Yegnasubramanian S, Lin X, Rogers CG, Mangold LA, et al. Prognostic value of preoperative serum cell-free circulating DNA in men with prostate

cancer undergoing radical prostatectomy. Clin Cancer Res. 2007;13(18):5361–7.
33. Schwarzenbach H, Alix-Panabières C, Müller I, Letang N, Vendrell JP, Rebillard X, et al. Cell-free tumor DNA in blood plasma as a marker for circulating tumor cells in prostate cancer. Clin Cancer Res. 2009;15(3):1032–8.
34. Van Neste L, Herman JG, Otto G, Bigley JW, Epstein JI, Van Criekinge W. The Epigenetic promise for prostate cancer diagnosis. Prostate. 2012;72(11):1248–61.
35. Tew KD, Manevich Y, Grek C, Xiong Y, Uys J, Townsend DM. The role of glutathione S-transferase P in signaling pathways and S-glutathionylation in cancer. Free Radic Biol Med. 2011;51(2):299–313.
36. Henrique R, Jerónimo C. Molecular detection of prostate cancer: A role for GSTP1 hypermethylation. Eur Urol. 2004;46(5):660–9.
37. Rosenbaum E, Hoque MO, Cohen Y, Zahurak M, Eisenberger MA, Epstein JI, et al. Promoter hypermethylation as an independent prognostic factor for relapse in patients with prostate cancer following radical prostatectomy. Clin Cancer Res. 2005;11(23):8321–5.
38. Maldonado L, Brait M, Loyo M, Sullenberger L, Wang K, Peskoe SB, et al. GSTP1 promoter methylation is associated with recurrence in early stage prostate cancer. J Urol. 2014;192(5):1542–8.
39. Richiardi L, Fiano V, Vizzini L, De Marco L, Delsedime L, Akre O, et al. Promoter methylation in APC, RUNX3, and GSTP1 and mortality in prostate cancer patients. J Clin Oncol. 2009;27(19):3161–8.
40. Litovkin K, Van Eynde A, Joniau S, Lerut E, Laenen A, Gevaert T, et al. DNA methylation-guided prediction of clinical failure in high-risk prostate cancer. PLoS One. 2015;10(6):1–22.
41. Richiardi L, Fiano V, Grasso C, Zugna D, Delsedime L, Gillio-Tos A, et al. Methylation of APC and GSTP1 in Non-Neoplastic Tissue Adjacent to Prostate Tumour and Mortality from Prostate Cancer. PLoS One. 2013;8(7):1–7.
42. Vasiljević N, Ahmad AS, Thorat MA, Fisher G, Berney DM, Møller H, et al. DNA methylation gene-based models indicating independent poor outcome in prostate cancer. BMC Cancer. 2014;14(1):1–10.
43. Henrique R, Ribeiro FR, Fonseca D, Hoque MO, Carvalho AL, Costa VL, et al. High promoter methylation levels of APC predict poor prognosis in sextant biopsies from prostate cancer patients. Clin Cancer Res. 2007;13(20):6122–9.
44. Yegnasubramanian S, Kowalski J, Gonzalgo ML, Zahurak M, Piantadosi S, Walsh PC, et al. Hypermethylation of CpG Islands in Primary and Metastatic Human Prostate Cancer. Cancer Res. 2004;64(6):1975–86.
45. Alumkal JJ, Zhang Z, Humphreys EB, Bennett C, Mangold LA, Carducci MA, et al. Effect of DNA Methylation on Identification of Aggressive Prostate Cancer. Urology [Internet]. 2008;72(6):1234–9. Available from: http://dx.doi.org/10.1016/j.urology.2007.12.060
46. Moritz R, Ellinger J, Nuhn P, Haese A, Müller SC, Graefen M, et al. DNA hypermethylation as a predictor of PSA recurrence in patients with low- and intermediate-grade prostate cancer. Anticancer Res. 2013;33(12):5249–54.
47. Vasiljević N, Wu K, Brentnall AR, Kim DC, Thorat MA, Kudahetti SC, et al. Absolute quantitation of DNA methylation of 28 candidate genes in prostate cancer using pyrosequencing. Dis Markers. 2011;30(4):151–61.
48. Schnekenburger M, Karius T, Diederich M. Regulation of epigenetic traits of the

glutathione S-transferase P1 gene: From detoxification toward cancer prevention and diagnosis. Front Pharmacol. 2014;5 JUL(July):1–7.
49. Goessl C, Krause H, Miller M, Heicappell R, Schrader M, Sachsinger J, et al. Fluorescent methylation-specific polymerase chain reaction for DNA-based detection of prostate cancer in bodily fluids. Cancer Res. 2000;60(21):5941–5.
50. Cairns P, Esteller M, Herman JG, Sidransky D, Schoenberg M, Sidransky D, et al. Molecular detection of prostate cancer in urine by GSTP1 hypermethylation. Clin Cancer Res. 2001;7(9):2727–30.
51. Jatkoe TA, Karnes RJ, Freedland SJ, Wang Y, Le A, Baden J. A urine-based methylation signature for risk stratification within low-risk prostate cancer. Br J Cancer. 2015;112(5):802–8.
52. Wu T, Giovannucci E, Welge J, Mallick P, Tang WY, Ho SM. Measurement of GSTP1 promoter methylation in body fluids may complement PSA screening: A meta-analysis. Br J Cancer. 2011;105(1):65–73.
53. Baylin SB, Esteller M, Rountree MR, Bachman KE, Schuebel K, Herman JG. Abberant patterns of DNA methylation, chromatin formation and gene expression in cancer. Hum Mol Genet. 2001;10(7):687–92.
54. Majumdar S, Buckles E, Koochekpour JE and S. Aberrant DNA Methylation and Prostate Cancer. Curr Genomics. 2011;12(716):486–505.
55. Chan TA, Glockner S, Joo MY, Chen W, Van Neste L, Cope L, et al. Convergence of mutation and epigenetic alterations identifies common genes in cancer that predict for poor prognosis. PLoS Med. 2008;5(5):0823–37.
56. Gerda Egger, Gangning Liang AA& PAJ, Departments. Epigenetics in human disease and prospects for epigenetic therapy. Nature. 2004;429(May):246–7.
57. Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer. 2003;3(4):253–66.
58. Issa JP. Cancer prevention: Epigenetics steps up to the plate. Cancer Prev Res. 2008;1(4):219–22.
59. Constâncio V, Nunes SP, Henrique R, Jerónimo C. DNA Methylation-Based Testing in Liquid Biopsies as Detection and Prognostic Biomarkers for the Four Major Cancer Types. Cells. 2020;9(3).
60. Müller I, Urban K, Pantel K, Schwarzenbach H. Comparison of genetic alterations detected in circulating microsatellite DNA in blood plasma samples of patients with prostate cancer and benign prostatic hyperplasia. Ann N Y Acad Sci. 2006;1075:222–9.
61. Sunami E, Shinozaki M, Higano CS, Wollman R, Dorff TB, Tucker SJ, et al. Multimarker circulating DNA assay for assessing blood of prostate cancer patients. Clin Chem. 2009;55(3):559–67.
62. Papadopoulou E, Davilas E, Sotiriou V, Georgakopoulos E, Georgakopoulou S, Koliopanos A, et al. Cell-free DNA and RNA in plasma as a new molecular marker for prostate and breast cancer. Ann N Y Acad Sci. 2006;1075:235–43.
63. Goessl C, Müller M, Miller K. Methylation-specific PCR (MSP) for detection of tumour DNA in the blood plasma and serum of patients with prostate cancer. Prostate Cancer Prostatic Dis. 2000;3:7852.
64. Usadel H, Henrique RUI, Silva C, Oliveira J, Jero C, Lopes C, et al. Quantitative Gstp1 Hypermethylation in Bodily. Urology. 2002;
65. Jochen Reibenwein, Dietmar Pils, Peter Horak, Birgit Tomicek, Gregor Goldner, Nina Worel, Katarzyna Elandt and MK. PromoterHypermethylationofGSTP1, AR, and14-3-

3s in SerumofProstateCancer Patients andits Clinical Relevance. Prostate. 2007;67(April):427–32.
66. Bryzgunova OE, Morozkin ES, Yarmoschuk S V., Vlassov V V., Laktionov PP. Methylation-specific sequencing of GSTP1 gene promoter in circulating/extracellular DNA from blood and urine of healthy donors and prostate cancer patients. Ann N Y Acad Sci. 2008;1137:222–5.
67. Bastian PJ, Palapattu GS, Yegnasubramanian S, Rogers CG, Lin X, Mangold LA, et al. CpG Island hypermethylation profile in the serum of men with clinically localized and hormone refractory metastatic prostate cancer. J Urol. 2008;179(2):529–35.
68. Esteller M, Corn PG, Urena JM, Gabrielson E, Baylin SB, Herman JG. Inactivation of glutathione S-transferase P1 gene by promoter hypermethylation in human neoplasia. Cancer Res. 1998;58(20):4515–8.
69. Esteller M. Erratum: Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum dna from non-small cell lung cancer patients (Cancer Research (January 1, 1999) (67-70)). Cancer Res. 1999;59(15):3853.
70. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. Methylation-specific PCR: A novel PCR assay for methylation. Reactions [Internet]. 1996;93(September):9821–6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC38513/pdf/pnas01522-0532.pdf
71. Wong IHN, Lo YMD, Zhang J, Liew CT, Ng MHL, Wong N, et al. Detection of aberrant p16 methylation in the plasma and serum of liver cancer patients. Cancer Res. 1999;59(1):71–3.
72. Chen XQ, Stroun M, Magnenat JL, Nicod LP, Kurt AM, Lyautey J, et al. Microsatellite alterations in plasma DNA of small cell lung cancer patients. Nat Med. 1996;2(9):1033–5.
73. Goessl C, Müller M, Heicappell R, Krause H, Miller K. DNA-based detection of prostate cancer in blood, urine, and ejaculates. Ann N Y Acad Sci. 2001;945:51–8.
74. Lin X, Tascilar M, Lee WH, Vles WJ, Lee BH, Veeraswamy R, et al. GSTP1 CpG island hypermethylation is responsible for the absence of GSTP1 expression in human prostate cancer cells. Am J Pathol [Internet]. 2001;159(5):1815–26. Available from: http://dx.doi.org/10.1016/S0002-9440(10)63028-3
75. Yegnasubramanian S, De Marzo AM, Nelson WG. Prostate cancer epigenetics: From basic mechanisms to clinical implications. Cold Spring Harb Perspect Med. 2019;9(4):1–22.
76. Jerónimo C, Varzim G, Henrique R, Oliveira J, Bento MJ, Silva C, et al. I105V polymorphism and promoter methylation of the GSTP1 gene in prostate adenocarcinoma. Cancer Epidemiol Biomarkers Prev. 2002;11(5):445–50.
77. Costa VL, Henrique R, Jerónimo C. Epigenetic markers for molecular detection of prostate cancer. Dis Markers. 2007;23(1–2):31–41.
78. Singh S. Cytoprotective and regulatory functions of glutathione S-transferases in cancer cell proliferation and cell death. Cancer Chemother Pharmacol. 2015;75(1):1–15.
79. Sawers L, Ferguson MJ, Ihrig BR, Young HC, Chakravarty P, Wolf CR, et al. Glutathione S-transferase P1 (GSTP1) directly influences platinum drug chemosensitivity in ovarian tumour cell lines. Br J Cancer [Internet]. 2014;111(6):1150–8. Available from: http://dx.doi.org/10.1038/bjc.2014.386
80. Goering W, Kloth M, Schulz WA. DNA methylation changes in prostate cancer. Methods Mol Biol. 2012;863:47–66.

81. Schnekenburger M, Karius T, Diederich M. Regulation of epigenetic traits of the glutathione S-transferase P1 gene: From detoxification toward cancer prevention and diagnosis. Front Pharmacol. 2014;5 JUL.
82. Ng MCY, Shriner D, Chen BH, Li J, Chen WM, Guo X, et al. Meta-Analysis of Genome-Wide Association Studies in African Americans Provides Insights into the Genetic Architecture of Type 2 Diabetes. PLoS Genet. 2014;10(8).
83. Zelic R, Fiano V, Zugna D, Grasso C, Delsedime L, Daniele L, et al. Global hypomethylation (LINE-1) and gene-specific hypermethylation (GSTP1) on initial negative prostate biopsy as markers of prostate cancer on a rebiopsy. Clin Cancer Res. 2016;22(4):984–92.
84. Shepard TF, Platz EA, Kantoff PW, Nelson WG, Isaacs WB, Freije D, et al. No association between the I105V polymorphism of the glutathione S-transferase P1 gene (GSTP1) and prostate cancer risk: A prospective study. Cancer Epidemiol Biomarkers Prev. 2000;9(11):1267–8.
85. Bremnes RM, Sirera R, Camps C. Circulating tumour-derived DNA and RNA markers in blood: A tool for early detection, diagnostics, and follow-up? Lung Cancer. 2005;49(1):1–12.

Chapter 4: TMPRSS2-ERG Gene Fusion in Prostate Cancer of Moroccan Patients

Chapter 4: TMPRSS2-ERG Gene Fusion in Prostate Cancer of Moroccan Patients

I. Introduction

Prostate cancer (PCa) is one of the most common adult malignancies (1). Cancer presents many genetic alterations, including gene mutations, DNA copy number changes, and gene rearrangements. For prostate cancer, the most common alteration found in patients is the fusion gene between androgen-regulated *TMPRSS2* and the oncogene *ERG* (1).

The androgen receptor (AR) is involved in the development and function of the normal prostate and for the growth of PCa. For a long time, it was unknown which AR regulated genes were involved in tumor growth.

The fusion of *TMPRSS2* to *ETS* family genes in prostate cancer, was first described by Tomlins et al. (2005) (2). The gene fusion *TMPRSS2-ETS* causes the *ERG/ETV* genes to be controlled by androgen regulation. Gene fusion could, therefore, predict the response to hormonal treatment, and some studies have been conducted to address this possible correlation, such as the in vitro study by Hermans et al. (2006) (3). Currently, androgen-regulated *TMPRSS2* controls the expression of the *ERG* or *ETV1* oncogene and has a large impact on androgen-dependent growth and survival of PCa (2).

The aim of this study is to identify *TMPRSS2-ERG* fusion transcript variants in the prostate cancer of Moroccan patients. The proposed work will allow the detection of this molecular subtype of prostate cancer, which may have prognostic and/or therapeutic relevance. And could be embraced as a new molecular test in the Moroccan clinical routine.

II. Background

II.1 TMPRSS2 Gene

Transmembrane Serine Protease 2 (*TMPRSS2*) is a protein coding gene that belongs to the serine protease family (4). This gene has gained a central importance in the molecular biology of cancer and has raised great potential as a valuable biomarker for this disease (4). The *TMPRSS2* gene promoter contains AR response elements and it seems to be expressed in normal prostate and was demonstrated to be upregulated by androgenic hormones in prostate cancer cells and downregulated in androgen-independent prostate cancer tissue (5). Its expression is

higher in PCa and hyperplasia than in normal prostate gland (6). The *TMPRSS2* gene is involved in prostate carcinogenesis and works as an activator of protease-activated receptor 2, playing a partial role in a signal transduction pathway related to inflammation, metastasis and invasion (Figure 4.1) (7).

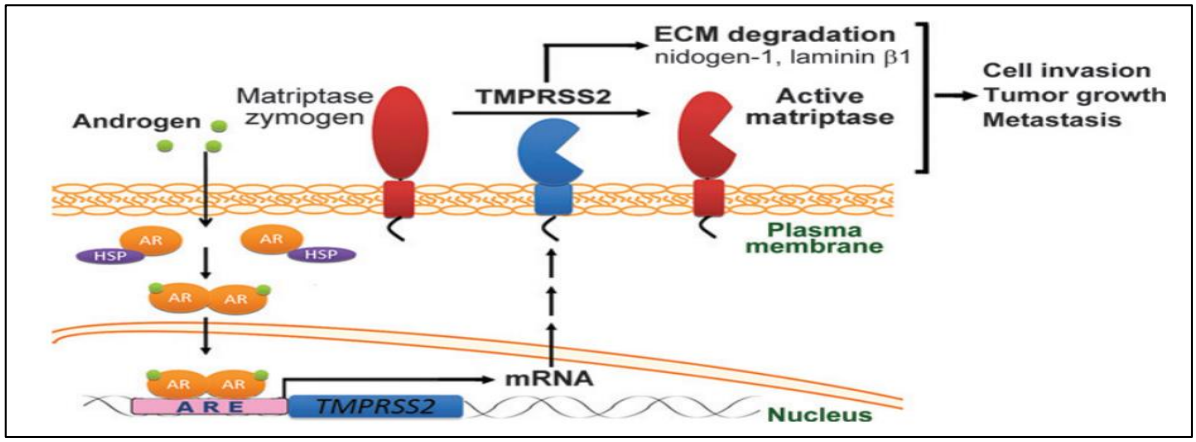


Figure 4. 1: Schematic model for TMPRSS2-promoted prostate cancer cell invasion, tumor growth, and metastasis. In prostate cancer cells, androgens can induce TMPRSS2 expression via activating AR to bind the androgen response element (ARE). Androgen-induced TMPRSS2 initiates a pericellular proteolytic cascade to activate matriptase that promotes prostate cancer cell invasion. Moreover, TMPRSS2 can also promote prostate cancer tumorigenicity and metastasis, which occurs along with degradation of ECM components, nidogen-1, and laminin b1. Thus, TMPRSS2 can mediate matriptase activation by androgens and disrupt the ECM network, leading to prostate cancer cell invasion, tumor growth, and metastasis (8).

TMPRSS2 protein is selectively expressed in normal prostate tissues and is overexpressed in the neoplastic prostatic epithelium (9–12). The expression of the *TMPRSS2* gene seems to be regulated by androgen-responsive elements (AREs) in a promoter (12,13), and it has been shown that androgen stimulation can induce the overexpression of *ERG* in a *TMPRSS2–ERG*–positive cell line (2). The dysregulation of *ETS* transcription factors may affect PCa progression though AREs, which as a result, could affect cell differentiation, proliferation and apoptosis, among other processes (2,14).

II.2. ERG Gene

In prostate cancer, the translocation partner of *TMPRSS2* gene is the *ERG* gene. *ERG* is a member of *ETS* transcription factor family (15) and consists of 17 exons (16). The overexpression of this gene, which is due to fusion with *TMPRSS2*, is one of the early causes of PCa. *ERG* activates transcriptional program and induces oncogenesis by upregulation *MYC*, *EZH2* and *SOX 9* and repressing *NKX3* (17,18). The *ERG* gene is believed to play a role as a

key regulator of many physiological mechanisms, such as, cell proliferation, differentiation, angiogenesis, apoptosis and vascular inflammation (7).

Additional *ETS* gene fusions have been identified including *ETV1*, *ETV4* (19), *ETV5* (20) and *FLI1* (21). However, the frequency of these variants is much lower than that of ERG rearrangements but still accounts for 2-5% of all *ETS* gene fusions found in PCa (22,23). Nevertheless, it is not yet understood how *ETS* gene rearrangements are indeed capable of initiating prostate tumorigenesis whereas its role in mediating progression toward more aggressive disease is well accepted (24,25).

Understanding the mechanisms of *ETS* gene expression and function will help to identify new therapeutic targets for early and late-stage prostate cancer. *ERG* overexpression is a key tool for the diagnosis and treatment of prostate cancer (7).

II.3. TMPRSS2-ERG Gene fusion

TMPRSS2 and *ERG* are located on chromosome 21 q 22 (26). *TMRSS2* usually fuses to ERG by replacing the 5' end of *ERG* with the 5' untranslated region of *TMPRSS2* (21q22.2) (Figure 4.2). Different transcripts of *TMPRSS2-ERG* are detected by quick amplification of cDNA ends and RT PCR assays. The most common types of *TMPRSS2-ERG* transcripts are: *TMPRSS2* (exon 1 or exon 2) combined with *ERG* exons 2,3,4, and 5. The less repetitive one is *TMPRSS2* (exons 4, 5) combined with *ERG* exon 4 or 5. The most common variant is T1-E4, which is approximately 75 to 80% of *ERG* variants in PCa patients. This *TMPRSS2-ERG* fusion shows that this region is an important spot for chromosomal rearrangement and deletion in PCa (27).

These variants are useful in postulating the prognostic outcomes of each patient by detecting rearrangement and *ERG* overexpression (6). Mostly, *TMPRSS2-ERG* fusions are frequently accompanied by loss of the entire chromosome 21 sequence between *TMPRSS2* and *ERG* genes (interstitial deletion) (28).

There are 18 distinct fusion transcripts of TMPRSS2-ERG that are distinguished (Figure 4.3).

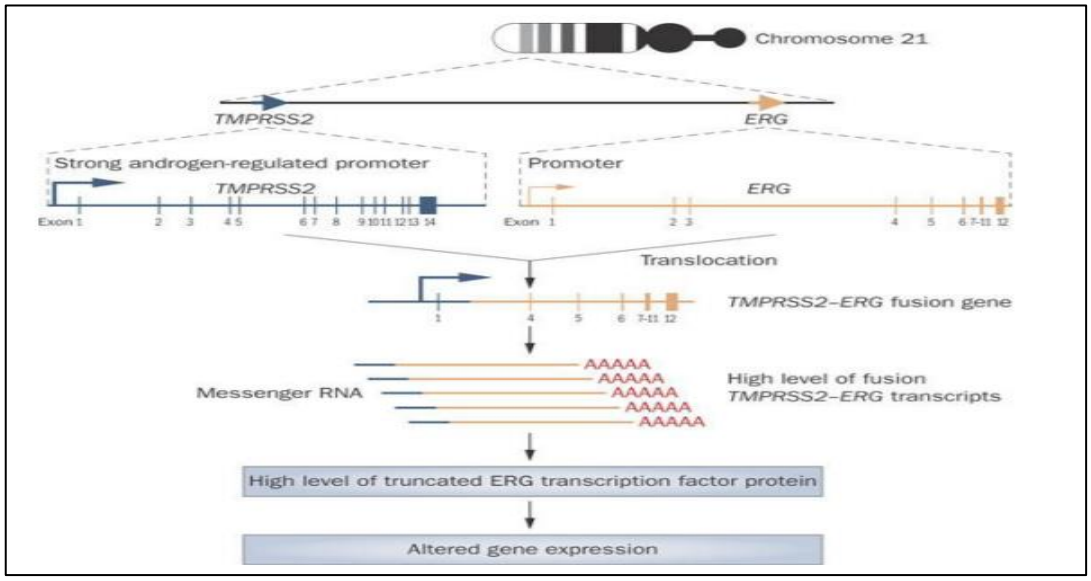


Figure 4. 2: The Arrangement of TMPRSS2-ERG Fusion Gene by Translocation (29,30).

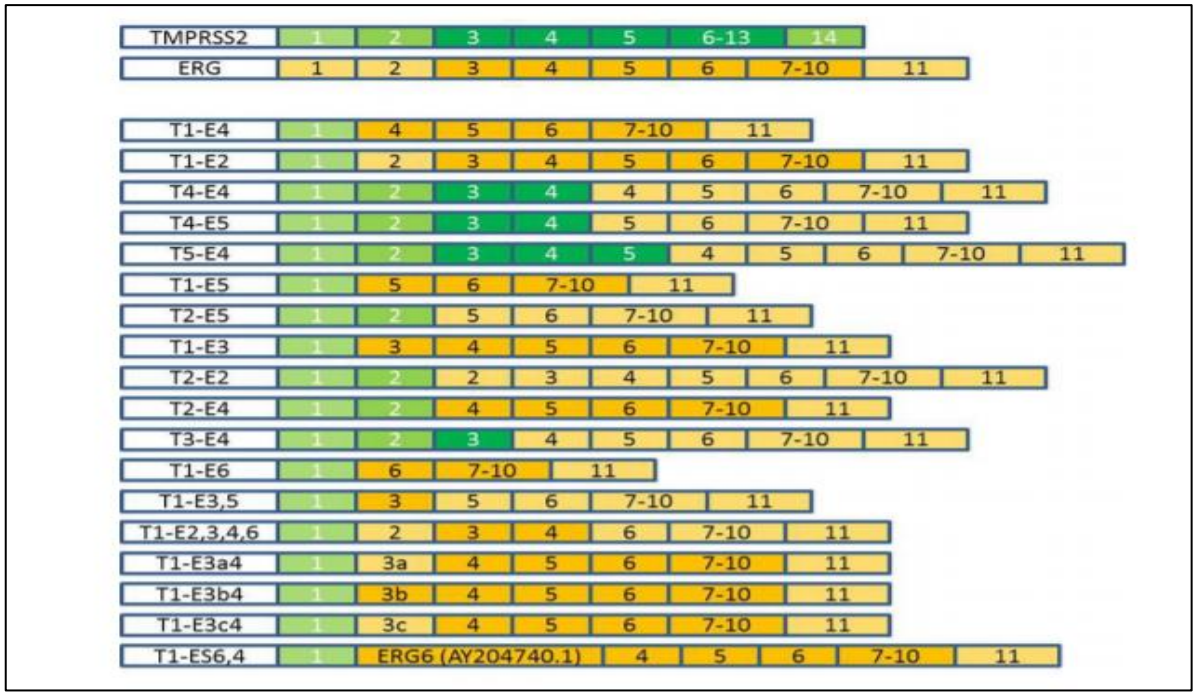


Figure 4. 3: Different mRNA Splice Variants of the TMPRSS2-ERG Fusion Gene (31).

In prostate cancer, multiple *TMPRSS2-ERG* fusion variants have been identified at the transcript level, most of which involve only the 5' untranslated sequence of *TMPRSS2* (2,32–35). Recurrent gene fusion between the androgen-regulated gene *TMPRSS2* and members of the *ETS* transcription factor family, most commonly *ERG*, are present in about 50% of prostate

cancer cases (36).

The *ERG* oncogene is one of the earliest genetic variations identified as a transcription factor of the *ETS* family, which is located on chromosomal band 21q22 and is overexpressed in over 50% of PCa (37). *TMPRSS2* is a prostate-specific and androgen regulated gene, located very close to *ERG* on chromosome 21. *ERG* overexpression in major PCa cases is driven by fusion with *TMPRSS2* (25). In recent studies, *TMPRSS2-ERG* gene fusion is the pervasive variant in approximately 40% to 70% of PCa cases (38). The gene fusion could be a promising biomarker of PCa (36).

Furthermore, there are other fusions between the *TMPRSS2* gene and *ETS* family genes, such as *TMPRSS2-ETV1*, *TMPRSS2-ETV4*, and *TMPRSS2-ETV5* and between other genes and *ETS* family genes, such as *SLC45A3-ETV5* (6).

A large variation in the prevalence of TMPRSS2-ERG gene fusion in prostate cancer has been observed among different ethnicities (Figure 4.4) (39).

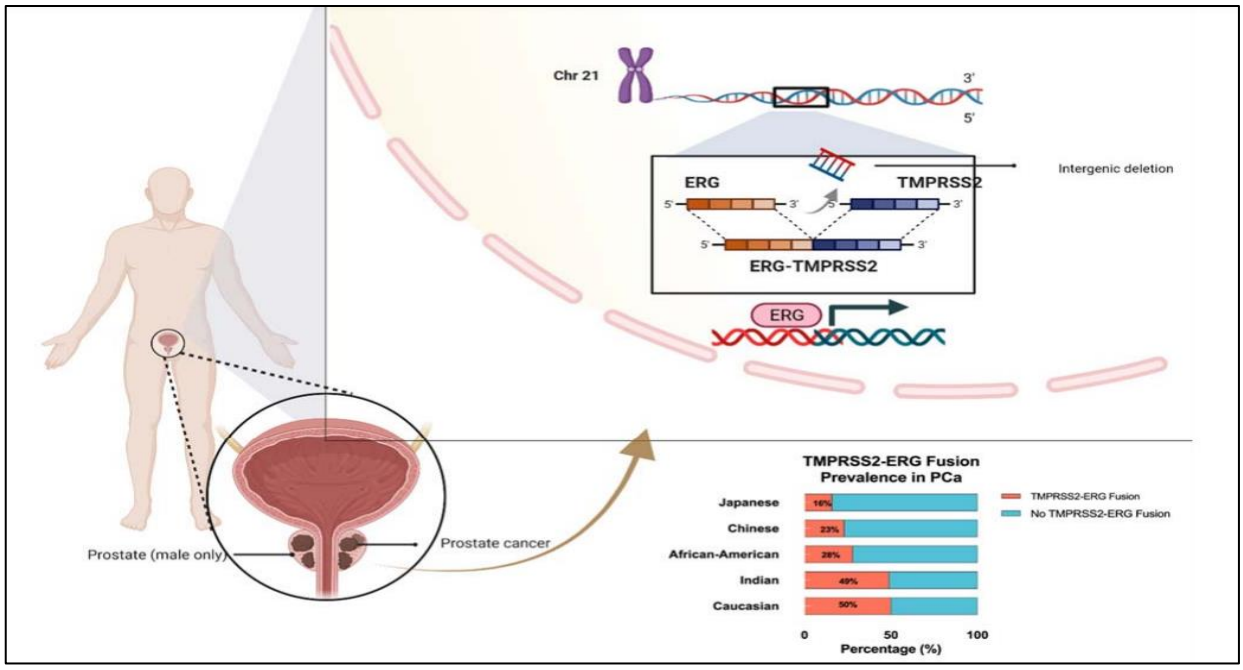


Figure 4. 4: Schematic diagram of TMPRSS2-ERG gene fusion on chromosome 21 and the ethnic prevalence of TMPRSS2-ERG gene fusion in prostate cancer patients (39).

II.4. Role of *TMPRSS2-ERG* fusion gene in PCa

The role of the *TMPRSS2-ERG* fusion gene in PCa has been widely studied. Delliaux & colleagues (2020) found that the expression of the fusion gene in bone-metastasized prostate

cancer increased the properties of bone cell-like phenotype acquisition, or osteomimicry, and increased the formation of osteoblastic lesions, while decreasing the bone resorption (40). The presence of the fusion gene could increase the density of nerves in prostate tumors, and mRNAs of the NRP/PLXN/SEMA family network could be targeted by the *ERG* transcription factor, which might affect innervation by increasing nerve density (41). Moreover, when there is *TMPRSS2-ERG* fusion gene present in prostate tumors, distinct intratumoral androgen profiles will be shown in patients, such as an altered intratumoral dihydrotestosterone (DHT) to testosterone ratio and an altered androgen response (42). The *TMPRSS2-ERG* fusion gene is also associated with non-communicable disease, such as obesity and diabetes. Interestingly, Graff & her coworkers found that the risk of *TMPRSS2-ERG* fusion gene with *ERG*-positive prostate cancer was associated with taller height. In contrast, developing the fusion gene with *ERG*-positive PCa was negatively associated with obesity (43). As a result, they found that insulin and insulin-like growth factor 1 (IGF-1) could decrease the formation of the fusion gene, but the number of fusion genes could be increased by hyperglycemia and insulin-like growth factor binding protein-2 (IGFBP2) (44).

TMRPSS2-ERG fusion has been reported as the most common genomic alteration that drives prostate tumorigenesis, and this androgen-upregulating fusion is known to correlate with higher grades of the disease (45). Although men of African ancestry are known to present with higher disease grade, only three studies have examined the *TMRPSS2-ERG* fusion on the continent (45–47). This fusion often results from either a chromosomal translocation or an interstitial deletion, and these studies reported rates that were less than 20% in Ghanaian and Black South African patients (45–47). Prostate cancer continues to be the leading cause of cancer morbidity and mortality among African men (48,49).

Currently, urinary *TMPRSS2-ERG* marker is used as a novel screening and diagnostic tool to detect prostate tumors in additionother markers, such as serum PSA and urinary *PCA3*, that can also serve as prognostic markers (50,51).

II.5 Molecular Pathway of *TMPRSS2-ERG* gene fusion

Due to the activity of androgen receptor (AR), the regulatory region of the androgen-regulated gene *TMPRSS2* is bound to the coding sequence of *ERG*, which recruits the upregulation of *ERG* transcription (or other *ETS* gene family members, i.e., *ETV1*, *ETV4*, and *ETV5*). Under

the control of *AR*, the *TMPRSS2-ERG* fusion transcript is overproduced and translated into wild-type *ERG* or N-terminally truncated *ERG* that retains all important functional domains. ERG binds to DNA at specific sequences and is targeted by several post-translational modifications (acetylation, methylation, ubiquitination, and phosphorylation) controlled by diverse proteins (i.e. P300, EZH2, Cullin3-SPOP, AKT, ERK2..). Together with its interactors (i.e., *AP-1, FOXO1*,..), *ERG* regulates target gene expression and processes that positively affect prostate tumorigenesis and prostate cancer maintenance. Known direct target genes of *ERG* in *TMPRSS2-ERG* fusion positive tissues include *MMP3*, *PLAU*, *LAMC2*, *KCNS3*, *PLA1A*, *C-MYC*, *GNMT*, *SARDH*, *CXCR4*, *ADAMTS1*,*TFF3*, *ERG*, *PLAT*, *MMP9*, *NDRG1*, *CUTL2*, *AR*, *KLK3 (PSA)*,*KLK2*, *SLC43A1*, *FKBP5*, *EZH2*, *ZBTB16*, *HPGD*, *ZEB1*, *SPINT1*, *IL1R2*, *PSMA*, and *OPN*. These and other up or down regulated genes in fusion positive cancers could facilitate PCa progression (52) (53). Furthermore, perturbation of the PI3K pathway (PTEN loss and AKT overexpression) promotes the progression of invasive carcinoma in *TMPRSS2-ERG* positive prostate tissues (Figure 4.5) (54).

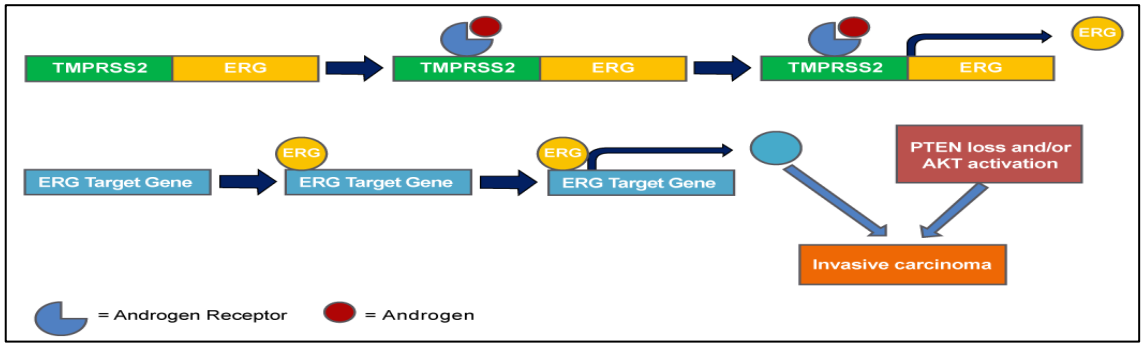


Figure 4. 5 : Biology of TMPRSS2-ERG gene fusions in prostate cancer (54)

III. Materials and Methods

Fresh-frozen radical prostatectomy specimens and blood were collected from the tumor regions in the prostate. The RNA was extracted using different methods and commercial RNA extraction kits to obtain the best yield needed for the following steps. Then, RT-polymerase chain reaction (PCR) was performed before proceeding to the molecular cloning and sequencing of TMPRSS2-ERG variants steps (Figure 4.6).

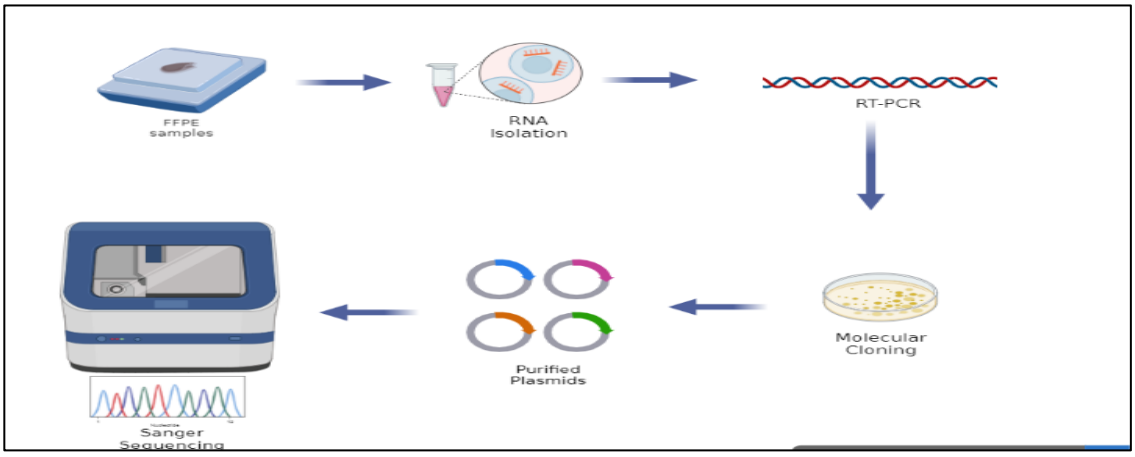


Figure 4. 6: Flow chart of the method section

A. Cohort constitution

We recruited patients from consenting persons with prostate cancer from Northeast Morocco (mainly Oujda University hospitals that cover patients from one of large areas in Morocco). Forty-five formalin-fixed paraffin-embedded tissue samples were collected.

B. Clinical Characteristics

The following table presents the principal clinical and pathological characteristics of our 19 prostate cancer patients:

Table 4. 1: Clinical Characteristics of Prostate Cancer Patients

Sample	Type of PCa	Gleason Score	PSA ng/ml
P1	Adenocarcinoma	7(3+4)	23
P2	Adenocarcinoma	7(3+4)	424
P3	Adenocarcinoma	7(3+4)	500
P4	Adenocarcinoma	7(3+4)	123
P5	Adenocarcinoma	8 (4+4)	880
P6	Adenocarcinoma	7(3+4)	37,7
P7	Adenocarcinoma	8 (4+4)	412
P8	Adenocarcinoma	8 (4+4)	649
P9	Adenocarcinoma	7(3+4)	55
P10	Adenocarcinoma	7(3+4)	140
P11	Adenocarcinoma	7(3+4)	65
P12	Adenocarcinoma	8 (4+4)	62
P13	Adenocarcinoma	8 (4+4)	70
P14	Adenocarcinoma	6(3+3)	30
P15	Adenocarcinoma	8 (4+4)	380

P16	Adenocarcinoma	7(3+4)	40
P17	Adenocarcinoma	6(3+3)	27
P18	Adenocarcinoma	7(3+4)	74
P19	Adenocarcinoma	8 (4+4)	400

C. Reagents

As part of the genomic characterization, we extracted the RNA for the RT-PCR and sequencing from formalin-fixed paraffin-embedded tissue collected from Moroccan patients. Formalin-fixed paraffin-embedded (FFPE) prostate tissue blocks were sliced 2 slides per sample by the Pathology Core and GW Biorepository at The George Washington University, Washington DC, USA.

Primers for RT-polymerase chain reaction (PCR) were designed using PrimerBlast (National Center for Biotechnology Information (NCBI), Bethesda, MD). Primer pairs were selected for a length of 15-21 base pairs, PCR product length of 100-150 bp, melting temperature 58-62oC, G/C content 20-70%, and zero 3' self-complementarity. Human specificity for each sequence was verified using BLAST (NCBI). Primers were produced by GenScript (Piscataway, NJ).

Table 4. 2: RT-PCR primers for the detection of TMPRSS2-ERG fusions

Gene	**Primer 5'-3'**
TMPRSS2-1A	TAGGCGCGAGCTAAGCAGGAG
TMPRSS2-1A1	CGCGAGCTAAGCAGGAGGC
ERG-4B	GTAGGCACACTCAAACAACGACTGG
ERG-4B1	TGGTCCTCACTCACAACTGATAAG
ERG-5B	TCATCCCAACGGTGTCTGGGCTG
ERG-5B1	CAACGGTGTCTGGGCTGCCCACC
TMPRSS2 Exon8-A	AAGCAGTGGTTTCTTTACGCT
TMPRSS2 Exon9-B	AGAGCAGGATTGTGGGCG

Gene names, and forward and reverse primer sequences used in RT-PCR reactions. All 'A' primers are forward primers, 'B' primers are reverse primers; 'A1' and 'B1' primers are nested primers, and TMPRSS2(exon 8-exon 9) is the wild type (WT)

D. RNA Extraction

Different RNA extraction methods have been used before adopting TRIzol protocol because extracting RNA from an FFPE biopsy was challenging since the RNA is degraded easily.

- Deparaffinization of the FFPE slides was performed using an optimized protocol from the RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE (ThermoFisher Scientific) (Appendix 1). RNA was isolated via QIAgen miRNeasy kit (Valencia, CA) per the manufacturer's protocol or RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE (ThermoFisher Scientific) and stored at −80°C. RNA integrity was measured via 1% agarose gel electrophoresis, and the RNA concentration was determined using a Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific). → the curve and the concentration of the RNA was not good.

- Deparaffinization of the FFPE slides was performed using a standard xylene and ethanol procedure prior to macrodissection of prostate tumor and adjacent histologically normal tissue (ADJNT). Total RNA was isolated and extracted by TRIzol (Invitrogen, Catalog # 15596018) following the manufacturer's instructions and using a powerful homogenizer to homogenize the tissue. The quantity and quality of FFPE RNA was measured using the Agilent 2100 Bioanalyzer system (RNA 6000 Pico Kit). → good yield of RNA and is the one that has been used in the following steps

- *TRIzol protocol:* Deparaffinized prostate cancer tissue pellet was resuspended and sheared in 1 ml TRIzol using a 25 G needle. After incubation in QIAzol for 5 min at room temperature, 200 µl of chloroform was added and the suspension was vortexed vigorously for 15 sec. The suspension was incubated for 3min at room temperature and then centrifuged at 10,000xg for 15 min at 4°C. The clear upper aqueous phase was carefully transferred to a fresh 1.7 ml Eppendorf tube. Then, 500 µl of isopropanol was added and the sample was stored in -80°C overnight. After thawing, the sample was centrifuged at 10,000xg for 15 min at 4°C. The isopropanol was carefully removed and 1 ml 70% ethanol was added to wash the pellet. The sample was centrifuged at 10,000xg for 5 min at 4°C. The ethanol was gently dumped out and the Eppendorf was left open on the bench top for 5min to allow any remaining ethanol to evaporate. The RNA pellet was resuspended in 30-40 µl of RNase-free water.

- *QIAgen miRNeasy protocol*: Deparaffinized prostate cancer tissue pellet was resuspended and sheared in 700 µl QIAzol using a 25 G needle. After incubation in QIAzol for 5 min at room temperature, 200 µl of chloroform was added and the suspension was vortexed vigorously for 15 sec. The suspension was incubated for 3 min at room temperature and then centrifuged at 12,000xg for 15 min at 4°C. The clear upper aqueous phase was transferred to a fresh 1.7 ml Eppendorf tube and mixed with 525 µl 100% ethanol. The supernatant was added to the RNeasy column, centrifuged at 8,000xg for 30 sec, and washed once with 500 µl RWT buffer. DNase I (10 µl in 70 µl RDD Buffer) was added to the column and incubated for 15min at room temperature, followed by washes with 500 µl RWT and 700 µl RPE buffers, with a 30 sec 8,000xg spin between each wash. The spin column was placed in a new 2 ml elution tube and centrifuged at max speed for 1 min. RNA was eluted in 40 µl of RNase-free water in a fresh 1.7 ml Eppendorf tube.
- *RecoverAll™ Total Nucleic Acid Isolation protocol*: Deparaffinized prostate cancer tissue pellet was resuspended in 4 µl of protease and 200 µl of digestion buffer. After incubation in heat blocks for 15 min at 50°C, then 15 min at 80°C, 790 µl of isolation additive/ethanol mixture was added and the suspension was mixed by pipetting up and down. The suspension was Passed through a filter cartridge and then centrifuged at 10,000xg for 30sec. The flow-through was discarded and the filter cartridge was re-inserted into the same collection tube. Then it was washed twice with 700 µl then 500 µl of the wash solutions. 60 µl of the DNase mix was added to the center of the Filter Cartridge. After incubation for 30 min at room temperature (22–25°C), 700 µl of Wash1 solution, then 500 µl of Wash2/3 solution were added respectively with a 30 sec 10,000xg spin between each wash. The filter cartridge column was placed in a fresh collection tube, and RNA was eluted in 60 µl of RNase-free water.

E. Reverse transcription-polymerase chain reaction

Reverse transcription-polymerase chain reaction consists of converting RNA into its complementary DNA sequence by reverse transcriptase, to synthesize a second strand with DNA polymerase. The cDNA was then used as a template for exponential amplification using PCR.

Reverse transcription (RT) of 1 µg purified RNA to cDNA was performed using the Taq Man Reverse Transcription kit (Applied Biosystems, Foster City, CA), random hexamer primers according to manufacturers' instructions, and the following reaction mix and incubation protocol:

Table 4. 3: Reverse transcription reaction mix and incubation protocol

10X Reaction buffer	5 µl
MgCl2 solution (25 mM)	11 µl
dNTP mixture (10 nM)	10 µl
Random hexamers (50 µM)	2.5 µl
RNase inhibitor (20U/µl)	1 µl
Reverse Transcriptase II (50U/µl)	1,25 µl
1µg RNA	
Nuclease/RNase free water	Up to 50 µl

25°C	10 min
42°C	30 min
95°C	5 min
4°C	∞

PCR was performed using the 2X GoTaq Green Master Mix (Promega, Madison, WI) using 1 µl of cDNA with the following reaction mix and cycling conditions:

Table 4. 4: PCR reaction mix and incubation protocol

2X Go Green Taq Master Mix	12.5 µl
Primer_Forward (10 µM)	1 µl
Primer_Reverse (10 µM)	1 µl
cDNA template	1 µl
Nuclease/RNase free water	9.5 µl

95°C	3min	
95°C	30 sec	X35 cycle
55°C	30 sec	
72°C	30 sec	
72°C	6 min	
4°C	∞	

PCR products were analyzed via 1% agarose gel run at 120 V for 20-30 min. Housekeeping gene Eukaryotic Translation Initiation Factor 1A, X-Linked 1A (EIF1AX) was used as a loading control.

F. Molecular cloning and sequencing of TMPRSS2-ERG variants

- ***Molecular Cloning Principle***

TOPO cloning is a highly efficient technique, for the direct insertion of Taq polymerase-amplified PCR products into a plasmid vector. Without requirement for DNA ligase, post-PCR procedures, or PCR primers containing specific sequences. The TOPO cloning technique relies on the ability of adenine (A) and thymine (T) on different DNA fragments to either hybridize or ligate together in the presence of ligase or topoisomerase.

TOPO vectors are designed with a specific sequence 5´-(C/T) CCTT-3' at the two linear ends with the topoisomerase enzyme covalently attached to both of its strands' free 3' ends. When this vector is mixed with PCR products, the free 5' ends of the PCR product strands will attach to the topoisomerase 3' end of each vector strand. Then, the strands will be linked by the bound topoisomerase.

- ***Molecular Cloning Method***

RNA isolation, cDNA generation, and PCR amplification were performed as described previously. *TMPRSS2(exon 1)-ERG(exon 4)* and *TMPRSS2(exon 1)-ERG(exon 5)* and *TMPRSS2 (exon8-9)* were amplified using 5’ and 3’ rapid amplification of cDNA ends in the following sections: *TMPRSS2 (exon 1)-ERG (exon 4)* (primers exon1_F exon 4_R); *TMPRSS2 (exon 1)-ERG (exon 5)* (primers exon1_F exon 5_R); *TMPRSS2 (exon8-9)* (primers exon8_F exon 9_R). *EIF1AX* was used as housekeeping gene (control).

After PCR products were run on 1% agarose cells, bands were cut out using a razor blade. DNA was purified from the agarose using the QIAgen Gel Extraction kit. Briefly, the gel slice was weighed and three volumes of Buffer QG were added, and the gel slice was incubated at 50 °C for 10 min or until the gel was fully dissolved. One volume of isopropanol was added and the sample was mixed well before being transferred to a spin column. The column was centrifuged at 13,000 rpm for 1 min. Two wash steps, 1X 500 µl Buffer QG and 1X 750 µl Buffer PE, were performed with a 1 min centrifugation step between each wash. After the flow through from the last wash was discarded, the column was centrifuged for an additional minute. The column was then transferred to a fresh 1.7ml Eppendorf and 50 µl of water was added to elute the DNA. Next, DNA was cloned into the pCR2.1 TOPO vector (Life Technologies, Grand Island, NY) by incubating 1 µl DNA with 1 µl salt solution, 3 µl water, and 1 µl of vector on ice for 5min. Two microliters of the cloning reaction were added to a vial of One-Shot TOP10 Chemically Competent *E. coli* cells (Thermo Fisher Scientific) and incubated on ice for 5-30 min. Cells were then heat shocked in a 42°C water bath for 30 sec and immediately placed on ice for 2 min. 250 µl of room temperature super optimal broth with catabolite repression (SOC) media was added and the cells were placed horizontally to shake at 220 rpm at 37°C for 1 hr. Then, 5-10 µl were spread on a pre-warmed Luria broth (LB) agar plate (with 50 µg/ml carbenicillin) autoclaved using glass beads. The plate was incubated at 37°C overnight. The following day, about 10 individually picked colonies were spread on a new LB place and PCR was carried out to confirm that the sequence of interest was present. PCRs were performed using nested primers using the cycling conditions listed previously (exception: initial denaturing step was run for 5 min). Two to three positive colonies were inoculated into 2 ml of LB+carbenicillin (50 µg/ml) and placed in a 37°C shaker at 220 rpm overnight.

The following day, 1 ml of bacteria was added to a cryovial with 160 µl glycerol. This glycerol stock was stored in -80 °C. The remaining bacteria was transferred to a 1.7ml Eppendorf tube to use in Qiagen's Miniprep kit to isolate the inserted vector. The following protocol steps were conducted at room temperature and all centrifuge steps are carried out at 13,000 rpm unless otherwise noted. Bacteria were centrifuged at 8,000 rpm for 3 minutes at room temperature. Bacterial pellets were resuspended in 250 µl Buffer P1 and 250 µl Buffer P2 was added. Tubes were inverted four to six times and 250 µl Buffer N3 was added. Tubes were inverted four to six times before centrifuging for 10 min. The supernatant was transferred to a spin column and centrifuged for 1min. The spin column was washed 1X with 500 µl Buffer PB and 1X 750 µl

Buffer PE with one 30 sec centrifuge steps between each wash and following the PE wash. The column was transferred to a fresh Eppendorf and 40 µl water was added to elute the DNA. The column was allowed to stand for 1min before centrifuging for 1min. The DNA concentration was determined by a Nanodrop ND-1000 spectrophotometer**.** The insert of interest was confirmed by running PCR using nested primers.

- ***Sequencing of TMPRSS2-ERG variants***

Purified plasmids were submitted to Sequetech (Mountain View, CA) for Sanger sequencing (3 µg per run, forward and reverse reads using vector M13 primer pair). Sequences were analyzed for clean chromatograms via Chromas Lite (Technelysium Pty Ltd, South Brisbane, Australia), sequence similarity to *TMPRSS2-ERG and TMPRSS2* via BLAST (NCBI), and 100% alignment to the reference sequence. Two to three colonies with forward and reverse reads were generated for each sequence of interest and used to generate a final consensus sequence.

IV. Results and Discussion

We have obtained RNA from 28 out of 45 FFPE samples, 17 cases showed poor RNA quality. We analyzed a series of 28 prostate cancer samples for the expression of *TMPRSS2-ERG* fusion transcripts. Figure 4.7 represents 4 agarose gel profiles indicating RT-PCR screening of the different TMPRSS2-ERG fusion transcripts. The gel profile (A) represents an example profile of a patient positive for T1/E4 fusion, where lane 1 refers to ladder, lane 2 refers to T1/E4 transcript, lane 3 refers to T1/E5, lane 4 refers to T8/T9 and Lane 5 refers to the housekeeping gene EIF1AX. The gel profile (B) represents an example profile of patient positive for T1/E5 fusion, where lane 1 refers to ladder, lane 2 refers to T1/E5 transcript, lane 3 refers to T1/E4, lane 4 refers to T8/T9 and lane 5 refers to the housekeeping gene EIF1AX. The gel profile (C) represents an example profile of patient positive for both T1/E4 and T1/E5 fusions, where lane 1 refers to ladder, lane 2 refers to T1/E4 transcript, lane 3 refers to T1/E5, lane 4 refers to T8/T9 and lane 5 refers to the housekeeping gene EIF1AX. The gel profile (D) represents an example profile of patients positive for both T1/E4 and T1/E5 fusions and presents the wild type of TMPRSS2 gene, where lane 1 refers to ladder, lane 2 refers to T1/E4 transcript, lane 3 refers to T1/E5, lane 4 refers to T8/T9 and lane 5 refers to the housekeeping gene *EIF1AX*. Therefore, we were able to classify cases into categories (fusion gene present/absent). RT-PCR analysis revealed *TMPRSS2-ERG* fusion transcripts in 19 of the 28 tumors (60%). The vast diversity of

fusion transcripts is a unique feature of the *TMPRSS2-ERG* fusion. Two types of *TMPRSS2-ERG* chimeric transcripts were detected, consisting of *TMPRSS2* exon 1 fused to either *ERG* exon 4 (9 cases) or *ERG* exon 5 (5 cases) or both *ERG* exon4 and *ERG* exon 5 (3 cases) or *TMPRSS2* (exon 8-exon 9) which is the wild type (2 cases).

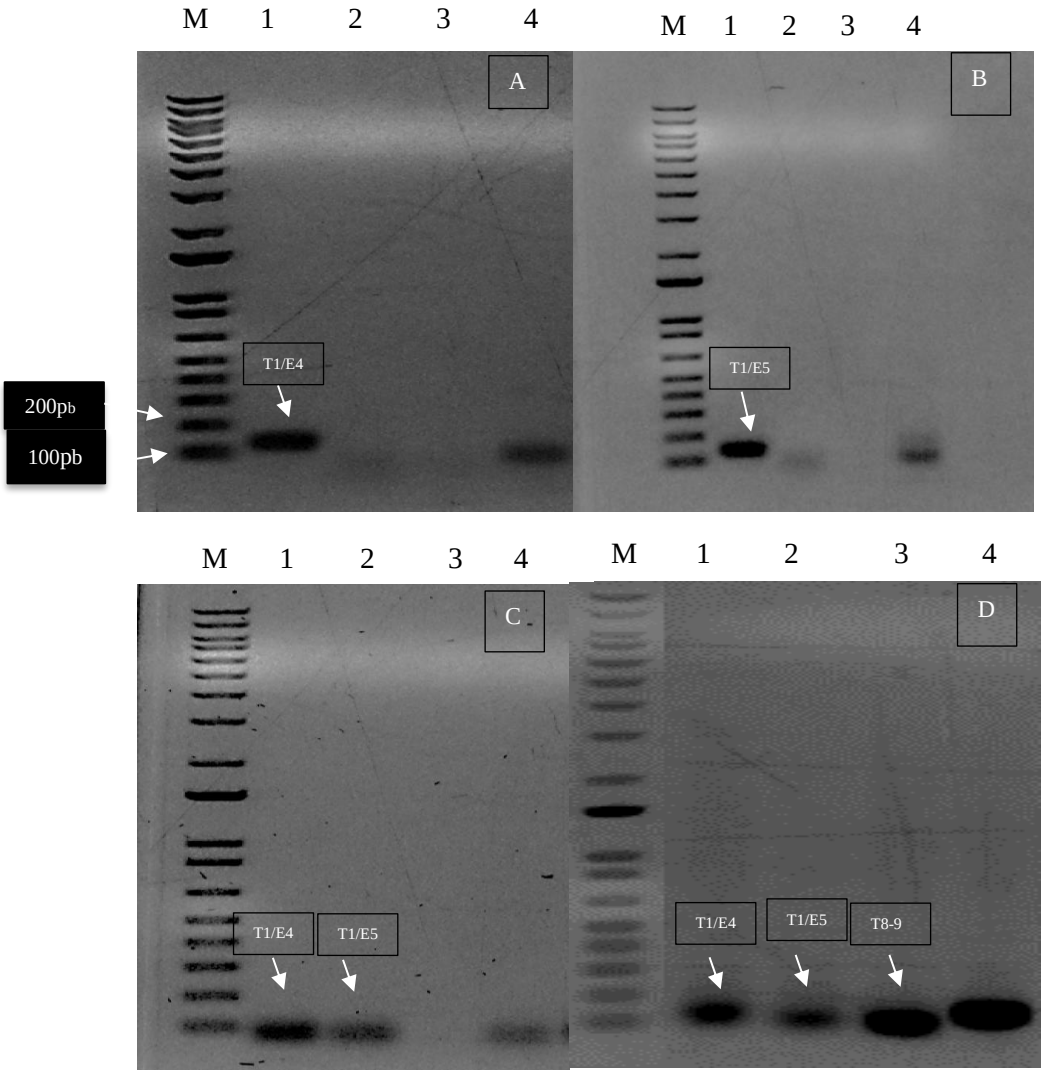


Figure 4. 7: Agarose gels showing RT-PCR screening of TMPRSS2-ERG fusion transcripts

(A): The arrows indicate the following transcript variants in this panel: T1/E4 (lanes 2)(126pb), T1/E5 (lane 3)(146pb) and T8-9 (lane 4) (75pb) and EIF1AX (lane 5). Lanes 3and 4 were negative for gene fusion. The DNA band in lane1 was sequenced and shown to be related to TMPRSS2 and ERG fusion gene.(example profile of patient positive for T1/E4 fusion)

(B): The arrows indicate the following transcript variants in this panel: T1/E5 (lanes 2), T1/E4 (lane 3) and T8-9 (lane 4) and EIF1AX (lane 5). Lanes 3 and 4 were negative for gene fusic The DNA band in lane1 was sequenced and shown to be related to TMPRSS2 and ERG fusion gene. (example profile of patient positive for T1/E5 fusion)

(C): The arrows indicate the following transcript variants in this panel: T1/E4 (lanes 2), T1/E5 (lane 3) and T8-9 (lane 4) and EIF1AX (lane 5). Lanes 4 was negative for TMPRSS2 wild type. The DNA bands in lanes2 and 3 was sequenced and shown to be related to TMPRSS2 and ERG fusion gene. (example profile of patient positive for both T1/E4 and T1/E5 fusions)

(D): The arrows indicate the following transcript variants in this panel: T1/E4 (lanes 2), T1/E5 (lane 3) and T8-9 (lane 4) and EIF1AX (lane 5). The DNA bands in lanes 2 and 3 was sequenced and shown to be related to TMPRSS2 and ERG fusion gene. (example profile of patient positive for both T1/E4 and T1/E5 fusions and present the wild type of TMPRSS2 gene)

All fusion transcripts were confirmed by nucleotide sequence analysis of the PCR products (Sanger sequencing) (Figure 4.8).

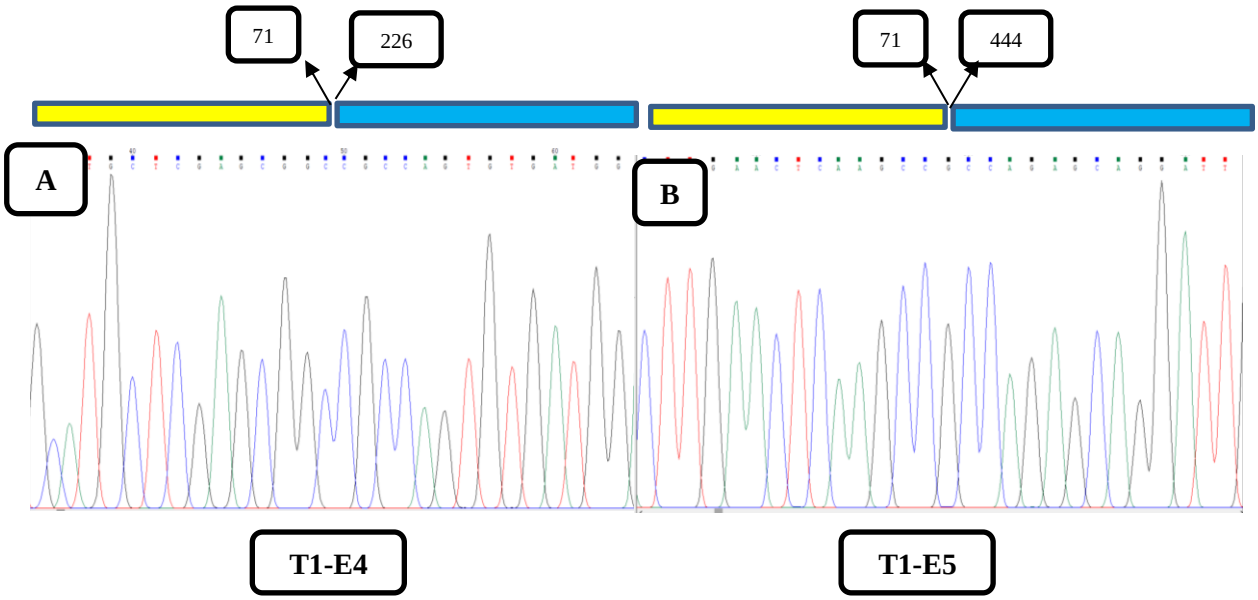


Figure 4. 8: Representative chromatograms of TMPRSS2-ERG fusion transcript sequencing. Exons (boxes) for TMPRSS2 (yellow) and ERG (blue) are numbered according to GenBank reference sequences (NM_005656, NM_004449). Line graphs show the position and automatic DNA sequencing result of fusion point. For each type of fusion, numbers above the fusion site are the last base pair in TMPRSS2 and the first base pair in ERG. **(A):** Position no. 71 of TMPRSS2 exon 1 was fused to position no. 226 of EGR1 exon 4. **(B):** Position no. 71 of TMPRSS2 exon 1 was fused to position no. 444 of EGR1 exon 5.

Our findings show that the *TMPRSS2-ERG* fusion is common in prostate cancer. However, the frequency of *TMPRSS-ERG*-fusions in the present study is somewhat higher than previously observed, (55) which indicates heterogeneity with regard to the expression of *TMPRSS2-ERG* gene fusions in prostate cancer. Moreover, the fusion positive/negative tumors may play a significant role as a prognostic marker in prostate cancer in our population as it has been shown in previous clinical follow-up studies (55).

The present investigation confirms and extends previous data showing that the *TMPRSS2-ERG* fusion is common in prostate cancer (33,55).

In agreement with the literature, we observed that the T1/E4 variant is the predominant transcript observed (32). The frequency of *ERG*-fusions in our study is somewhat higher than that reported, Tomlins et al in their original study (2) found *TMPRSS2-ERG* fusion transcripts in 19 of 42 (45%) samples of localized (32 samples) and metastatic (10 samples) prostate cancer and Winnes et al. (2007) found *TMPRSS2-ERG* fusion transcripts in 18 of the 50 tumors (36%)

(56). However, Our findings were similarly observed by Soller et al. (2006) who found *TMPRSS2-ERG* fusion in 14/18 cases (78%) (33). Our findings indicate heterogeneity with regard to the expression of ETS-gene fusions in subsets of prostate cancers and were similar to a recently reported study with a small series of 18 prostate cancer samples where the *TMPRSS2-ERG* fusion was found in 45-79% of prostate cancer samples. Although, we confirmed in our own patient cohort the observation that T1/E5 coexisted with T1/E4 in most cases (32) which indicates that these two variants are presumably alternatively spliced products from a single fusion. The reason why *ERG*-fusions are much more common in prostate cancer than other *ETS*-fusions is unknown. The frequency of variant gene fusions may depend on a number of factors, including the transcriptional orientation and the exon-intron organization of the partner genes as well as the size of the introns in which the breakpoints occur (32). All *TMPRSS2-ERG* fusion transcripts, except for the T2/E4 variant, can only encode truncated ERG proteins. The translational efficiency of these aberrant RNAs and the biological activity of the truncated proteins in prostate cancer remain to be investigated (55).

Moreover, some studies suggested that fusion-positive tumors may be associated with better survival and lower Gleason scores than fusion-negative tumors and Demichelis et al. (2007) (15) suggested that *TMPRSS2-ERG* fusion-positive carcinomas might represent a more aggressive phenotype However, additional studies are therefore warranted before firm conclusions can be made regarding the clinical significance of a molecular classification of prostate cancers into fusion-positive and fusion-negative tumors. (15)

It is known that some prostate carcinomas may not cause any trouble or only progress very slowly. However, there is an increasing need to identify patients with tumors that tend to progress rapidly to spare patients with slowly growing tumors the strain of overtreatment. For that purpose, *TMPRSS2-ERG* is used as a panel of genes whose expression is significantly involved in PCa carcinogenesis (57–60). In the African continent, only three studies have been conducted on these gene fusions in prostate cancer (45,47,61).

Conclusion

To the best of our knowledge, this is the fourth study of its kind in the African continent. We evaluated *TMPRSS2-ERG* fusion transcript variants by RT-PCR and Sanger sequencing and

successfully found this fusion gene in the Moroccan population. This study confirms that the presence of the *TMPRSS -ERG* fusion gene is a very common marker in prostate cancer.

Previous studies have shown that *TMPRSS2-ERG* fusion isoforms are significantly correlated with more aggressive disease which is in agreement with the outcomes of our study. The findings from this study will be useful for future research, and show the increased interest in integrating such biomarkers for prostate cancer aggressiveness detection and prostate cancer diagnosis treatment processes.

In conclusion, a combination of high preoperative serum PSA (which includes PSA level, clinical stage and Gleason score) and high expression of *TMPRSS2-ERG* could be promising in distinguishing those tumors that are aggressive and together with patient status make it possible to decide if radical resection is beneficial (57).

References

1. Prashanth Rawla. Épidémiologie du cancer de la prostate. Article de revue. Oncology [Internet]. 2019;32(1):2–4. Available from: doi: https://doi.org/10.14740/wjon1191
2. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science (80-). 2005;310(5748):644–8.
3. Hermans KG, Van Marion R, Van Dekken H, Jenster G, Van Weerden WM, Trapman J. TMPRSS2:ERG fusion by translocation or interstitial deletion is highly relevant in androgen-dependent prostate cancer, but is bypassed in late-stage androgen receptor-negative prostate cancer. Cancer Res. 2006;66(22):10658–63.
4. Anuradha Gopalan1, Margaret A. Leversha2, Jaya M. Satagopan3, Qin Zhou3, Hikmat A. Al- Ahmadie1, Samson W. Fine1, James A. Eastham4, Peter T. Scardino4, Howard I. Scher5, Satish K. Tickoo1, Victor E. Reuter1, and William L. Gerald1Anuradha Gopalan1, Marga and WLG. TMPRSS2-ERG gene fusion is not associated with outcome in patients treated by prostatectomy. Cancer Res. 2009;69(4):1400–1406.
5. Clinckemalie L, Spans L, Dubois V, Laurent M, Helsen C, Joniau S, et al. Androgen regulation of the TMPRSS2 gene and the effect of a SNP in an androgen response element. Mol Endocrinol. 2013;27(12):2028–40.
6. Kumar-Sinha C, Tomlins SA, Chinnaiyan AM. Recurrent gene fusions in prostate cancer. Nat Rev Cancer. 2008;8(7):497–511.
7. Adamo P, Ladomery MR. The oncogene ERG: A key factor in prostate cancer. Oncogene. 2016;35(4):403–14.
8. Ko CJ, Huang CC, Lin HY, Juan CP, Lan SW, Shyu HY, et al. Androgen-induced TMPRSS2 activates matriptase and promotes extracellular matrix degradation, prostate cancer cell invasion, tumor growth, and metastasis. Cancer Res. 2015;75(14):2949–60.
9. Hooper JD, Clements JA, Quigley JP, Antalis TM. Type II transmembrane serine proteases. Insights into an emerging class of cell surface proteolytic enzymes [Internet]. Vol. 276, Journal of Biological Chemistry. J Biol Chem; 2001 [cited 2020 Nov 21]. p. 857–60. Available from: https://pubmed.ncbi.nlm.nih.gov/11060317/
10. Netzel-Arnett S, Hooper JD, Szabo R, Madison EL, Quigley JP, Bugge TH, et al. Membrane anchored serine proteases: A rapidly expanding group of cell surface proteolytic enzymes with potential roles in cancer [Internet]. Vol. 22, Cancer and Metastasis Reviews. Cancer Metastasis Rev; 2003 [cited 2020 Nov 21]. p. 237–58. Available from: https://pubmed.ncbi.nlm.nih.gov/12784999/
11. Vaarala MH, Porvari K, Kyllönen A, Lukkarinen O, Vihko P. The TMPRSS2 gene encoding transmembrane serine protease is overexpressed in a majority of prostate cancer patients: Detection of mutated TMPRSS2 form in a case of aggressive disease. Int J Cancer [Internet]. 2001 Dec 1 [cited 2020 Nov 21];94(5):705–10. Available from: https://pubmed.ncbi.nlm.nih.gov/11745466/
12. Afar DEH, Vivanco I, Hubert RS, Kuo J, Chen E, Saffran DC, et al. Catalytic cleavage of the androgen-regulated TMPRSS2 protease results in its secretion by prostate and prostate cancer epithelia. Cancer Res. 2001;61(4):1686–92.
13. Biaoyang Lin, Camari Ferguson, James T. White, Shunyou Wang, Robert Vessella, Lawrence D. True LH, Nelson and PS. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2 [Internet]. CANCER RESEARCH. 1999. p. 4180–4184. Available from:

https://pubmed.ncbi.nlm.nih.gov/10485450/

14. Seth A, Watson DK. ETS transcription factors and their emerging roles in human cancer. Eur J Cancer [Internet]. 2005 [cited 2020 Nov 21];41(16):2462–78. Available from: https://pubmed.ncbi.nlm.nih.gov/16213704/
15. Demichelis F, Fall K, Perner S, Andrén O, Schmidt F, Setlur SR, et al. TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. Oncogene. 2007;26(31):4596–9.
16. Albadine R, Latour M, Toubaji A, Haffner M, Isaacs WB, A Platz E, et al. TMPRSS2-ERG gene fusion status in minute (minimal) prostatic adenocarcinoma. Mod Pathol. 2009;22(11):1415–22.
17. Liu W, Laitinen S, Khan S, Vihinen M, Kowalski J, Yu G, et al. Copy number analysis indicates monoclonal. Cancer. 2010;15(xxxxx):559–65.
18. C Sun, A Dobi, A Mohamed, H Li, R Thangapazham, B Furusato, S Shaheduzzaman, S-H Tan, G Vaidyanathan, E Whitman, DJ Hawksworth, Y Chen, M Nau, V Patel, M Vahey, JS Gutkind, T Sreenath, G Petrovics, I A Sesterhenn, DG McLeod SS. TMPRSS2-ERG Fusion, a Common Genomic Alteration in Prostate Cancer Activates C-MYC and Abrogates Prostate Epithelial Differentiation. Oncogene. 2008;27(40):5348–5353.
19. Tomlins SA, Mehra R, Rhodes DR, Smith LR, Roulston D, Helgeson BE, et al. TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. Cancer Res. 2006;66(7):3396–400.
20. Helgeson BE, Tomlins SA, Shah N, Laxman B, Cao Q, Prensner JR, et al. Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer. Cancer Res. 2008;68(1):73–80.
21. Paula Paulo, Joa˜o D. Barros-Silva, Franclim R. Ribeiro, Joa˜o Ramalho-Carvalho, Carmen Jero´nimo, Rui Henrique, Guro E. Lind, Rolf I. Skotheim RAL and MRT. FLI1 Is a Novel ETS Transcription Factor Involved in Gene Fusions in Prostate Cancer. Genes Chromosom Cancer. 2012;51:240–9.
22. Sashida G, Bazzoli E, Menendez S, Liu Y, Nimer SD. The oncogenic role of the ETS transcription factors MEF and ERG. Cell Cycle. 2010;9(17):3457–9.
23. Rosen P, Sesterhenn IA, Brassell SA, McLeod DG, Srivastava S, Dobi A. Clinical potential of the ERG oncoprotein in prostate cancer. Nat Rev Urol [Internet]. 2012;9(3):131–7. Available from: http://dx.doi.org/10.1038/nrurol.2012.10
24. Carver BS, Tran J, Chen Z, Carracedo-Perez A, Alimonti A, Nardella C, et al. ETS rearrangements do not initiate cancer in the prostate. Nature [Internet]. 2009;457(7231):E1–E1. Available from: http://www.nature.com/doifinder/10.1038/nature07738
25. Tomlins SA, Bjartell A, Chinnaiyan AM, Jenster G, Nam RK, Rubin MA, et al. ETS Gene Fusions in Prostate Cancer: From Discovery to Daily Clinical Practice [Internet]. Vol. 56, European Urology. Eur Urol; 2009 [cited 2020 Nov 21]. p. 275–86. Available from: https://pubmed.ncbi.nlm.nih.gov/19409690/
26. Rajput AB, Miller MA, De Luca A, Boyd N, Leung S, Hurtado-Coll A, et al. Frequency of the TMPRSS2:ERG gene fusion is increased in moderate to poorly differentiated prostate cancers. J Clin Pathol. 2007;60(11):1238–43.
27. Zong Y, Xin L, Goldstein AS, Lawson DA, Teitell MA, Witte ON. ETS family transcription factors collaborate with alternative signaling pathways to induce carcinoma from adult murine prostate cells. Proc Natl Acad Sci U S A. 2009;106(30):12465–70.

28. Hermans KG, Van Marion R, Van Dekken H, Jenster G, Van Weerden WM, Trapman J. TMPRSS2:ERG fusion by translocation or interstitial deletion is highly relevant in androgen-dependent prostate cancer, but is bypassed in late-stage androgen receptor-negative prostate cancer. Cancer Res [Internet]. 2006 Nov 15 [cited 2020 Nov 21];66(22):10658–63. Available from: https://pubmed.ncbi.nlm.nih.gov/17108102/
29. Manzoor S. CAPITAL UNIVERSITY OF SCIENCE AND The Study of TMPRSS-ERG Gene Fusion in Prostate Cancer in Pakistan by. 2018;
30. Murakami K, Mavrothalassitis G, Bhat NK, Fisher RJ, Papas TS. Human ERG-2 protein is a phosphorylated DNA-binding protein--a distinct member of the ets family. Oncogene. 1993 Jun;8(6):1559–66.
31. G Attard, J Clark1, L Ambroisine, G Fisher, G Kovacs, P Flohr, D Berney, CS, Foster, A Fletcher, WL Gerald, H Moller, V Reuter, JS De Bono, P Scardino, J G Attard, J Clark, L Ambroisine, G Fisher, G Kovacs, P Flohr, D Berney, CS Cuzick and CC on behalf of the TPG. Duplication of the fusion of TMPRSS2 to ERG sequences identifies fatal human prostate cancer. Oncogene. 2008;27(4):253–263.
32. Clark J, Merson S, Jhavar S, Flohr P, Edwards S, Foster CS, et al. Diversity of TMPRSS2-ERG fusion transcripts in the human prostate. Oncogene. 2007;26(18):2667–73.
33. Panagopoulos I, Soller MJ, Isaksson M, Elfving P, Soller W, Lundgren R. Confirmation of the high frequency of the TMPRSS2/ERG fusion gene in prostate cancer [Internet]. Vol. 45, Genes Chromosomes and Cancer. Genes Chromosomes Cancer; 2006 [cited 2020 Nov 21]. p. 717–9. Available from: https://pubmed.ncbi.nlm.nih.gov/16575875/
34. Wang J, Cai Y, Ren C, Ittmann M. Expression of Variant TMPRSS2 / ERG Fusion Messenger RNAs Is Associated with Aggressive Prostate Cancer. 2006;(17):8347–52.
35. Yoshimoto M, Joshua AM, Chilton-MacNeill S, Bayani J, Selvarajah S, Evans AJ, et al. Three-color FISH analysis of TMPRSS2/ERG fusions in prostate cancer indicates that genomic microdeletion of chromosome 21 is associated with rearrangement. Neoplasia [Internet]. 2006 [cited 2020 Nov 21];8(6):465–9. Available from: /pmc/articles/PMC1601467/?report=abstract
36. St. John J, Powell K, Katie Conley-LaComb M, Chinni SR. TMPRSS2-ERG fusion gene expression in prostate tumor cells and its clinical and biological significance in prostate cancer progression [Internet]. Vol. 4, Journal of Cancer Science and Therapy. J Cancer Sci Ther; 2012 [cited 2020 Nov 21]. p. 94–101. Available from: https://pubmed.ncbi.nlm.nih.gov/23264855/
37. Hessels D, Smit FP, Verhaegh GW, Witjes JA, Cornel EB, Schalken JA. Detection of TMPRSS2-ERG fusion transcripts and prostate cancer antigen 3 in urinary sediments may improve diagnosis of prostate cancer. Clin Cancer Res [Internet]. 2007 Sep 1 [cited 2020 Nov 21];13(17):5103–8. Available from: https://pubmed.ncbi.nlm.nih.gov/17785564/
38. Perner S, Mosquera JM, Demichelis F, Hofer MD, Paris PL, Simko J, et al. TMPRSS2-ERG fusion prostate cancer: An early molecular event associated with invasion. Am J Surg Pathol [Internet]. 2007 Jun [cited 2020 Nov 21];31(6):882–8. Available from: https://pubmed.ncbi.nlm.nih.gov/17527075/
39. Kish EK, Choudhry M, Gamallat Y, Buharideen SM, Dhananjaya D, Bismar TA. The Expression of Proto-Oncogene ETS-Related Gene (ERG) Plays a Central Role in the Oncogenic Mechanism Involved in the Development and Progression of Prostate Cancer. Int J Mol Sci. 2022;23(9).

40. Delliaux C, Tian T, Bouchet M, Fradet A, Flourens A, Deplus R, et al. TMPRSS2 : ERG gene fusion expression regulates bone markers and enhances the osteoblastic phenotype of prostate cancer bone metastases To cite this version : HAL Id : hal-03060052. 2020;
41. Hänze J, Rexin P, Jakubowski P, Schreiber H, Heers H, Lingelbach S, et al. Prostate cancer tissues with positive TMPRSS2-ERG-gene-fusion status may display enhanced nerve density. Urol Oncol Semin Orig Investig. 2020;38(1):3.e7-3.e15.
42. Knuuttila M, Mehmood A, Ryberg H, Knaapila J, Ettala O, Ohlsson C, et al. INTRATUMORAL ANDROGEN LEVELS ARE LINKED TO TMPRSS2-. 2018;(May):1–30.
43. Graff RE, Ahearn TU, Pettersson A, Ebot EM, Gerke T, Penney KL, et al. Height, obesity, and the risk of TMPRSS2: ERG-Defined prostate cancer. Cancer Epidemiol Biomarkers Prev. 2018;27(2):193–200.
44. Holly JMP, Broadhurst J, Mansor R, Bahl A, Perks CM. Hyperglycemia promotes TMPRSS2-ERG gene fusion in prostate cancer cells via upregulating insulin-like growth factor-binding protein-2. Front Endocrinol (Lausanne). 2017;8(NOV):1–11.
45. Blackburn J, Vecchiarelli S, Heyer EE, Patrick SM, Lyons RJ, Jaratlerdsiri W, et al. TMPRSS2-ERG fusions linked to prostate cancer racial health disparities: A focus on Africa. Prostate. 2019;79(10):1191–6.
46. Wang Y, Freedman JA, Liu H, Moorman PG, Hyslop T, George DJ, et al. Associations between RNA splicing regulatory variants of stemness-related genes and racial disparities in susceptibility to prostate cancer. Int J Cancer. 2017;141(4):731–43.
47. Abdel-Hady A, El-Hindawi A, Hammam O, Khalil H, Diab S, Abd El-Aziz S, et al. Expression of ERG protein and TMRPSS2-ERG fusion in prostatic carcinoma in Egyptian patients. Open Access Maced J Med Sci. 2017;5(2):147–54.
48. Adeloye D, David RA, Aderemi AV, Iseolorunkanmi A, Oyedokun A, Iweala EEJ, et al. An estimate of the incidence of prostate cancer in Africa: A systematic review and meta-analysis. PLoS One. 2016;11(4):1–18.
49. Odedina FT, Akinremi TO, Chinegwundoh F, Roberts R, Yu D, Reams RR, et al. Prostate cancer disparities in Black men of African descent: A comparative literature review of prostate cancer burden among Black men in the United States, Caribbean, United Kingdom, and West Africa. Infect Agent Cancer. 2009;4(SUPPL. 1):1–8.
50. Kretschmer A, Tilki D. Biomarkers in prostate cancer – Current clinical utility and future perspectives. Crit Rev Oncol Hematol [Internet]. 2017;120(September):180–93. Available from: https://doi.org/10.1016/j.critrevonc.2017.11.007
51. Tian JY, Guo FJ, Zheng GY, Ahmad A. Prostate cancer: Updates on current strategies for screening, diagnosis and clinical implications of treatment modalities. Carcinogenesis. 2018;39(3):307–17.
52. Angeles A, Bauer S, Ratz L, Klauck S, Sültmann H. Genome-Based Classification and Therapy of Prostate Cancer. Diagnostics [Internet]. 2018 Sep 2 [cited 2020 Nov 21];8(3):62. Available from: /pmc/articles/PMC6164491/?report=abstract
53. Lorenzin F, Demichelis F. Past, Current, and Future Strategies to Target ERG Fusion-Positive Prostate Cancer. Cancers (Basel). 2022;14(5):1–22.
54. Singareddy R, Semaan L, Katie Conley-Lacomb M, John JS, Powell K, Iyer M, et al. Transcriptional regulation of CXCR4 in prostate cancer: Significance of TMPRSS2-ERG fusions. Mol Cancer Res. 2013;11(11):1349–61.
55. Tu JJ, Rohan S, Kao J, Kitabayashi N, Mathew S, Chen Y. Gene fusions between TMPRSS2 and ETS family genes in prostate cancer : frequency and transcript variant

analysis by RT-PCR and FISH on paraffin-embedded tissues. 2007;29:921–8.

56. Winnes M, Lissbrant E, Damber JE, Stenman G. Molecular genetic analyses of the TMPRSS2-ERG and TMPRSS2-ETV1 gene fusions in 50 cases of prostate cancer. Oncol Rep. 2007;17(5):1033–6.
57. Kulda V, Topolcan O, Kucera R, Kripnerova M, Srbecka K, Hora M, et al. Prognostic significance of tmprss2-erg fusion gene in prostate cancer. Anticancer Res. 2016;36(9):4787–93.
58. Shahana Majid, Altaf A Dar, Sharanjot Saini, Sumit Arora, Varahram Shahryari, Mohd Saif Zaman, Inik Chang, Soichiro Yamamura, Yuichiro Tanaka, Guoren Deng and RD. MicroRNA-23b represses proto-oncogene Src kinase and functions as methylation-silenced tumor suppressor with diagnostic and prognostic significance in prostate cancer. Cancer Res. 2012;72(24):6435–46.
59. Kato M, Goto Y, Matsushita R, Kurozumi A, Fukumoto I, Nishikawa R, et al. MicroRNA-26a/b directly regulate La-related protein 1 and inhibit cancer cell invasion in prostate cancer. Int J Oncol. 2015;47(2):710–8.
60. Kneitz B, Krebs M, Kalogirou C, Schubert M, Joniau S, Van Poppel H, et al. Survival in patients with high-risk prostate cancer is predicted by mir-221, which regulates proliferation, apoptosis, and invasion of prostate cancer cells by inhibiting IRF2 and SOCS3. Cancer Res. 2014;74(9):2591–603.
61. Zhou CK, Young D, Yeboah ED, Coburn SB, Tettey Y, Biritwum RB, et al. TMPRSS2:ERG gene fusions in prostate cancer of west African men and aMeta-analysis of racial differences. Am J Epidemiol. 2017;186(12):1352–61.

Chapter 5: Whole Exome Sequencing of Moroccan Prostate Cancer Patients

Chapter 5: Whole Exome Sequencing of Moroccan Prostate Cancer Patients

I. Introduction

Prostate cancer (PCa) is genetically very heterogeneous. Genetic factors play a crucial role and have been estimated to explain almost 58% of the cancer risk (1). The heterogeneity of this cancer has made the identification of genetic susceptibility factors extremely challenging. Only a few high-risk candidate genes and variants have been found, and the risk effect of the more common variants is typically low (1). Today, whole exome sequencing is the most efficient method to identify novel and unique disease-causing mutations, because most known mutations that cause disease occur in exons.

Whole exome sequencing (WES) of tumors (cases) and matched normal tissue (controls) has robustly examined the genomic landscape of primary and metastatic prostate cancer (2). However, the spectrum of statistically mutated genes occurring at lower frequencies and associated with prostate cancer is not completely known. Larger WES studies have been suggested to identify those genes involved in prostate cancer that harbor substantial interpatient genomic heterogeneity (3). The results of meta-analyses of WES data have been integrated and transferred to a clinical interpretation of germline genetics (4).

Advances in next generation sequencing (NGS) have led to a spectacular transformation in the diagnosis and treatment of this disease due to the understanding of the molecular mechanisms that underlie cancer. In this sense, a new paradigm has been developed called ‘Precision Medicine’ which is a personalized treatment for each patient based on his genetic profile and is capable of attacking a specific tumor with higher efficiency. Genomics and bioinformatics are two essential pillars in this applied field. The massive data (Big Data) generated by these technologies, require powerful analytical tools and trained scientists to perform, analyze, integrate and transfer that genomic information to physicians.

The paradigm of precision medicine has been supplemented by all the recent advances in genomics and bioinformatics. The strategy of this discipline requires the genomic and bioinformatics fields to determine the molecular mechanisms of tumorigenesis and tumor resistance to the potential applied treatment. In the last decade, progress in bioinformatics and computational biology has allowed moving the precision medicine model from the conceptual

to the pragmatic scenario (5).

In this study, we proposed using WES and bioinformatics to identify novel variants and unique genetic features in Moroccan prostate cancer patients. The study aimed to identify the genomic variant pattern of PCa in the Moroccan population, which will lead to identification of better prognostic markers and therapeutic targets for personalized treatment. We performed WES to cover all possible markers in order to obtain a good overview of prostate cancer genomics in the Moroccan population. DNA was extracted from tissue samples for WES and genomic analysis was performed to determine the different variants of twelve Moroccan prostate cancer patients with high-risk disease and identify new markers, mutated genes and oncogenic signaling pathways to characterize the abnormal exome regions using bioinformatics tools. This will be the first time such an approach has been conducted in Morocco and will provide a genomic landscape of prostate cancer in the Moroccan population that can be applied to novel prognostic and predictive assays in prostate cancer management for Moroccan patients and can help in stratified treatment strategies and personalized follow-up for patients. To the best of our knowledge, this is the largest prostate cancer exome study of the Moroccan population that provides a robust characterization of the somatic landscape within Moroccan prostate cancer, notwithstanding the challenges and limitations of this study.

II. Background

II.1 Prostate Cancer Genomics

In the past 25 years advancements in genomics have placed a major emphasis on the causative role of somatic aberrations in tumorigenesis. The ability to study the underlying genomic alterations that distinguish cancers from normal tissue has been increased due to the progression of NGS technologies. This has led to a widespread paradigm shift in the way in which both cancer tumorigenesis and cancer therapeutics are conceptualized. In 2000, Hanahan & Weinberg delineated the currently accepted mechanistic steps involved in cancer development (6).

For prostate cancer, some criteria have been investigated as multiple somatic events that occur as normal or benign tissues transform into malignant disease including :self-sufficiency in growth signals (7), evading apoptosis (8), sustained angiogenesis (9), limitless replicative

potential (10), tissue invasion and metastasis (11).

Due to genetic and phenotypic heterogeneity, the identification of prostate cancer susceptibility genes and variants associated with an increased cancer risk has been challenging. Gene rearrangements and copy-number changes characterize much of the genomic landscape of this disease (12,13). In addition, PCa has a low mutational rate compared to other cancers (14). As a result, some recurrent mutations in genes observed in other cancers such as *KRAS*, *IDH1*/*2*, *EGFR*, and *PIK3CA* are absent or relatively uncommon in the landscape of PCa (14).

Several prostate cancer genes and mutations have been discovered through WGS and WES. A recent study using exome sequencing found low levels of recurrent mutations in *BRAF*, *RET*, and multiple androgen receptor cofactor genes, including *FOXA1*, *MLL2*, and *MLL* (15). Moreover, the ubiquitin ligase *SPOP* gene has been discovered through WGS of PCa (13), and found to be mutated in ~5% of all prostate cancers (~10% of ETS-negative cancers).

Amplification of several oncogenes also helps to drive prostate cancer pathogenesis. The *cMYC* oncogene is frequently amplified as part of an arm-level copy number gain of Chr8q, leading to its transcriptional upregulation in cancer (16).

II.1.1 Genome-wide association studies and Prostate Cancer

The genome-wide association studies (GWAS) are effective in finding common disease alleles. One of the first studies that applied GWAS in prostate cancer genetics reported a disease-associating variant on 8q24 and subsequent analyses confirmed the association (17).

Since 2006, a vast number of GWASs and meta-analyses combining the results from individual studies have been performed and numerous prostate cancer-associated single nucleotide polymorphisms (SNPs) have been published. The findings are listed in a manually curated, quality controlled GWAS Catalog (www.ebi.ac.uk/gwas/), developed in collaboration between the National Human Genome Research Institute (NHGRI) and the European Bioinformatics Institute (EMBL-EBI) (18). Currently, the catalog contains results from 28 GWASs reporting 193 SNPs that associate statistically significantly ($p \leq 1.0 \times 10^{-5}$) with prostate cancer (accessed: 26 Nov, 2015). According to the GWAS Catalog, prostate-cancer-associated SNPs have been detected on all chromosomes except for the Y chromosome.

Furthermore, CNVs can promote tumorigenesis by several mechanisms: a tumor suppressor gene or a DNA repair gene can be deleted, an oncogene can be amplified, or a regulatory

element can be removed or introduced to a new genomic location, thereby leading to aberrant gene expression (19,20). The involvement of germline CNVs in prostate cancer susceptibility has also been investigated, and a few statistically significant associations have been identified. Two of these loci, 2p24.3 and 20p13, were shown to be associated with an aggressive form of the disease (21,22).

In contrast to inherited germline changes, the number of somatic CNVs detected in prostate tumors is remarkably high. A recent meta-analysis combined the results from eleven CNV studies performed on 662 primary or advanced prostate tumors and reported 14 recurrent deletions and five recurrent gains in the dataset. The most frequent somatic copy number change was the deletion of chromosome 8p. Other common CNVs included gain of 8q, losses at 2q, 3p, 5q, 6q, 13q, 16q, 17p and 18q, deletions involving the *PTEN* gene, and *TMPRSS2-ERG* gene fusions (23).

To date, genome-wide association (GWA) studies, which are focused on investigating the association between common genetic variation and disease, have searched for SNPs that affect the progression from normal prostate tissue to tumor tissue (Figure 5.1). SNPs on chromosome 8q24 were one of several highly replicated SNPs for prostate cancer that have been detected. However, the potential use of these SNPs at diagnosis remains unclear. An increasing frequency of somatic gene fusions between the transmembrane protease serine 2 gene (*TMPRSS2*) and members of the erythroblast-transformation specific (*ETS*) gene family, has been observed across the different steps of prostate cancer progression. This *TMPRSS2-ERG* fusion has been labelled ‘signature’ of prostate cancer aggressiveness. The common *TMPRSS2–ERG* fusion has similar frequencies among localized and metastatic tumors, although the fusions in metastatic tumors primarily exhibit interstitial deletions. Further studies are needed to completely understand and distinguish which germline and somatic factors affect the development of prostate cancer after it is diagnosed, which will help to choose the best treatment for each patient (24).

.

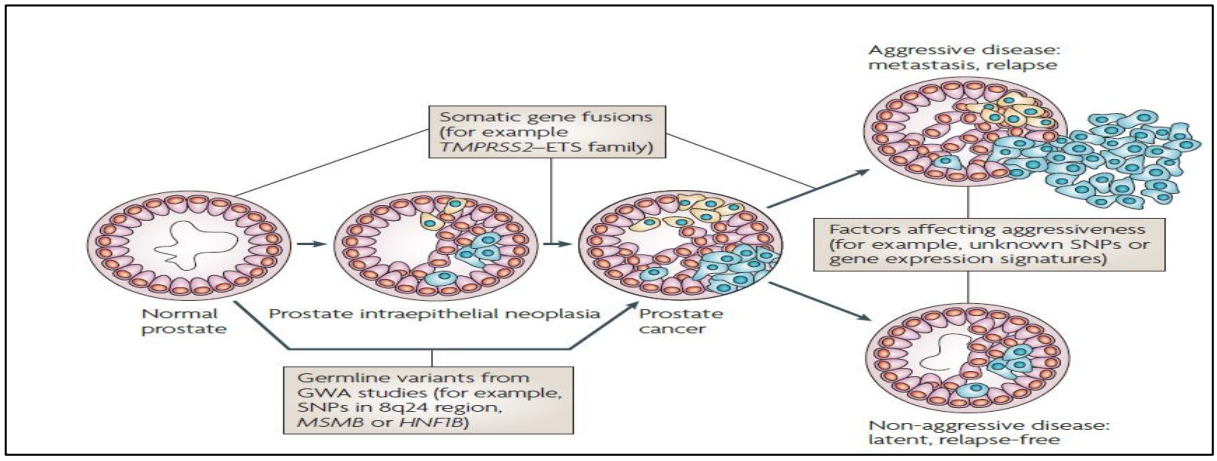


Figure 5. 1: Etiologic evolution of prostate cancer and key steps affected by germline variants and somatic fusions; MSMB: β-microseminoprotein; HNF1B: HNF1 homeobox B (24).

I.2 Next-generation sequencing technologies and applications

Today, the emergence of the big-data era has seen the day due to the availability of genetic information and expansion in biomedical publications offered by NGS technologies in the field of cancer genomics (25).

Next-generation sequencing has played an enormous role in the progress of medical research by enabling studies generating multi-omics data such as proteome, epigenome, and microbiome data. The integrative analysis of multi-omics data provides a more comprehensive view of biological processes leading to a better understanding of these systems compared to single-layer analysis (26,27).

The NGS platforms enable the simultaneous analysis of multiple genes of interest at substantially lower costs. Currently, the sequencing of the entire human genome can be done in a few hours, and, according to NHGRI's Genome Sequencing Program, the average cost is 1,363 US dollars (www.genome.gov/sequencingcosts). The costs are expected to decrease even further as Illumina (San Diego, CA, USA), a leader in the DNA sequencing industry, has recently presented the new HiSeq X Ten Sequencing System, which is able to sequence 18,000 human genomes per year (49 genomes per day) at the price of 1,000 dollars per genome (www.illumina.com/systems/hiseq-X-sequencingsystem/).

NGS data analysis begins with base-calling, the translation of the sequencing signal into base sequences. Then, sequence reads are either assembled *de novo* (built from scratch) or aligned to a reference sequence or genome. The next step, variant calling, aims to identify genomic

alterations by comparing the correctly targeted reads to their reference sequence (28).

Different variant calling algorithms may be required to reliably recognize different types of variations, but in principle, NGS methods are able to detect single-nucleotide variants, small and large insertions and deletions, copy number variants and gene fusions simultaneously (29). NGS methods typically provide tens or hundreds of sequence reads representing each target region, which increases the sensitivity and reliability of mutation detection.

Sequencing coverage or depth describes the average number of times a base pair has been sequenced. To overcome biases resulting from sequencing errors and uneven read distribution across the reference sequence, a coverage of approximately 30x to 40x is recommended for the accurate identification of variants (30).

I.3 Whole Exome Sequencing

Whole-exome sequencing (WES) has become a very efficient technology that can be performed for less than $1,000 and limited to the protein-coding regions only. These protein accounts for approximately 1% of the entire genome (12). Compared to WGS, for which the cost still remains high, WES is a cost-effective and highly sensitive mutation detection approach, as it covers only limited exon coding sequences (31). Indeed, useful genomic data for disease diagnosis, such as identifying the genes responsible for Mendelian disorders, have already been provided by WES (32).

According to Spans & colleagues, more than 200 prostate cancer-related exome sequencing reports have already been published. Most of these studies have investigated the different stages of prostate cancer tumorigenesis as well as the progression of the disease to CRPC by sequencing tumor cell exomes (33). WES has also been used for the exploration of genetic predisposition to prostate cancer, and has identified susceptibility genes and variants such as: germline missense variants in the *BTNL2 (Butyrophilin-like 2)* gene (34).

Henceforth, WES may be replaced by WGS with the speed of technological advances, which will be cheaper in practice and require less experimental manipulation. However, such a scenario will certainly increase the complexity of the bioinformatic analysis (35).

There are some factors that have made WES more practical in clinical and diagnostic settings than WGS in addition to the cost. Which are the compacted size of the aligned whole exome BAM file with the aligned whole genome file, which means lower data storage cost and lower computational power are needed for alignment (36). Moreover, the WES strategy is often

employed to only identify those variants in protein coding regions which are the most relevant because it has been estimated that 85% of disease causing variants are within protein coding or regulatory regions (37).

I.4 Bioinformatics and cancer genomics

The progress in NGS technologies has increased the availability and cost of these technologies. Although, the large amount of data that NGS platforms generate makes genomics and bioinformatics fall within the disciplines that require big data approaches (38).

The term "bioinformatics" is a field that gathered different fields including: biology, medicine, computer science, mathematics, statistics, and information technology. In biomedical sciences, bioinformatics tools and databases constitute an essential component of the current research process (39).

The current next generation sequencing platforms (NGS), allow the epigenomics, genomics, and transcriptomics characterization of a specific tumor. The integrity and concentration of DNA and RNA obtained either from tissue tumor biopsies, or liquid biopsies (circulating cell-free DNA or RNA derived from tumor exosomes) are determine before carrying out the quality control. Specific library construction protocols are needed for the DNA/RNA samples according to the user needs [e.g., whole genome sequencing (WGSeq), whole exome sequencing (WESeq), TargetedSeq, MethylSeq, RNAseq, small RNAseq] using different sequencing platforms (Illumina: MySeq, HiSeq, and NovaSeq series; ThermoFisher: PGM, Ion Proton, Ion S5 series, etc.) (40). The FASTQ files undergo pre-processing, which includes demultiplexing, quality control and adapter trimming. These processed files are put through a pipeline that keeps information that is deemed important, such as pathways and variants that are deleterious (Figure 5.2).

Terabytes of raw data (FASTQ files) are generated from whole-exome or transcriptome sequencing for only a small project (10 to 30 samples). These data are pre-processed and aligned against a reference genome (BAM files for binary version of sequence alignment) that are usually on the gigabyte scale depending on the sequencing coverage (average number of unique reads that include a given nucleotide in the reconstructed sequence) and the length of reads (38).

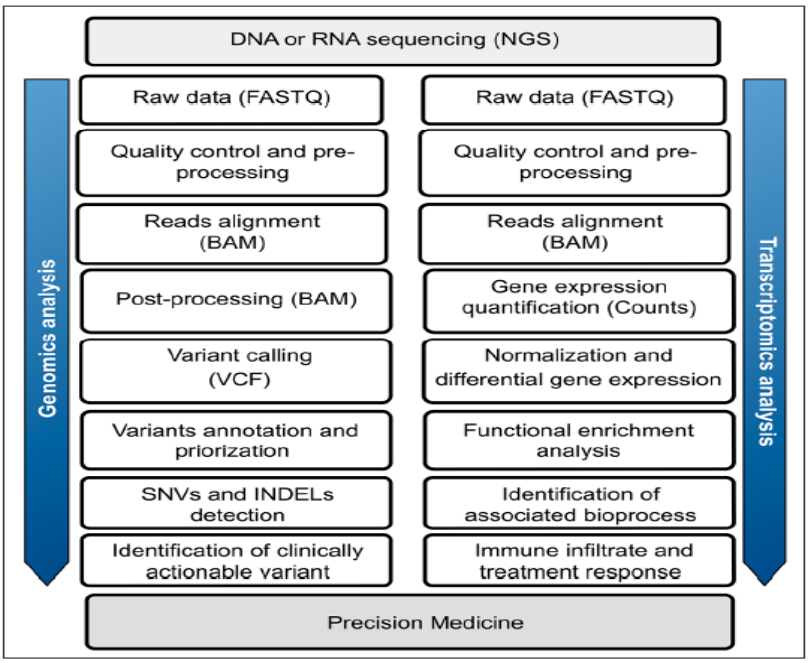


Figure 5. 2: Bioinformatics workflows for NGS (Next Generation Sequencing) data analysis (Exome-Seq in the left and RNA-Seq in the right). FASTQ: TXT file with the reads sequence and quality, BAM: binary alignment map; SNV: single nucleotide variants; INDEL: insertion or deletion variants (5).

II.4.1 Phases of NGS data analysis

There are varying demands of computational capability, the most being required by pre-processing and alignment, and the least by other downstream bioinformatics actions. Various public consortiums and companies are developing cloud-based bioinformatics platforms to facilitate the implementation of the required workflows with on-demand computational capability. The implementation of these workflows and computational infrastructure needed for personalized genome and exome -and other omics- analysis will be a crucial step into establishing precision medicine in the health system (Figure 5.3).

There are two phases of NGS data analysis, namely primary and secondary analysis. Primary analysis generally defines the process by which instrument-specific sequencing measures are transformed into FASTQ files comprising the short-read sequence data and sequencing run quality control metrics. Secondary analysis involves alignment of these sequence reads to a human reference genome and detection of variations between the patient sample and the reference. The most frequently employed secondary analysis approach assimilates five sequential steps. These include: initial read alignment (42), elimination of duplicate reads (deduplication) (43), local realignment around known indels (44), recalibration of the base quality scores (45), and variant detection as well as genotyping (46).

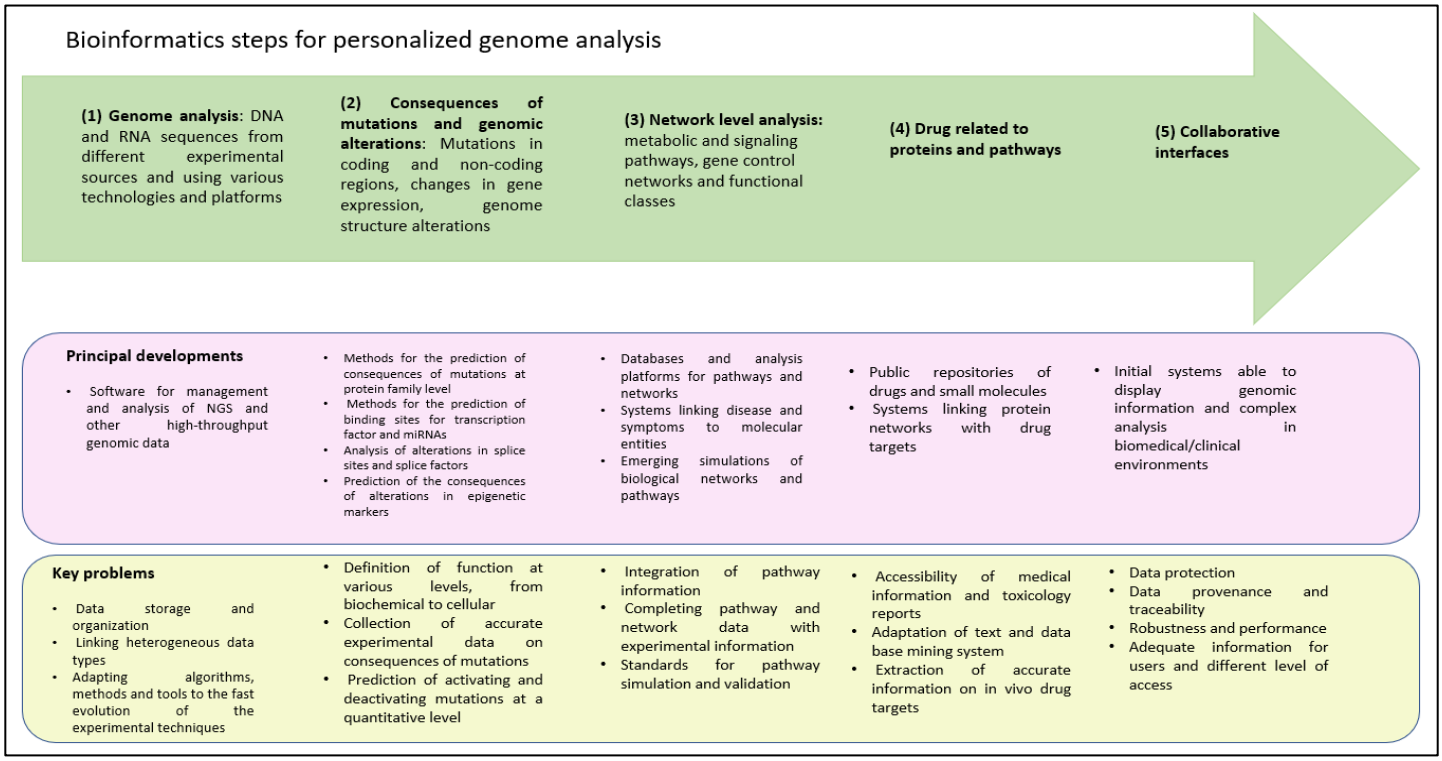


Figure 5. 3: Scheme of a comprehensive bioinformatics pipeline to analyze personalized genomic information. The five steps in the pipeline are shown in the top row, with the main methods that have so far been developed for each step the middle and outstanding problems in the bottom row. (1) Revision of genomic information. In this rapidly developing area methods and software are continuously changing to match the improvements in sequencing technologies. (2) Analysis of the consequences of specific mutations and genomic alterations. The analysis needs go from the area of point mutation prediction in proteins to the much more challenging area of prediction of mutations in non-coding regions, including promoter regions and TF binding sites. Other genetic alterations important in cancer must also be taken into consideration, such as copy number variation, modification of splice sites and altered splicing patterns. (3) Mapping of gene/protein variants at the network level. At this point, the relationships between individual components (genes and proteins) are analyzed in terms of their involvement in gene control networks, protein interaction maps and signaling/metabolic pathways. It is clearly necessary to develop a network analysis infrastructure and analysis methods capable of extracting information from heterogeneous data sources. (4) Translation of the information into potential drugs or treatments. The pharmacogenomic analysis of the information is essential to identify potential drugs or treatments. The analysis at this level integrates genomic information with that obtained from databases linking drugs and potential targets, combining it with data on clinical trials drawn from text or web sources. Toxicogenomics information adds an interesting dimension that enables additional exploration of the data. (5) Finally, it is essential to make the information extracted by the systems accessible to the end users in adequate conditions, including geneticists, biomedical scientists and clinicians (41).

A variant call format file is the final output of this process, which will be analyzed to discover the clinically relevant variants. Of the stages of human genome sequencing data analysis, secondary analysis is by far the most computationally demanding. This is a result of the size of the files that must be altered, to establish optimum alignments of millions of reads to the human reference genome, and sequential alignments for variant calling and genotyping. Abundant

software tools have been generated to carry out secondary analysis steps, each with conflicting strengths and weaknesses. Of the multitudinous aligners available (47), the Burrows-Wheeler transform based alignment algorithm (BWA), is most frequently used due to its precision, speed and ability to output Sequence Alignment/Map (SAM) formats (48). Usually, for the post alignment processing steps, Picard and SAMtools are the most used tools and they output SAM binary (BAM) format files (48). Considerable statistical methods have been devised for variant calling and genotyping in NGS studies (49), with the Genome Analysis Toolkit (GATK), being the most prominent (46). The bulk of NGS studies integrate BWA, Picard, SAMtools and GATK to pinpoint and genotype variants (42). Despite all these incredible advancements, there remain scientific questions pertaining to sample enrichment, sequencing methodologies and variant discovery and calling algorithms, which still require careful scrutiny in order to validate the analytical steps of NGS techniques for clinical applications. Nonetheless, the fundamental predictive challenge lies within the elucidation of the clinical significance of the variants observed in a given patient, and their significance for family members and for other patients. Every step in the variant clarification process has impediments and complications, such as their contributions to false positive and negative results. There is no single piece of information enough on its own to make concrete conclusions regarding the pathogenicity and disease causality of a given variant (50).

The molecular mechanisms of tumorigenesis are a real challenge for the development of new bioinformatics tools for diagnostics and systems biology modeling. The trend in recent medical genomics studies is the integration of bioinformatics data on gene expression, mutagenesis, and pathway interaction analysis (51,52).

I.5 Databases and Web Tools for Cancer Genomics Study

The effective use of massive amounts of cancer genome data remains a challenge due to the limitations of computational methodologies and insufficient collaboration and sharing (Figure 5.4) (53). These components present basically the future of cancer research because the integration, collaboration and sharing of the genomic and omics data in general will complete the genomic landscape and the racial disparities picture between the populations worldwide.

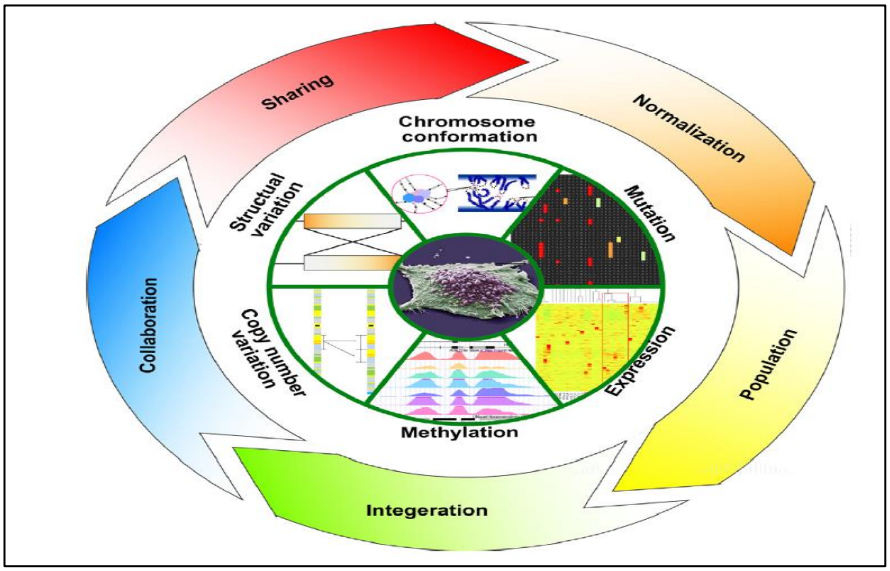


Figure 5. 4: The future of cancer research (53)

There are several popular and effective web-based cancer genomics data repositories, along with tools and resources (Table 5.1) to manage and analyze these data (53).

Table 5. 1 : Major web resources for cancer genomics research (53)

Name	Link	Main features	Refs.
CGHub	https://cghub.ucsc.edu/	Comprehensive data repository; huge data size	(54)
EGA	https://www.ebi.ac.uk/ega/	Comprehensive data repository; huge data size	
COSMIC	http://cancer.sanger.ac.uk	Largest somatic mutation database; genome sequencing paper curation	(55)
CPRG	http://www.broadinstitute.org/software/cprg	Interface for cancer program resources	.
GDAC	http://gdac.broadinstitute.org/	Data analysis; automatic pipelines; user-friendly reports	(56)
SNP500Cancer	http://snp500cancer.nci.nih.gov	Sequence and genotype verification of SNPs	(57)
canEvolve	www.canevolve.org/	Comprehensive analysis of tumor profile; Data from 90 studies involving more than 10,000 patients	(58)
MethyCancer	http://methycancer.psych.ac.cn	Relationship among DNA methylation, gene expression and cancer	(59)
SomamiR	http://compbio.uthsc.edu/SomamiR/	Correlation between somatic mutation and microRNA; genome-wide displaying	(60) .
cBioPortal	http://www.cbioportal.org/public-portal/	Graphical summaries; gene alteration; processed data; visualization	(61)
UCSC Cancer Genomics Browser	https://genome-cancer.soe.ucsc.edu/	Clinical information; gene expression; copy number variation; visualization	(62)

CGWB	https://cgwb.nci.nih.gov/	Visualization; gene mutation and variation; automated analysis pipeline	(63)
GDSC	http://www.cancerrxgene.org	Drug sensitivity information; drug response information	(64)
canSAR	https://cansar.icr.ac.uk/	Multidisciplinary information; drug discovery	(65)
NONCODE	http://www.noncode.org/	ncRNAs; lncRNAs; up-to-date and comprehensive resource	(66)

There are many newly available resources in different areas where the extremely popular fields, cancer research and bioinformatics overlap. However, these areas may be unfamiliar to most cancer researchers aiming to use bioinformatics tools to analyze their data, and to bioinformaticians looking for real data to develop new bioinformatic tools and workflows and test algorithms (67).

III. Materials and Methods

In this section, a detailed description of the methods used to perform the different experiments will be given starting from the cohort constitution to the DNA extraction, and WES until the interpretation of the results obtained.

A. Cohort constitution

We collected both formalin-fixed paraffin-embedded tissue samples and fresh tissue samples in addition to blood samples due to the lack of samples so we had to collect different types of tissue to obtain enough samples for our study (fresh and FFPE).

❖ Formalin-fixed paraffin-embedded tissue

Sixty-five formalin-fixed paraffin-embedded tissue samples were collected from the tumor regions in the prostate at the University Hospital Mohamed VI, Oujda in Morocco. This study was approved by the Ethics Committees of the University Hospital Hassan II.

❖ Fresh tissue

Fifteen fresh tissue and blood samples were collected from the tumor regions in the prostate at the University Hospital Mohamed VI, Oujda and from the CHU of Casablanca in Morocco. This study was approved by the Ethics Committees of the University Hospital Hassan II. The 15 samples of fresh-frozen radical prostatectomy specimens (and blood that was used as a

control) were collected in RNAlater tissue protect tubes (Qiagen). Tissue samples were stored at –80°C.

B. Clinical Characteristics

The following table presents the principal clinical and pathological characteristics of our twelve prostate cancer patients:

Table 5. 2: Clinical Characteristics of Prostate Cancer Patients

Sample	Age	Type of PCa	Gleason Scor	PSA ng/m
2T	70	Adenocarcinoma	7(3+4)	80
5T	68	Adenocarcinoma	7(3+4)	123
4T	75	Adenocarcinoma	8 (4+4)	400
6T	66	Adenocarcinoma	6(3+3)	23
8T	80	Adenocarcinoma	8 (4+4)	645
10_3_tumor	55	Adenocarcinoma	7(4+3)	100
11_6_tumor	70	Adenocarcinoma	8(4+4)	139
12_7_tumor	54	Adenocarcinoma	7 (4+3)	100
13_9_tumor	72	Adenocarcinoma	8(4+4)	129
14_10_tumor	64	Adenocarcinoma	6(3+3)	34
14_10_tumor	59	Adenocarcinoma	6(3+3)	100
15_17_tumor	74	Adenocarcinoma	9(4+5)	381
9_1_tumor	81	Adenocarcinoma	9(4+5)	584

C. DNA Extraction

C.1 Formalin-fixed paraffin-embedded tissue

Formalin-fixed paraffin-embedded (FFPE) prostate tissue blocks were sliced into two slides per sample by the Pathology Core and GW Biorepository at George Washington University, Washington DC, USA. Then, they were examined to select the high-density cancer region from noncancerous material to ensure high purity of cancer DNA. Two different deparaffinization methods were used. The DNA was isolated using TRIzol, which is a complete reagent for the isolation of high-quality total RNA, DNA, and protein simultaneously.

- ***Deparaffinization***

Deparaffinization of the FFPE slides was performed using an optimized protocol from RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE (ThermoFisher Scientific). → the

curve and the concentration of the DNA was not good.

Then, deparaffinization of the FFPE slides was performed using a standard xylene and ethanol procedure prior to macrodissection of the prostate tumor and adjacent histologically normal tissue (ADJNT). Total DNA was isolated and extracted by TRIzol (Invitrogen, Catalog # 15596018) (See detailed protocol in Chapitre4. RNA extraction section) following the manufacturer's instructions and using a powerful homogenizer to homogenize the tissue. The quantity and quality of FFPE DNA was measured using the Agilent 2100 Bioanalyzer system (RNA 6000 Pico Kit). → good yield of DNA and it is the one that has been used in the following steps.

C.2 Fresh tissue

The 15 samples of fresh-frozen radical prostatectomy specimens were collected in RNAlater tissue protect tubes (Qiagen). Tissue samples were stored at –80°C.

The fresh tissue was homogenized using a clean homogenizer probe (Tissue Treator) by washing for 1 min in RNase, 1min 70% ethanol, and 1 min in molecular grade water at speed 3 (in 15 ml conicals). DNA was extracted using Mag Max DNA/RNA ultra-kit (Applied Biosystems by Life Technologies) (Appendix 4). The quantity and quality of DNA were measured using Qubit Flurometer.

D. Exome library preparation

The DNA concentration was measured by Qubit™ 1X dsDNA-HS assay kit (Invitrogen, cat # Q32231). Due to the limited amount of DNA for few samples 27-200 ng of DNA were sheared in 20 ul of 1X Low TE buffer by Covaris Instrument (E220 PLUS) to ~ 175 bp. Shearing was performed in an 8 AFA-TUBE TPX Strip (Covaris p/n 520292) with the following parameters: duty factor 25%, Peak Incident power 220, cycle burst 50, time-200 sec at 4°C.

The Agilent SureSelect XT Target Enrichment System for Illumina paired-end sequencing library protocol was used for library preparation (Agilent Technologies, Santa Clara, CA). SureSelect XT Human All Exon V7 Exon capture kit # G9507N was used. DNA Lo Bind Tubes, 1.5-mL PCR clean (Eppendorf # 022431021) or 96 well plates were used to process the samples.

Agilent SureSelectXT-Low input library prep ILM Kit (Agilent # G9507N) was used to prepare

the library for each sheared DNA sample. DNA fragment's ends were repaired and adenylated at the 3' end and then indexed with individual Indexes and adaptor ligated. These libraries were then amplified (pre-capture PCR amplification, 98°C for 2 minutes, 8-11 cycles (for 100-200 ng starting material and 11 cycles for 27-50 ng starting material)– 98°C for 30 seconds, 60°C for 30 seconds, 72°C for 1 minute then 72°C for 5 minutes, 4°C hold) by Herculase II fusion enzyme provided with the Agilent library preparation kit. Adaptor ligated amplified libraries were purified by Ampure XP beads (Beckmann Coulter Genomics # A63882).

Samples were analyzed by Tapestation DNA-1000 screen tape (Agilent # 5067-5582) to check the size of the libraries (Agilent 4200 TapeStation # G2991 AA). The concentration of each library was determined by integrating the 200-500 bp area.

Then each library (~500- 1000 ng) was hybridized with biotinylated Human All Exon V7 capture RNA baits (catalog # G9507N, 96 reactions) in the presence of blocking oligonucleotides. Hybridization was done for 2 hours: 95°C for 5 minutes, 65°C for 10 minutes, 65°C for 1 minute, 60 cycles– 65°C for 1 minute, 37°C for 3 sec then 65°C hold. V7 baits were added to the libraries after pausing the thermal cycler after 10 minutes at 65°C incubation. Capture and post capture steps were performed by automation on an Agilent-BRAVO Liquid handler (Model: G5563A). Bait-target hybrids were captured by streptavidin-coated magnetic beads (Dynabeads MyOne Streptavidin T1, Life Technologies, # 6560) for 30 minutes at room temperature. Then, after a series of washes to remove the non-specifically bound DNA, captured library was eluted in nuclease free water and amplified (post-capture PCR amplification: 98°C for 2 minutes, 9 cycles– 98°C for 30 seconds, 60°C for 30 seconds, 72°C for 1 minute then, 72°C for 5 minutes, 4°C hold). Tapestation High sensitivity DNA screen Tape (Agilent # 5067-5587) was used to validate the size of the libraries.

Finally, 40 ng of DNA per sample was used for sequencing.

E. Exome Sequencing

Exome sequencing is the process to determine the nucleotide sequence primarily of the exonic (or protein-coding) regions of an individual's genome and related sequences, representing approximately 1% of the complete DNA sequence. but contains ~85% of known disease-related variants, (4–5 Gb of sequencing per exome compared to ~90 Gb per whole human genome). There are 3 steps in WES: library preparation, sequencing and data analysis.

The ready to sequence DNA libraries were pooled equimolar and then quantified by qPCR using KAPA Library Quantification Kit from Roche, before cluster generation. DNA was then subjected to exonic hybrid capture using SureSelect v7 Exome bait (Agilent) with tumor samples at 150 x coverage, and normal samples at 50x coverage before sequencing on an Illumina NovaSeq 6000.

F. Data Processing

The samples (4T/4N,6T/6N,8T/8N) have 210 to 293 million pass filter reads with more than 91% of bases above the quality score of Q30. The samples (2N/2T & 5T/5N) have 171 M to 187 M pass filter reads, with Q30 above 93%. The fresh tissue samples had 109 M to 200 M pass filter reads, with Q30 above 90%. Reads of all the samples were trimmed for adapters and low-quality bases using Cutadapt (68) before alignment with the reference genome Human – hg38 using BWA mapping software. Total alignment with the reference genome (Human – hg38) was 99% for all the samples and uniquely mapped reads were between 77% and 99.95%. Picard's MarkDuplicate utility was used to measure library complexity in terms of unique fragments with the mapped reads. The samples had 68-91% non-duplicate reads. The capture efficiency is measured by the percentage of the reads mapped to the target region or overlap with the target regions. There are between 83-88% reads mapped to the target. Coverage statistics were also measured using Dragen (Dynamic Read Analysis for GENomics) Bio-IT Platform. Raw sequencing depth coverage over the target region was between 396x and 570x and more than 97% of the target regions had the coverage above 20x to 30x or more.

The WES data basically FASTQ files underwent pre-processing, which includes demultiplexing, quality control and adapter trimming along with alignment before being processed using GATK best practices (69) which is the more standard variant caller. In the following step, Genome Analysis Toolkit (LeftAlignIndels, RealignerTargetCreator, IndelRealigner, BaseRecalibrator) was used to improve alignment accuracy.

G. Variant Calling

GATK HaplotypeCaller was used to call germline SNPs and indels via local re-assembly of haplotypes. Using the following references: hg38.fa; Homo_sapiens_assembly38.known_indels.vcf.gz; dbsnp_146.hg38.vcf.gz. Somatic single

nucleotide variants (SNVs) and small insertions and deletions for each pair were identified using MuTect2 (v2.9.0) (70) with default parameters. CNV was performed with cn.mops (Copy Number estimation by a Mixture Of PoissonS), which is a data processing pipeline for copy number variations and aberrations (CNVs and CNAs) from next generation sequencing (NGS) data. More CNV changes were detected in the Fresh tissue samples. Annovar/latest.(71) was used for variant annotation and vcf files were filtered (exclude variant allele read depth >=3) and converted using the vcf2maf tool (https://github.com/mskcc/vcf2maf). Mutation tables were imported into RStudio (V4.2.1) (https://www.R-project.org) for visualization and analysis using the Maftools (v2.4.10) (72) package (Figure 5.5).

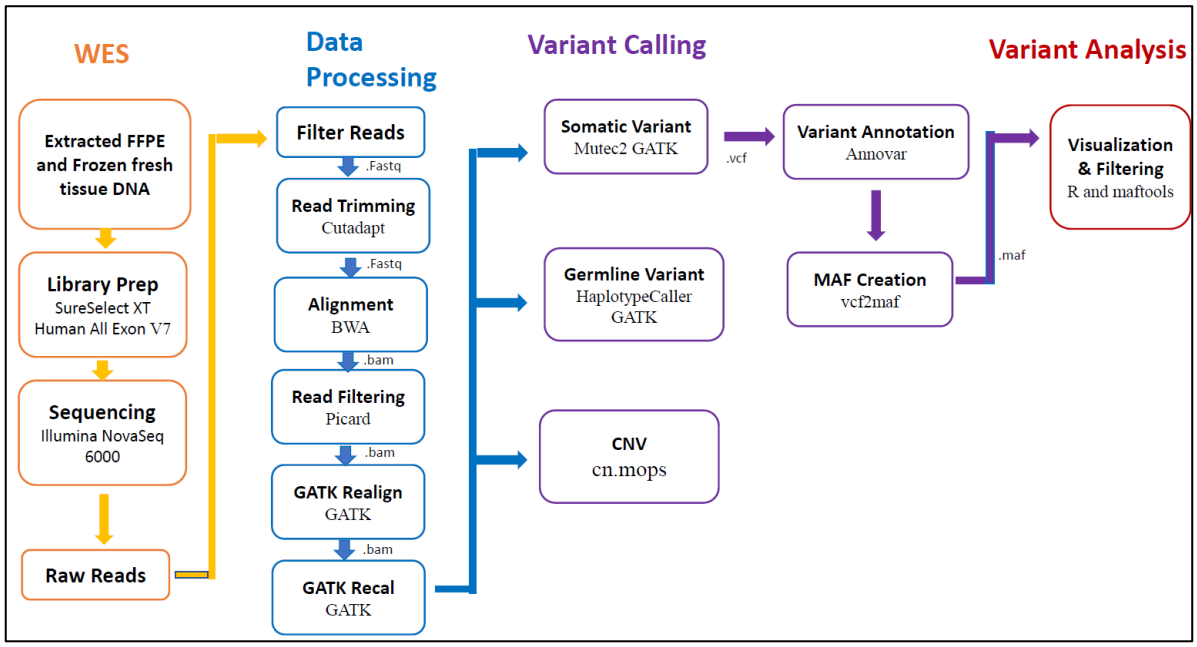


Figure 5. 5: WES Analysis Workflow. Moroccan PCa DNA samples were processed and analyzed using the GATK WES pipeline in tumor-normal somatic variant and germline modes. Output mutation tables (maf) and variant calls (vcf) were used for all downstream analyses.

IV. Results

To determine the somatic and germline variants and CNVs associated with prostate cancer in Moroccan PCa men, we collected samples from two hospital university institutions. Specifically, we collected 12 intermediate (Gleason scores 4 + 3) and high-grade (Gleason scores ≥ 4 + 3) tumors and 12 nontumor prostate samples and 12 blood samples. A total of 18 (9 tumor and 9 nontumor) samples were collected from Northeastern Morocco), 6 samples were collected from Central-western Morocco.

Somatic variant analysis of matched tumor and normal sequencing data lead to the

identification of genes known to be associated with prostate cancer, and novel prostate cancer genes. We identified 639 novel variants that were not found in either of the following databases: COSMIC (73), dbSNP (74) and ClinVar (75).

❖ *Mutational Landscape of PCa*

We analyzed the variant data using "Maftools". Figure 5.6 and Figure 5.7 illustrate the overview of the somatic variants found in the 12 PCa patients and describes the general features of the mutations across the PCa patients. The identified variants were comprised of 150,843 missense mutations, 6825 nonsense mutations, 10206 insertions, 10783 deletions, 1057 translation start site variants, 10962 splice site variants, and 1027 nonstop mutations, the most abundant is missense mutations (Figure 5.6A). The single nucleotide variant (SNV) is the most abundant type (Figure 5.6B). The main single nucleotide variant in PCa samples was T > C (35.29%) and C > T (33.52%) (Figure 5.6C).

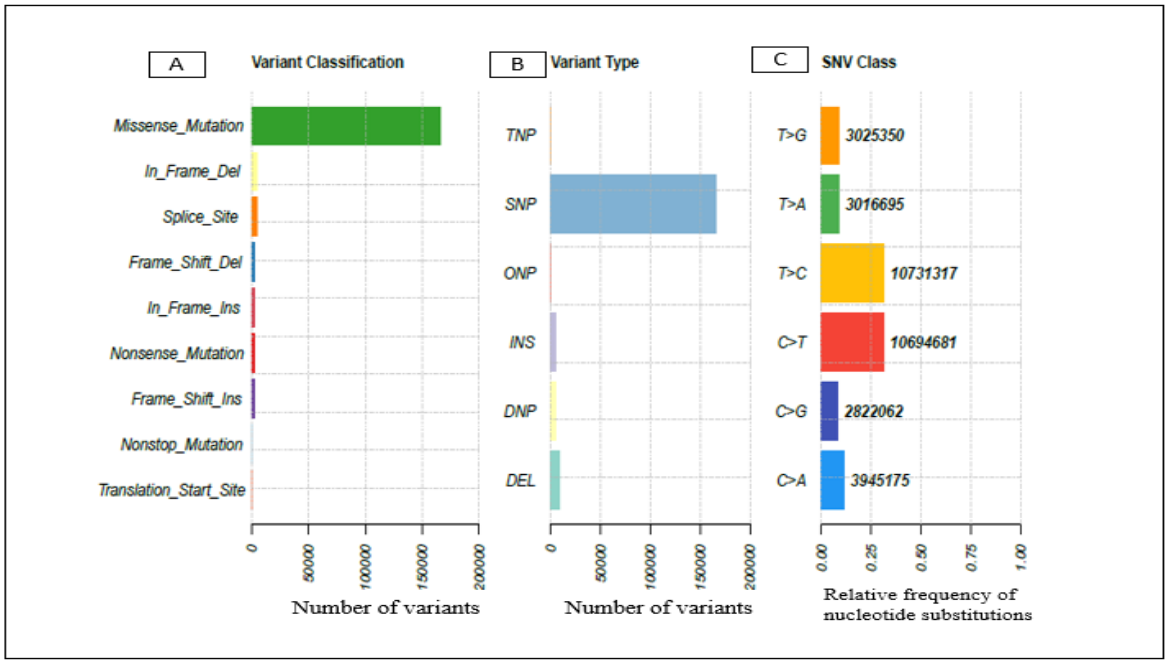


Figure 5. 6: Mutational spectrum of the prostate cancer samples. (A): Barplot of the variant classification, by using the variant effect predictor (VEP) Missense mutations are the most common form of mutations. (B) Distribution of SNV classes, the total number of mutation sites are respectively belonging to SNP (substitution), INS (insertion), DEL (deletion), Single nucleotide polymorphism (SNP) is the most common genetic variant type (C): Distribution of six basic base substitutions, among single nucleotide variation (SNV) transition, C > T and T > C are the most common.

There were significant differences among different variant types (Figure 5.7 A). The barplot in

figure 5.7B displayed the top ten mutated genes in our PCa patient. Each bar represents the number of mutations (on x-axis) for each gene and the stacked color indicates variant classification. The number on top of the bar represents the percentage of samples with mutation(s) in that particular gene.

Within our cohort, almost 363 somatic variants were present in 10 mutated genes. One prostate cancer–associated gene *NBPF1* (100%) and other three cancer-associated genes *CROCC, AK2, and CFAP74* were identified, which may be a new prostate cancer-associated gene (Figure 5.7B).

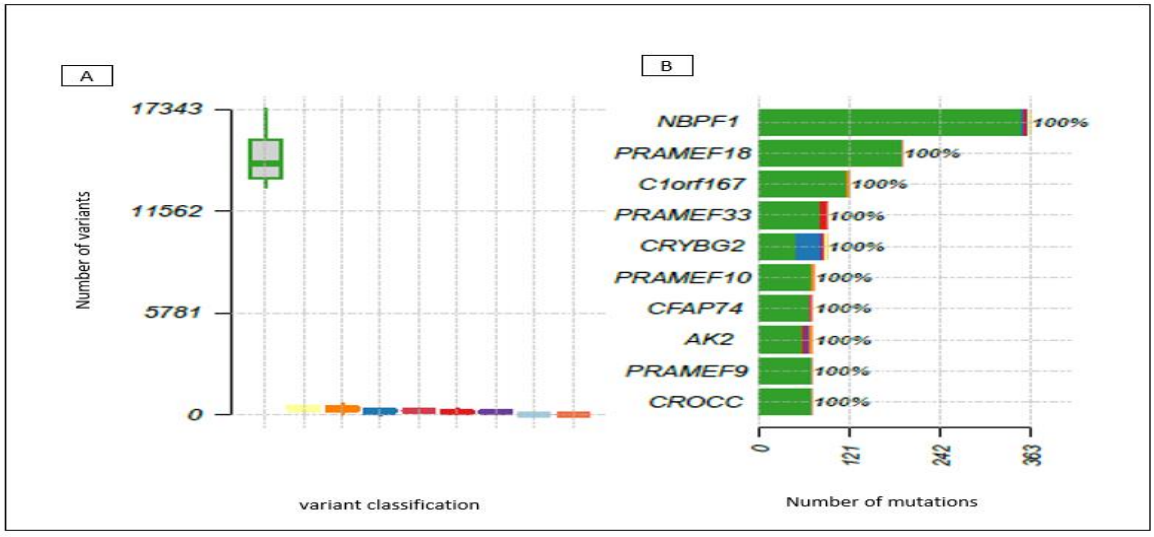


Figure 5. 7: Overview of the somatic mutations in 12 PCa patients. (A): Boxplot summarizing variant types by Variant Classification. (B): Distribution of top 10 mutated genes across our cohort. Each bar represents the number of mutations (on x-axis) for each gene and the stacked color indicates variant classification. The number on top of the bar represents the percentage of samples with mutation(s) in that particular gene.

The oncoplot in Figure 5.8 represents the top 20 mutated genes across the 12 samples. The oncoplot is a combination of gene-level summary, sample level summary, and the variant level summary. We identified four prostate cancer-associated genes: *NBPF1, FAM13C, HSPG2, and MYOM3*. Therefore, 6/20 of the top frequently mutated genes aside from the four-prostate cancer-associated genes in our PCa cohort were previously reported to be implicated in cancer, which are: *LUZP1, TAS1R2, CROCC, AK2, CFAP74, and UBXN11* genes, and maybe a new potential prostate cancer-associated genes.

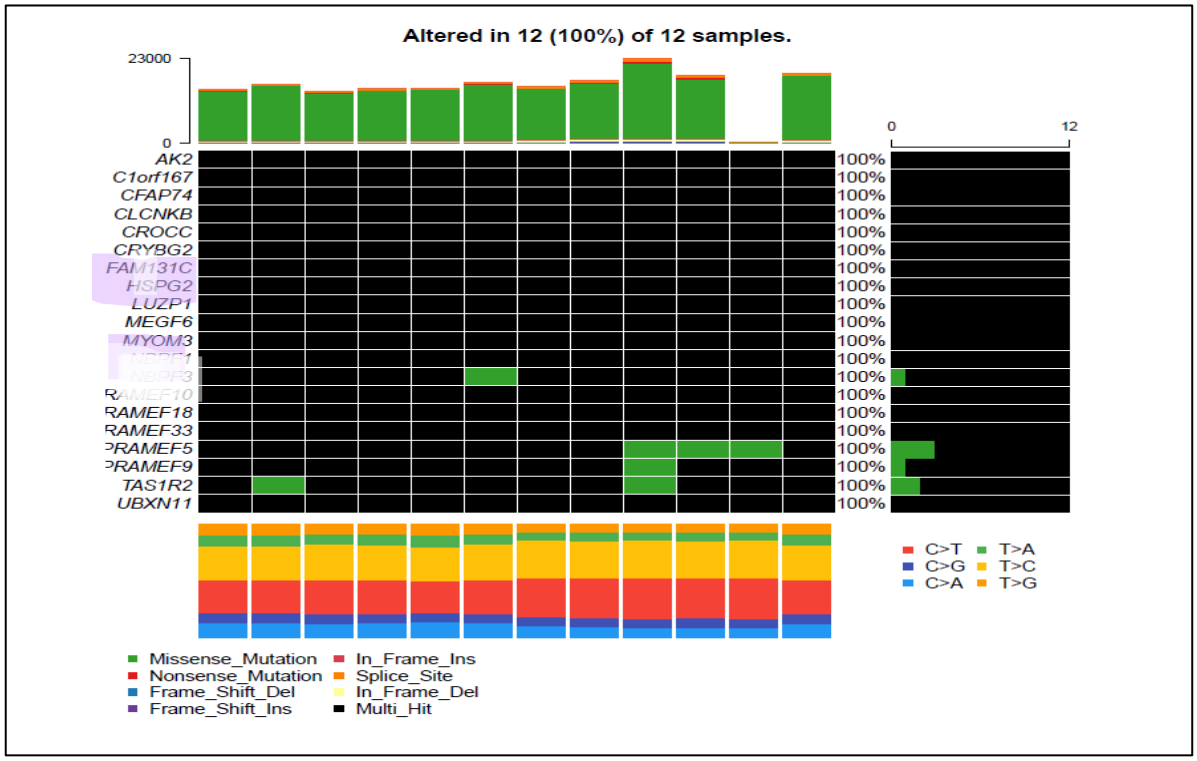


Figure 5. 8 Oncoplot showing the heatmap of variant classification for the top 20 genes across 12 PCa patients. The X-axis shows different samples and the Y-axis shows the top 20 genes. Genes are ordered by their mutation frequency. Several genes harbor mutations across all tumor samples showing their importance in oncogenesis. Variants annotated as Multi_Hit are those genes which are mutated more than once in the same sample.

- **Signature Pathways across PCa Samples**

The pathway analysis revealed some signature pathways in the PCa patients in this research study. Among the top affected pathways RTK-RAS and NOTCH were affected in all the PCa samples. Furthermore, WNT, PI3K, Hippo, MYC, Cell-cycle and TP53 were affected across a substantial number of PCa samples (Figure 5.9). All the affected pathways are already known for their involvement in PCa. Therefore, the most affected pathways discovered in our PCa cohort were in agreement with previous findings.

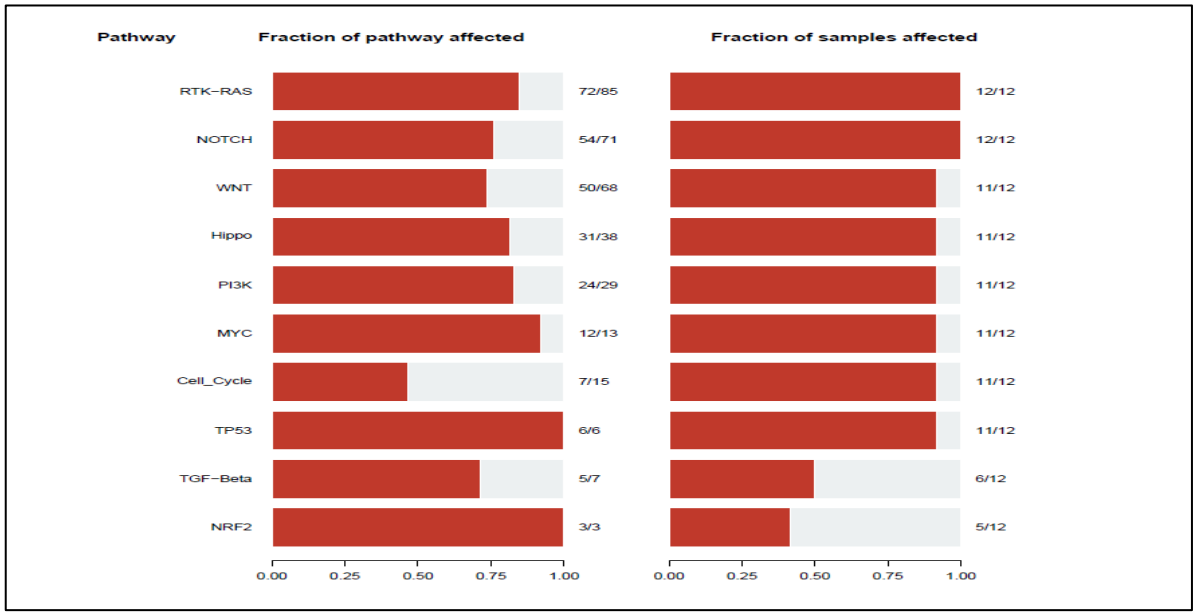


Figure 5. 9: Most affected pathways across the 12 PCa samples. checked for enrichment of known Oncogenic Signaling Pathways in TCGA cohorts. Affected pathways: the components of the signalling pathways that have been affected by the oncogenic mutations which caused a dysregulation of these signaling pathways that that control cell growth and division and cell death.

V. Discussion

In this study, we used whole exome sequencing to identify genomic features of prostate cancer in Moroccan men. This study is the first to perform WES to identify genomic features of prostate cancer in Moroccan men and one of the few attempts in the African continent using WES to analyze PCa. We analyzed the exomes of 12 Moroccan prostate cancer patients. Herein, we report the identification of three known prostate cancer-associated genes: *NBPF1*, *FAM13C*, *and HSPG2* (76–78), and six novel potential prostate cancer–associated genes : *LUZP1*, *TAS1R2, CROCC, AK2, CFAP74, and UBXN11*, that were initially associated with other types of cancers (79–83).

Neuroblastoma breakpoint family, member 1 (*NBPF1*) has been shown to have various traits as tumor suppressor for prostate cancer, neuroblastoma, cutaneous squamous cell carcinoma, and cervical cancer. The same gene has other functions as oncogene in other cancers (76). Several studies, demonstrated that expression of *HSPG2* (Perlecan) gene has a significant role in PCa by regulating this cancer cell proliferation (77). Moreover, the expression of *FAM13C* (Family with sequence similarity 13, Member C) - a gene of largely unknown function- has been recently found to be associated with *TMPRSS2-ERG* fusions. This gene is a strong and independent prognostic marker in prostate cancer and proposed for routine prognostic test in

PCa (78).

From Laura et al. (2021) (79) analysis, The Leucine Zipper Protein 1, LUZP1 emerges as a potential cancer-relevant gene, exhibiting genomic and transcriptional perturbation pattern suggestive of a tumor suppressive function (79). Furthermore, TAS1R2 is one of the three isoforms of *TAS1R* genes, which found to be expressed in some solid tumors where they regulate apoptosis and proliferation and could serve as biomarkers for cancer outcomes (80).

Our analysis identified several well-known signaling pathways that are involved in prostate cancer progression and tumorigenesis and are clinically targetable Including: RTK-RAS, NOTCH, WNT, PI3K and MYC pathways. It is worth noting that down-regulation of the *NOTCH* signaling pathway-which is one of the highest-ranking pathways in PCa- promotes tumorigenesis, invasion, and tumor progression by triggering the AKT, FoxM1, and other targets (84,85). However, there is a need for further analysis on *NOTCH* family genes (*NOTCH*s) as potential therapeutic targets and prognostic biomarkers for PCa (86). Furthermore, Wnt/β-catenin pathway is extremely important in the prostate tumor development and comprises of secreted glycoproteins that play a vital role in tissue homeostasis, cell proliferation, differentiation, and migration, and also activate WNT signaling in tumor cells and promote drug resistance and PCa progression (87,88). Although, another key molecular pathway that promotes PCa development and progression is the RTK-RAS or (Ras/Mitogen-Activated Protein Kinase) RAS/MAPK pathway. This pathway is one of the major branches of RTK signaling pathway, which is overactivated in PCa, display activating genetic alterations in more than 40% of primary tumors and in almost all metastatic prostate tumors (89). Likewise, the role of phosphoinositide-3-kinase pathway (PI3K) is recognized in the PCa tumorigenesis, which has led to the development of some PI3K inhibitor drugs which are still in clinical trials such as AZD8186 and GSK2636771. Moreover, the dysregulation of this pathway has been found in up to 70%-100% of advanced prostate cancer cases (90).

Another critical signaling pathway, which its overexpression leads to prostate cancer growth and can mediate resistance of PCa cells to chemotherapy and radiotherapy is the MYC signaling pathway. Recently, this signaling remains as a new emerging target in cancer therapy and can function as independent prognostic and diagnostic factor in PC patients (91). Prominently, the upregulation of Hippo signaling pathway downstream effectors, Yes-associated protein (YAP)/ transcriptional coactivator with PDZ-binding motif (TAZ), is well identified in variety of solid tumors including PCa. This pathway consists of an upstream serine-threonine kinase cascade,

which activates various tumor suppressor (92).

These findings represent the first and most comprehensive characterization of the Moroccan PCa exome to date. Because North African are underrepresented in genomic studies, our findings address a gap in the contribution of genetic variation to the incidence of prostate cancer and aggressiveness of the disease in the Moroccan and all North African population.

Prostate cancer sequencing studies have been comprised primarily of men of European ancestry, and there have been few large studies focused on men of African ancestry and especially afro-native populations (93). Although, Emerging evidence indicates that there are remarkable disparities in the prostate cancer molecular landscape among different population groups of European, North American, African, and Asian origin (2,94). It is known that African populations carry a larger number of rare variants than European populations. One aspect of rare variant discovery within individuals of African ancestry that warrants discussion is the potential effect of fine-scale population stratification (95,96). This indicates the need for further genomic studies to be conducted in the afro-native populations.

Huang et al. (2017) conducted a WES study on 102 African American prostate cancer patients. This study identified new insights into prostate cancer, including the identification of *ERF* as a prostate cancer gene, somatic copy-number alteration differences, and uncommon *PIK3CA* and *PTEN* alterations. This highlighted the importance of the inclusion of underrepresented minorities in cancer sequencing studies (93).

A recent study examining 82 risk variants in ∼4800 prostate cancer cases and ∼4700 controls of African ancestry found 83% of variants to have directionally consistent effect estimates, suggesting that the majority of common loci harbor risk alleles were identified by GWAS (97). Multiple studies have demonstrated the difficulty of identifying the genes contributing to PCa as this disease risk has a significant genetic component. Large-scale genome-wide association studies have suggested as many as 170 distinct risk loci, with many contributing a very small amount of risk (98).

In a very recent study conducted by White et al. (2022) (99) on a cohort of 45 advanced-stage, treatment-naïve Nigerian (NG) primary prostate cancer tumors and 11 unmatched normal tissues. WES has been used to identify West African ancestry-linked variant patterns. Germline alterations were found in *BRCA1*, *BARD1*, *BRCA2*, and *PMS2* in at least two samples. *BARD1* and *BRCA2* germline variant patterns seem to be specific to NG men. The somatic

variant analysis identified 905 variants across 25 genes known to be associated with prostate cancer, and 156 variants across 51 novel prostate cancer genes (99).

Next-generation sequencing technologies enable personalized, genome-based medicine in practice. NGS methods have revolutionized laboratory diagnostics for several genetic disorders, including cancer. Diverse targeted NGS strategies can be applied for the analysis of DNA sequence and copy number variation, even in hundreds of genes, simultaneously and at reasonable costs (29). Therefore, the idea of incorporating prostate cancer genetics in the clinic is worth revisiting in the next few years.

There are few publications and studies on prostate cancer in Africa, although genetics is a major risk factor for this disease (100). Therefore, several genetic variants that contribute to elevated circulating androgens, including androgen reduced clearance and upregulated activity of androgen receptor, have been reported in peer-reviewed publications on prostate cancer genetics in African populations. These include polymorphisms in cytochrome P450 genes such as *CYP3A4*, *CYP3A5*, *CYP1A1*, *CYP17* in Morocco, Tunisia, Nigeria, South Africa, and Senegal (101–105). In North Africa, Ivory Coast, and Nigeria, other alterations in CAG and GGN repeats in the androgen receptor gene have been reported as risk factors (106,107).

Our study has several limitations, including the missing or incomplete health record of some patients and the non-ideal process of deparaffinization of the formalin fixed paraffin embedded (FFPE) biopsy blocks which affected the quality of the DNA. Also, the relatively small sample size may not properly reflect the whole population, especially because this is one of the most heterogeneous diseases. Our results align with other studies suggesting that inclusion of sufficient numbers of patients in cancer genomic studies may give a better overview of the genomic features of a population and enable the discovery of new cancer genes and inform the inclusion of diverse populations toward precision cancer medicine (108).

Furthermore, this study permitted us to identify several barriers to the implementation of WES/WGS in Morocco and probably on the whole continent. The first major hurdle to overcome when analyzing a set of exome sequencing data is the sheer number of variants that are present compared to the reference sequences. In a single exome sequence, researchers can expect to be confronted with anywhere from 20-30,000 variants. Of these, approximately 10,000 variants will be predicted to result in nonsynonymous amino acid substitutions, splice-site alterations, insertions, or deletions (109). In addition, there remains a huge deficit in computational skills and infrastructure required to carry out this kind of research studies in

African nations to alleviate this knowledge gap (110). This will allow us to have a proper understanding of genetics and genomics among African populations and lead to a step forward into precision medicine for prostate cancer.

In the future, new mechanisms of carcinogenesis that inform both disease modeling and patient stratification for clinical trials of experimental agents may see the day due to this expanded genetic framework may articulate. With advances in technology and studies on the genome, exome, epigenome and transcriptome of prostate cancer, genetic alterations that lead to the disease will be discovered, as well as therapeutic target areas.

VI. Conclusion

In conclusion, this is, to our knowledge, the first study of WES among Moroccan prostate cancer patients and among North African populations. Further WES studies on prostate cancer in Afro-native populations are needed to identify their genomic features and understand their effects and to confirm their association with prostate cancer risk. Unfortunately, there are still several scientific and technical limitations that make difficult the direct implementation of such a strategy.

In the coming years, with advances in NGS and low cost of WES and WGS, it will be feasible to quickly and accurately obtain and utilize the germline genetic profile of a man with PCa, particularly in Morocco, at the time of diagnosis to set up a personalized treatment pathway according to the risk category indicated by the SNPs in his germline and the presence or absence of common or rare variants. Using whole exome sequencing to predict aberrations in tumors is a cost-effective alternative to whole genome sequencing, however it is predominantly used for variant detection and infrequently utilized for the detection of somatic copy number variation. Moreover, technical and human resources are two relevant components to support precision medicine implementation in the health system.

References

1. Laitinen V. Genetic Risk Factors for Hereditary Prostate Cancer in Finland.
2. Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, et al. The long tail of oncogenic drivers in prostate cancer. Nat Genet. 2018;50(5):645–51.
3. Maggiore RJ, Curran EK, Witt ME, Haraf DJ, Everett E, Cohen EEW. Discovery and saturation analysis of cancer genes across 21 tumor types. Nature. 2015;4(4):327–33.
4. Hanna EA et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2017;536(7167):285–91.
5. Canzoneri R, Lacunza E, Abba MC. Genomics and bioinformatics as pillars of precision medicine in oncology. Medicina (B Aires). 2019;79(6–1):587–92.
6. CRICK FH. On protein synthesis. Symp Soc Exp Biol. 1958;12:138–63.
7. CRICK FH, BARNETT L, BRENNER S, WATTS-TOBIN RJ. General nature of the genetic code for proteins. Nature. 1961 Dec;192:1227–32.
8. Thieffry D, Sarkar S. Forty years under the central dogma. Trends Biochem Sci. 1998 Aug;23(8):312–6.
9. Matsubara K, Okubo K. Identification of new genes by systematic analysis of cDNAs and database construction. Curr Opin Biotechnol. 1993 Dec;4(6):672–7.
10. Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, Quackenbush J. Gene index analysis of the human genome estimates approximately 120,000 genes. Nat Genet. 2000;25(2):239–40.
11. Hanahan D, Weinberg RA. The Halmarks of Cancer. Cell. 2000;100:57–70.
12. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74.
13. Berger MF, Lawrence MS, Demichelis F, Cibulskis K, Sivachenko AY, Sboner A, et al. The genomic complexity of primary human prostate cancer Michael. Nature. 2011;470(7333):214–20.
14. Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D, Stern HM, et al. Diverse somatic mutation patterns and pathway alterations in human cancers. Nature. 2010;466(7308):869–73.
15. Grasso CS, Wu Y, Robinson DR, Cao X, Saravana M, Khan AP, et al. The Mutational Landscape of Lethal Castrate Resistant Prostate Cancer. Nature. 2013;487(7406):239–43.
16. Kim JH, Dhanasekaran SM, Mehra R, Tomlins SA, Gu W, Yu J, et al. Integrative analysis of genomic aberrations associated with prostate cancer progression. Cancer Res. 2007;67(17):8229–39.
17. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, et al. A common variant associated with prostate cancer in European and African populations. Nat Genet. 2006;38(6):652–8.
18. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014;42(D1):1001–6.
19. Kuiper RP, Ligtenberg MJL, Hoogerbrugge N, Geurts van Kessel A. Germline copy number variation and cancer risk. Curr Opin Genet Dev. 2010 Jun;20(3):282–9.
20. Krepischi ACV, Pearson PL, Rosenberg C. Germline copy number variations and cancer predisposition. Futur Oncol. 2012;8(4):441–50.
21. Liu W, Sun J, Li G, Zhu Y, Zhang S, Kim ST, et al. Association of a Germ-Line copy

number variation at 2p24.3 and risk for aggressive prostate cancer. Cancer Res. 2009;69(6):2176–9.
22. Jin G, Sun J, Liu W, Zhang Z, Chu LW, Kim ST, et al. Genome-wide copy-number variation analysis identifies common genetic variants at 20p13 associated with aggressiveness of prostate cancer. Carcinogenesis. 2011;32(7):1057–62.
23. Williams JL, Greer PA, Squire JA. Recurrent copy number alterations in prostate cancer: an in silico meta-analysis of publicly available genomic data. Cancer Genet. 2014;207(10–12):474–88.
24. Witte JS. Prostate cancer genomics: Towards a New Understanding. Nat Rev Genet. 2009;10(2):90–3.
25. Xu J, Yang P, Xue S, Sharma B, Sanchez-Martin M, Wang F, et al. Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. Hum Genet. 2019;138(2):109–24.
26. Chari R, Coe BP, Vucic EA, Lockwood WW, Lam WL. An integrative multi-dimensional genetic and epigenetic strategy to identify aberrant genes and pathways in cancer. BMC Syst Biol. 2010;4:1–14.
27. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods. 2014 Mar;11(3):333–7.
28. Rizzo JM, Buck MJ. Key principles and clinical applications of "next-generation" DNA sequencing. Cancer Prev Res. 2012;5(7):887–900.
29. Luthra R, Chen H, Roy-Chowdhuri S, Singh RR. Next-generation sequencing in clinical molecular diagnostics of cancer: Advantages and challenges. Cancers (Basel). 2015;7(4):2023–36.
30. Lohmann K, Klein C. Next Generation Sequencing and the Future of Genetic Diagnosis. Neurotherapeutics. 2014;11(4):699–707.
31. Barbieri CE, Bangma CH, Bjartell A, Catto JWF, Culig Z, Grönberg H, et al. The mutational landscape of prostate cancer. Eur Urol. 2013;64(4):567–76.
32. Ng SB, Turner EH, Robertson PD, Flygare SD, Abigail W, Lee C, et al. Targeted Capture and Massicely Parallel Sequencing of twelve human exomes. Nature. 2010;461(7261):272–6.
33. Spans L, Clinckemalie L, Helsen C, Vanderschueren D, Boonen S, Lerut E, et al. The genomic landscape of prostate cancer. Int J Mol Sci. 2013;14(6):10822–51.
34. Fitzgerald LM, Kumar A, Boyle EA, Zhang Y, Laura M, Kolb S, et al. Germline Missense Variants in the BTNL2 Gene Are Associated with Prostate Cancer Susceptibility. 2014;1024:1–15.
35. Roach JC, Glusman G, Smit AFA, Huff CD, Shannon PT, Rowen L, et al. Analysis of Genetic Inheritance in a Family Quartet by Whole Genome Sequencing. Science (80-). 2011;328(5978):636–9.
36. Sun Y, Ruivenkamp CAL, Hoffer MJV, Vrijenhoek T, Kriek M, van Asperen CJ, et al. Next-Generation Diagnostics: Gene Panel, Exome, or Whole Genome? Hum Mutat. 2015;36(6):648–55.
37. Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. J Hum Genet. 2014 Jan;59(1):5–15.
38. He KY, Ge D, He MM. Big data analytics for genomic medicine. Int J Mol Sci. 2017;18(2):1–18.
39. Clément L, Emeric D, J GB, Laurent M, David L, Eivind H, et al. A data-supported

history of bioinformatics tools. 2018;1–14. Available from: http://arxiv.org/abs/1807.06808
40. Paulina K et al. The growing role of precision and personalized medicine for cancer treatment. technology. 2019;6(3–4):79–100.
41. Valencia A, Hidalgo M. Getting personalized cancer genome analysis into the clinic: the challenges in bioinformatics. Genome Med. 2012;4(8).
42. Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. Annu Rev Med. 2012;63:35–61.
43. Mardis ER. A decade's perspective on DNA sequencing technology. Nature. 2011;470(7333):198–203.
44. Brownstein CA, Beggs AH, Homer N, Merriman B, Yu TW, Flannery KC, et al. An international effort towards developing standards for best practices in analysis, interpretation and reporting of clinical genome sequencing results in the CLARITY Challenge. Genome Biol. 2014;15(3).
45. Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 2009;38(6):1767–71.
46. Depristo MA, Banks E, Poplin RE, Garimella K V, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next- generation DNA sequencing data. Nat Genet. 2011;43(5):491–8.
47. Schbath S, Martin V, Zytnicki M, Fayolle J, Loux V, Gibrat JF. Mapping reads on a genomic sequence: An algorithmic overview and a practical comparative analysis. J Comput Biol. 2012;19(6):796–813.
48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
49. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet [Internet]. 2011;12(6):443–51. Available from: http://dx.doi.org/10.1038/nrg2986
50. Quintáns B, Ordóñez-Ugalde A, Cacheiro P, Carracedo A, Sobrido MJ. Medical genomics: The intricate path from genetic variant identification to clinical interpretation. Appl Transl Genomics [Internet]. 2014;3(3):60–7. Available from: http://dx.doi.org/10.1016/j.atg.2014.06.001
51. Jung J, Hwang Y, Ahn H, Lee S, Yoo S. Precise characterization of genetic interactions in cancer via molecular network refining processes. Int J Mol Sci. 2021;22(20).
52. Brito C, Costa-silva B, Barral DC, Pojo M. Unraveling the relevance of arl gtpases in cutaneous melanoma prognosis through integrated bioinformatics analysis. Int J Mol Sci. 2021;22(17).
53. Yang Y, Dong X, Xie B, Ding N, Chen J, Li Y, et al. Databases and web tools for cancer genomics study. Genomics, Proteomics Bioinforma [Internet]. 2015;13(1):46–50. Available from: http://dx.doi.org/10.1016/j.gpb.2015.01.005
54. Wilks C, Cline MS, Weiler E, Diehkans M, Craft B, Martin C, et al. The Cancer Genomics Hub (CGHub): overcoming cancer through the power of torrential data. Database (Oxford). 2014;2014:1–10.
55. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, et al. COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015;43(D1):D805–11.
56. Marx V. Drilling into big cancer-genome data. Nat Methods. 2013;10(4):293–7.

57. Packer BR, Yeager M, Burdett L, Welch R, Beerman M, Qi L, et al. SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. Nucleic Acids Res. 2006;34(Database issue):617–21.
58. Samur MK, Yan Z, Wang X, Cao Q, Munshi NC, Li C, et al. canEvolve: A Web Portal for Integrative Oncogenomics. PLoS One. 2013;8(2):1–10.
59. He X, Chang S, Zhang J, Zhao Q, Xiang H, Kusonmano K, et al. MethyCancer: The database of human DNA methylation and cancer. Nucleic Acids Res. 2008;36(SUPPL. 1).
60. Bhattacharya A, Ziebarth JD, Cui Y. SomamiR: A database for somatic mutations impacting microRNA function in cancer. Nucleic Acids Res. 2013;41(D1):977–82.
61. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal Complementary Data Sources and Analysis Options. Sci Signal [Internet]. 2014;6(269):1–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23550210%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4160307
62. Goldman M, Craft B, Swatloski T, Ellrott K, Cline M, Diekhans M, et al. The UCSC cancer genomics browser: Update 2013. Nucleic Acids Res. 2013;41(D1):949–54.
63. Zhang J, Finney RP, Rowe W, Edmonson M, Sei HY, Dracheva T, et al. Systematic analysis of genetic alterations in tumors using Cancer Genome WorkBench (CGWB). Genome Res. 2007;17(7):1111–7.
64. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2013;41(D1):955–61.
65. Bulusu KC, Tym JE, Coker EA, Schierz AC, Al-Lazikani B. CanSAR: Updated cancer research and drug discovery knowledgebase. Nucleic Acids Res. 2014;42(D1):1040–7.
66. He S, Liu C, Skogerbø G, Zhao H, Wang J, Liu T, et al. NONCODE v2.0: Decoding the non-coding. Nucleic Acids Res. 2008;36(SUPPL. 1):2007–9.
67. Kihara D, Yang YD, Hawkins T. Bioinformatics resources for cancer research with an emphasis on gene function and structure prediction tools. Cancer Inform [Internet]. 2006;2:25–35. Available from: https://doi.org/10.1177/117693510600200020
68. Marcel M. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. EMBnet,journal. 2013;7(10):2803–9.
69. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Current Protocols in Bioinformatics. 2013. 1–33 p.
70. Sougnez C, Gabriel S, Meyerson M, Lander ES. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31(3):213–9.
71. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):1–7.
72. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28(11):1747–56.
73. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: The

Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47(D1):D941–7.
74. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. DbSNP: The NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308–11.
75. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: Improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062–7.
76. Li L, Chen S, Tang Y, Wu J, He Y, Qiu L. Oncogene or tumor suppressor gene: An integrated pan-cancer analysis of NBPF1. Front Endocrinol (Lausanne). 2022;13(August):1–17.
77. Datta MW, Hernandez AM, Schlicht MJ, Kahler AJ, DeGueme AM, Dhir R, et al. Perlecan, a candidate gene for the CAPB locus, regulates prostate cancer cell growth via the Sonic Hedgehog pathway. Mol Cancer. 2006;5:1–15.
78. Burdelski C, Borcherding L, Kluth M, Hube-Magg C, Melling N, Simon R, et al. Family with sequence similarity 13C (FAM13C) overexpression is an independent prognostic marker in prostate cancer. Oncotarget. 2017;8(19):31494–508.
79. Bozal-Basterra L, Gonzalez-Santamarta M, Muratore V, Martín-Martín N, Ercilla A, Rodríguez JA, et al. LUZP1 Controls Cell Division, Migration and Invasion Through Regulation of the Actin Cytoskeleton. Front Cell Dev Biol. 2021;9(April).
80. Carey RM, Kim TB, Cohen NA, Lee RJ, Nead KT. Impact of sweet, umami, and bitter taste receptor (TAS1R and TAS2R) genomic and expression alterations in solid tumors on survival. Sci Rep [Internet]. 2022;12(1):1–14. Available from: https://doi.org/10.1038/s41598-022-12788-z
81. Kim H, Jeong M, Na DH, Ryu SH, Jeong E Il, Jung K, et al. AK2 is an AMP-sensing negative regulator of BRAF in tumorigenesis. Cell Death Dis. 2022;13(5):1–11.
82. Gao Y, Zhang X, Wang T, Zhang Y, Wang Q, Hu Y. Hnrnpcl1, pramef1, cfap74, and dffb: Common potential biomarkers for sporadic and suspected lynch syndrome endometrial cancer. Cancer Manag Res. 2020;12:11231–41.
83. Wang L, Sun X, Jin C, Fan Y, Xue F. Identification of Tumor Microenvironment-Related Prognostic Biomarkers for Ovarian Serous Cancer 3-Year Mortality Using Targeted Maximum Likelihood Estimation: A TCGA Data Mining Study. Front Genet. 2021;12(June):1–11.
84. Li JL, Sainson RCA, Shi W, Leek R, Harrington LS, Preusser M, et al. Delta-like 4 Notch ligand regulates tumor angiogenesis, improves tumor vascular function, and promotes tumor growth in vivo. Cancer Res. 2007;67(23):11244–53.
85. Wang X De, Shou J, Wong P, French DM, Gao WQ. Notch1-expressing cells are indispensable for prostatic branching morphogenesis during development and re-growth following castration and androgen replacement. J Biol Chem [Internet]. 2004;279(23):24733–44. Available from: http://dx.doi.org/10.1074/jbc.M401602200
86. Qing Y, Wang Y, Hu C, Zhang H, Huang Y, Zhang Z, et al. Evaluation of NOTCH family genes' expression and prognostic value in prostate cancer. Transl Androl Urol. 2022;11(5):627–42.
87. Kühl SJ, Kühl M. On the role of Wnt/β-catenin signaling in stem cells. Biochim Biophys Acta - Gen Subj [Internet]. 2013;1830(2):2297–306. Available from: http://dx.doi.org/10.1016/j.bbagen.2012.08.010
88. Murillo-Garzón V, Kypta R. WNT signalling in prostate cancer. Nat Rev Urol [Internet]. 2017;14(11):683–96. Available from:

http://dx.doi.org/10.1038/nrurol.2017.144
89. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative Genomic Profiling of Human Prostate Cancer. Cancer Cell. 2010;18(1):11–22.
90. Cham J, Venkateswaran AR, Bhangoo M. Targeting the PI3K-AKT-mTOR Pathway in Castration Resistant Prostate Cancer: A Review Article. Clin Genitourin Cancer [Internet]. 2021;19(6):563.e1-563.e7. Available from: https://doi.org/10.1016/j.clgc.2021.07.014
91. Faskhoudi MA, Molaei P, Sadrkhanloo M, Orouei S, Hashemi M, Bokaie S, et al. Molecular landscape of c-Myc signaling in prostate cancer: A roadmap to clinical translation. Pathol Res Pract. 2022 May;233:153851.
92. Salem O, Hansen CG. The hippo pathway in prostate cancer. Cells. 2019;8(4):1–24.
93. Huang FW, Mosquera JM, Garofalo A, Oh C, Baco M, Amin-Mansour A, et al. Exome sequencing of African-American prostate cancer reveals loss-of-function ERF mutations. Cancer Discov. 2017;7(9):973–83.
94. Lindquist KJ, Paris PL, Hoffmann TJ, Cardin NJ, Kazma R, Mefford JA, et al. Mutational landscape of aggressive prostate tumors in African American men. Cancer Res. 2016;76(7):1860–8.
95. Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, et al. Demographic history and rare allele sharing among human populations. Proc Natl Acad Sci U S A. 2011;108(29):11983–8.
96. Mathieson I, McVean G. Differential confounding of rare and common variants in spatially structured populations. Nat Genet. 2012;44(3):243–6.
97. Han Y, Signorello LB, Strom SS, Kittles RA, Rybicki BA, Stanford JL, et al. Generalizability of established prostate cancer risk variants in men of African ancestry. Int J Cancer. 2015;136(5):1210–7.
98. Benafif S, Kote-Jarai Z, Eeles RA. A review of prostate cancer Genome-Wide Association Studies (GWAS). Cancer Epidemiol Biomarkers Prev. 2018;27(8):845–57.
99. White JA, Kaninjing E, Adeniji KA, Jibrin P, Obafunwa JO, Ogo CN, et al. Abstract 1507: Tumor only analysis of whole exome sequencing from a multi-institutional Nigerian prostate cancer cohort reveals DNA repair genes associated with African ancestry. Cancer Res. 2022;82(12_Supplement):1507–1507.
100. Rotimi SO, Rotimi OA, Salhia B. A Review of Cancer Genetics and Genomics Studies in Africa. Front Oncol. 2021;10(February):1–24.
101. Novillo A, Romero-Lorca A, Gaibar M, Bahri R, Harich N, Sánchez-Cuenca D, et al. Genetic diversity of CYP3A4 and CYP3A5 polymorphisms in North African populations from Morocco and Tunisia. Int J Biol Markers. 2015;30(1):e148–51.
102. Fernandez P, Zeigler-Johnson CM, Spangler E, van der Merwe A, Jalloh M, Gueye SM, et al. Androgen Metabolism Gene Polymorphisms, Associations with Prostate Cancer Risk and Pathological Characteristics: A Comparative Analysis between South African and Senegalese Men. Prostate Cancer. 2012;2012:1–8.
103. Souiden Y, Mahdouani M, Chaieb K, Bakhrouf A, Mahdouani K. Lack of Association of *CYP1A1* Polymorphism with Prostate Cancer Susceptibility of Tunisian Men. Genet Test Mol Biomarkers [Internet]. 2012;16(7):661–6. Available from: http://online.liebertpub.com/doi/abs/10.1089/gtmb.2011.0212
104. Souiden Y, Mahdouani M, Chaieb K, Elkamel R, Mahdouani K. CYP17 gene polymorphism and prostate cancer susceptibility in a Tunisian population. Cancer

Epidemiol [Internet]. 2011;35(5):480–4. Available from: http://dx.doi.org/10.1016/j.canep.2010.11.008

105. Fernandez P, de Beer PD, van der Merwe LD, Heyns CF. Genetic variations in androgen metabolism genes and associations with prostate cancer in South African men. South African Med J. 2010;100(11):741–5.
106. Akinloye O, Gromoll J, Simoni M. Variation in CAG and GGN repeat lengths and CAG/GGN haplotype in androgen receptor gene polymorphism and prostate carcinoma in Nigerians. Br J Biomed Sci. 2011;68(3):138–42.
107. Esteban E, Rodon N, Via M, Gonzalez-Perez E, Santamaria J, Dugoujon JM, et al. Androgen receptor CAG and GGC polymorphisms in Mediterraneans: Repeat dynamics and population relationships. J Hum Genet. 2006;51(2):129–36.
108. Manrai AK, Funke BH, Rehm HL, Olesen MS, Maron BA, Szolovits P, et al. Genetic Misdiagnoses and the Potential for Health Disparities. N Engl J Med. 2016;375(7):655–65.
109. Robinson P, Krawitz P, Mundlos S. Strategies for exome and genome sequence data analysis in disease-gene discovery projects. Clin Genet. 2011;80(2):127–32.
110. Peprah E, Wiley K, Sampson U, Narula J. A New Age for African-Driven Genomics Research: Human Heredity and Health in Africa (H3Africa). Vol. 12, Global heart. England; 2017. p. 67–8.

GENERAL LIMITATIONS OF THESE STUDIES

The different parts of this thesis had several limitations. First, there was a small number of patients in each study, even though prostate cancer is the most diagnosed cancer after lung cancer among men in Morocco. This is due to cultural beliefs; this disease is still seen as taboo in Moroccan culture, and many men tend to deny it and refuse to take treatment even when they present at an advanced stage of this cancer. Another contributing factor may be the choice of treating PCa with herbal medicine rather than visiting a clinic or hospital, both because it is cost-effective and more efficient from their point of view. Delay in diagnosis is another reason that may explain the presentation of the patients with advanced disease. First, there is a lack of prostate cancer screening programs, and patients tend to present for medical care only when symptomatic. Furthermore, many patients, with poor socio-economic status, struggle to attend referral visits to the hospital, miss appointments and are lost to follow-up.

Second, it was very difficult to convince the patients to be part of this research study because for them, this is a taboo disease, and they struggle to receive the therapy, which is crucial so for them there is no way they would accept being part of any further research studies.

The third challenge was DNA isolation from FFPE tissue blocks. The process of paraffinization of the tissue blocks in Morocco is not ideal for a molecular study. There is a specific process that the anatomo-pathologist must follow to preserve DNA and RNA as much as possible for the molecular studies.

Fourth, we observed a deficiency of clinical data encountered in the regional hospital registries. There is no national registry for cancer in Morocco that provides an updated, detailed and complete data on the presentation of prostate cancer throughout the country. However, it stands as a call to arms in the management of prostate cancer in Morocco.

GENERAL CONCLUSIONS & PERSPECTIVES

This thesis was concerned mainly with identifying the genomic features of Moroccan men with prostate cancer (PCa) and investigating some biomarkers of this disease in the Moroccan population for a better diagnosis and prognosis.

This matter was addressed from three important angles, investigating *TMPRSS2-ERG* gene fusion as a common biomarker in prostate cancer, evaluating the presence/absence of *GSTP1* hypermethylation in Moroccan prostate cancer subjects which is a diagnostic and prognostic biomarker, and identifying novel variants and unique genetic features in Moroccan PCa patients by using whole-exome sequencing. In addition to an epidemiological study about this disease in Morocco.

Firstly, the epidemiological study was the first part we conducted to have an overview of the epidemiological status of this type of cancer in Morocco, which showed that Moroccan men presenting PCa tend to have intermediate and high-grade disease at diagnosis and delayed diagnosis. This highlights the need for further research and routine screening for early detection of this disease. Furthermore, there was a lack of information in the medical records of the patients and lack of an updated National Prostate Cancer Registry for this disease.

Second, we explored, for the first time in the Moroccan population, the most common epigenetic alteration in PCa, which is the hypermethylation of *GSTP1* that many studies shown its relevant role as an efficient biomarker for the diagnosis and prognosis of prostate cancer. Our study results suggest that blood *GSTP1* methylation is a promising source for a non-invasive biomarker that may therefore be used in the clinical management routine of this disease to stratify prostate cancer patients with negative initial biopsy results.

Third, we investigated the most common fusion gene in PCa, which is the *TMPRSS2-ERG* fusion gene. This study is the first study of its kind in the Moroccan population and confirmed as a result the presence of this common biomarker of prostate cancer in almost 60% of cases in the Moroccan PCa samples. This will be useful for future research, and will show the increased interest in integrating such biomarkers for prostate cancer aggressiveness detection and prostate cancer diagnosis treatment processes.

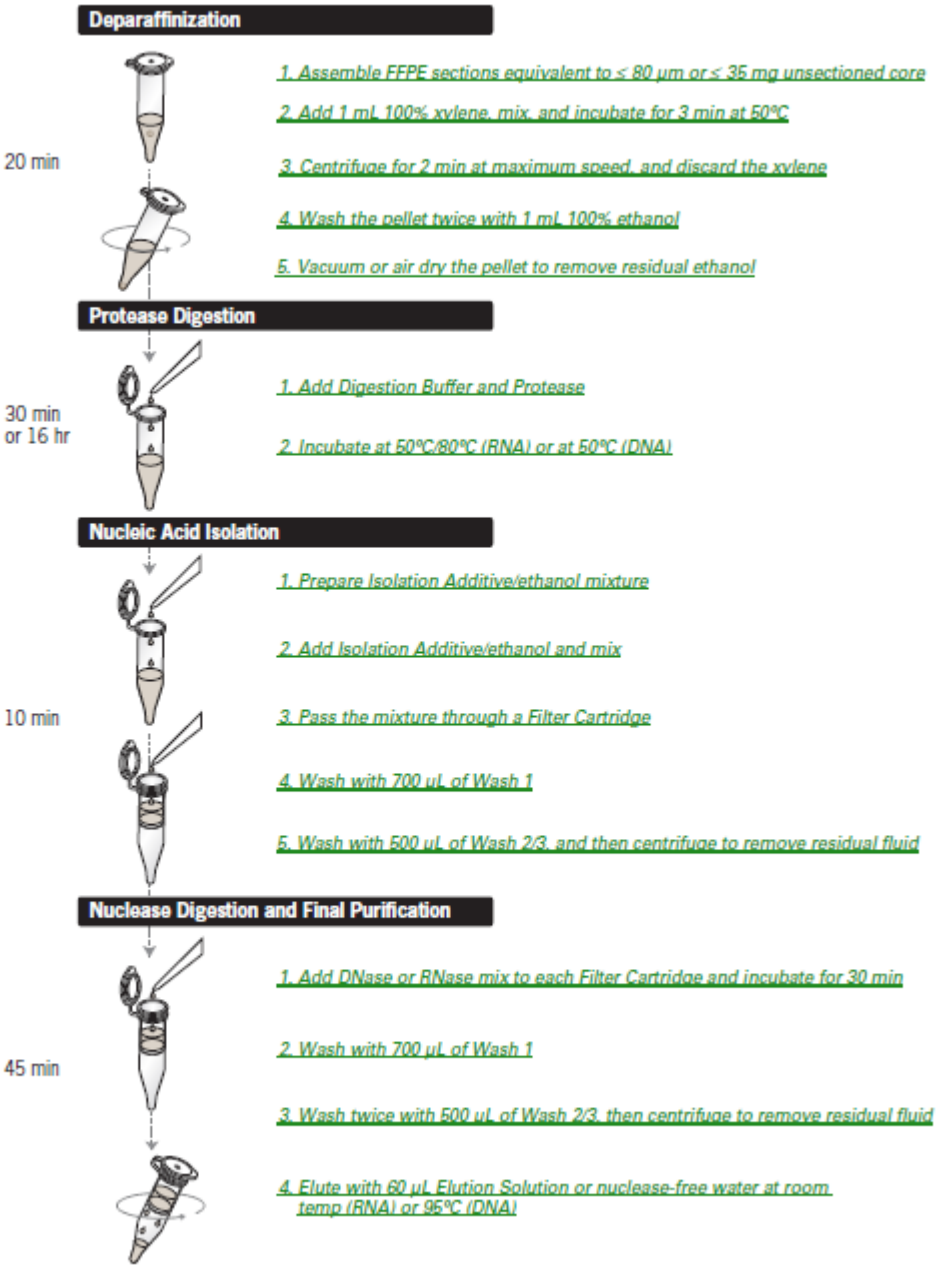
Finally, the last study identified some genomic variants of prostate cancer patients by whole-exome sequencing. We performed a somatic variant analysis on the generated data and we

found a relevant number of novel variants and mutated genes as well as many affected signaling pathways. This was the first study of its kind in the Moroccan population. Further analysis will be performed in order to identify all the germline mutation with their frequencies that will help to characterize the full mutation spectrum and catalog the somatic CNV in Moroccan men, which will provide key knowledge on the development and progression of PCa on genomic level. In addition, to novel and unique variants associated with prostate cancer in the Moroccan population aside with previously unrecognized gene coding variants with the potential to influence tumor behavior. This further analysis will help to identify new markers in order to characterize the abnormal exome regions of prostate cancer in the Moroccan population.

All these studies can be a step forward in helping stratify treatment strategies and personalize follow-up for patients.

APPENDIX

Appendix 1: Overview of the RecoverAll Total Nucleic Acid Isolation Procedure



Appendix 2: Preparation of 2% Agarose Gel for Electrophoresis

- Mix 2 g of Agarose powder with 100 ml of 1xTAE (diluted from a 10x stock) and dispense into the beaker
- Mix by swirling gently and transfer the beaker into the microwave at medium high to heat and boil for 1min
- Let agarose solution cool down for 1-2min at Room temperature
- Add 2.5 µL of Ethidium bromide into the gel while it's still liquid and swirl gently to mix
- Assemble the electrophoretic gel casting tray on a level surface and put the combs in place
- Pour the gel gently into the tray to avoid bubbles
- Leave gel to set for 20min to polymerize
- Remove the comb from the solidified gel
- Transfer the gel tray with the solidified gel to a horizontal electrophoretic gel box that is filled with 1X TAE buffer to completely cover the gel.
- Load a molecular weight ladder into the first lane of the gel and the PCR product of the samples mixed with 2ug of BET into the rest of lanes.

Appendix 3: Thermo Scientific EpiJET Bisulfite Conversion Kit: REAGENT PREPARATION

• **Preparation of Modification Reagent**

Modification Reagent is supplied as a dry mixture and needs to be dissolved before use. The amount of Modification Reagent in each vial provided is sufficient for 10 DNA conversion reactions. Prepare only as many tubes as needed for current experiment. For 10 conversions reactions:

1. Add 0.9 mL of molecular biology grade water, 200 µL of Modification Solution I and 60 µL of Modification Solution II to one vial of dry Modification Reagent.
2. Dissolve Modification Reagent by inverting for about 10 min (trace amounts of undissolved bisulfite may remain).

For best results, the prepared reagent should be used immediately.

Modification Reagent solution can be stored for up to one week at 4ºC or 4 weeks in the dark at -20ºC. Modification Reagent solution taken from the refrigerator must be warmed to

room

temperature and vortexed prior to use.

• **Preparation of Wash Buffer**: add 25 mL of 96-100% ethanol to the 9 mL of Wash Buffer.

• **Preparation of Desulfonation Buffer**: add 10 mL of 96-100% ethanol to the 3.5 mL of Desulfonation Buffer

Appendix 4 : MagMAX DNA/RNA Ultra Kit Protocol

- Add Protease Solution to each sample according to the following table

Sections	Protease Solution volume
≤40 µm	110 µL
>40 µm	210 µL

- Gently flick the tube to mix and to immerse the tissue.
- Incubate at 55°C for 1 hour or longer, then centrifuge briefly to collect any condensation droplets.
- Incubate at 90°C for 1 hour, then centrifuge briefly to collect any condensation droplets.
- Mix for 5 minutes according to the following table.

Sections	Mixing method
≤40 µm	Shake at speed 8 or 1000 rpm
>40 µm	Shake at speed 10 or 1150 rpm

- Place the sample on the magnetic stand for 2 minutes or until the solution clears and the beads are pelleted against the magnet.
- Carefully transfer the RNA-containing supernatant with a pipette to a new tube
- Wash the beads with DNA Wash Buffer according to the following table

Sections	DNA Wash Buffer volume
≤40 µm	200 µL
>40 µm	400 µL

- Shake for 1–2 minutes at speed 9 or 1100 rpm until the mixture is thoroughly chocolate brown in color.
- Place the sample on the magnetic stand for 2 minutes or until the solution clears and

the beads are pelleted against the magnet.

- Carefully discard the supernatant with a pipette.
- Wash the beads with Wash Solution 2 according to the following table.

Sections	Wash Solution 2 volume
≤40 µm	200 µL
>40 µm	500 µL

- Shake for 1 minute at speed 10 or 1150 rpm until the mixture is thoroughly chocolate brown in color.
- Place the sample on the magnetic stand for 2 minutes or until the solution clears and the beads are pelleted against the magnet.
- Carefully discard the supernatant with a pipette.
- Shake for 1–3 minutes at speed 10 or 1150 rpm to dry the beads.
- Add 50 µL of Elution Solution to the beads.
- Shake for 5 minutes at speed 10 or 1150 rpm until the mixture is thoroughly chocolate brown in color.
- Place the sample on the magnetic stand for 2 minutes or until the solution clears and the beads are pelleted against the magnet.
- The supernatant contains the purified DNA
- The purified DNA is ready for immediate use. Store at –20°C or –80°C for long-term storage.

OMICS A Journal of Integrative Biology
Volume 25, Number 4, 2021
© Mary Ann Liebert, Inc.
DOI: 10.1089/omi.2021.0004

Review Articles

Human OMICs and Computational Biology Research in Africa: Current Challenges and Prospects

Yosr Hamdi,[1,2] Lyndon Zass,[3,*] Houcemeddine Othman,[4,*] Fouzia Radouani,[5] Imane Allali,[3,6] Mariem Hanachi,[7,8] Chiamaka Jessica Okeke,[9] Melek Chaouch,[7] Maureen Bilinga Tendwa,[9] Chaimae Samtal,[10,11] Reem Mohamed Sallam,[12,13] Nihad Alsayed,[14] Michael Turkson,[15] Samah Ahmed,[14] Alia Benkahla,[7] Lilia Romdhane,[1,8] Oussema Souiai,[7] Özlem Tastan Bishop,[9] Kais Ghedira,[7,†] Faisal Mohamed Fadlelmola,[14,†] Nicola Mulder,[3,†] Samar Kamal Kassim[12,†]

Downloaded by 196.74.155.135 from www.liebertpub.com at 10/06/22. For personal use only.

Abstract

Following the publication of the first human genome, OMICs research, including genomics, transcriptomics, proteomics, and metagenomics, has been on the rise. OMICs studies revealed the complex genetic diversity among human populations and challenged our understandings of genotype–phenotype correlations. Africa, being the cradle of the first modern humans, is distinguished by a large genetic diversity within its populations and rich ethnolinguistic history. However, the available human OMICs tools and databases are not representative of this diversity, therefore creating significant gaps in biomedical research. African scientists, students, and publics are among the key contributors to OMICs systems science. This expert review examines the pressing issues in human OMICs research, education, and development in Africa, as seen through a lens of computational biology, public health relevant technology innovation, critically-informed science governance, and how best to harness OMICs data to benefit health and societies in Africa and beyond. We underscore the disparities between North and Sub-Saharan Africa at different levels. A harmonized African ethnolinguistic classification would help address annotation challenges associated with population diversity. Finally, building on the existing strategic research initiatives, such as the H3Africa and H3ABioNet Consortia, we highly recommend addressing large-scale multidisciplinary research challenges, strengthening research collaborations and knowledge transfer, and enhancing the ability of African researchers to influence and shape national and international research, policy, and funding agendas. This article and analysis contribute to a deeper understanding of past and current challenges in the African OMICs innovation ecosystem, while also offering foresight on future innovation trajectories.

[1]Laboratory of Biomedical Genomics and Oncogenetics, Institut Pasteur de Tunis, Université Tunis El Manar, Tunis, Tunisia.
[2]Laboratory of Human and Experimental Pathology, Institut Pasteur de Tunis, Tunis, Tunisia.
[3]Computational Biology Division, Department of Integrative Biomedical Sciences, Institute of Infectious Disease and Molecular Medicine, CIDRI Africa Wellcome Trust Centre, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa.
[4]Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa.
[5]Chlamydiae and Mycoplasmas Laboratory, Institut Pasteur du Maroc, Casablanca, Morocco.
[6]Laboratory of Human Pathologies Biology, Department of Biology, Faculty of Sciences, and Genomic Center of Human Pathologies, Faculty of Medicine and Pharmacy, Mohammed V University in Rabat, Rabat, Morocco.
[7]Laboratory of Bioinformatics, Biomathematics and Biostatistics, Institut Pasteur de Tunis, Université Tunis El Manar, Tunis, Tunisia.
[8]Faculty of Science of Bizerte, Zarzouna, University of Carthage, Tunis, Tunisia.
[9]Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Makhanda, South Africa.
[10]Laboratory of Biotechnology, Environment, Agri-food and Health, Faculty of Sciences Dhar El Mahraz–Sidi Mohammed Ben Abdellah University, Fez, Morocco.
[11]University of Mohamed Premier, Oujda, Morocco.
[12]Department of Medical Biochemistry and Molecular Biology, Faculty of Medicine, Ain Shams University, Cairo, Egypt.
[13]Department of Basic Medical Sciences, Faculty of Medicine, Galala University, Suez, Egypt.
[14]Centre for Bioinformatics and Systems Biology, Faculty of Science, University of Khartoum, Khartoum, Sudan.
[15]The National Institute for Mathematical Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.
*These authors contributed equally to this work.
†Joint senior authors.

Review

For reprint orders, please contact: reprints@futuremedicine.com

Personalized Medicine

A review of clinical pharmacogenetics Studies in African populations

Fouzia Radouani[1], Lyndon Zass[2], Yosr Hamdi[3], Jorge da Rocha[4], Reem Sallam[5], Sonia Abdelhak[3], Samah Ahmed[6,7], Maryame Azzouzi[1], Ichrak Benamri[1,8], Alia Benkahla[9], Balkiss Bouhaouala-Zahar[10], Melek Chaouch[9], Haifa Jmel[3], Rym Kefi[3], Ayoub Ksouri[9,10], Judit Kumuthini[11], Phumlani Masilela[2], Collen Masimirembwa[4,12], Houcemeddine Othman[4], Sumir Panji[2], Lilia Romdhane[3,13], Chaimae Samtal[14,1516], Rania Sibira[6,17], Kais Ghedira[9], Faisal Fadlelmola[6], Samar Kamal Kassim[‡,5] & Nicola Mulder*[,‡,2]

[1]Research Department, Chlamydiae & Mycoplasmas Laboratory, Institut Pasteur du Maroc, Casablanca 20360, Morocco
[2]Computational Biology Division, Department of Integrative Biomedical Sciences, IDM, CIDRI Africa Wellcome Trust Centre, University of Cape Town, South Africa
[3]Laboratory of Biomedical Genomics & Oncogenetics, Institut Pasteur de Tunis, Université Tunis El Manar, 13, Place Pasteur BP 74, 1002 Tunis, Belvédère, Tunisie
[4]Sydney Brenner Institute for Molecular Bioscience, University of The Witwatersrand, Johannesburg, South Africa
[5]Medical Biochemistry & Molecular Biology Department, Faculty of Medicine, Ain Shams University, Abbaseya, Cairo 11381, Egypt
[6]Centre for Bioinformatics & Systems Biology, Faculty of Science, University of Khartoum, 321 Khartoum, Sudan
[7]Faculty of Clinical & Industrial Pharmacy, National University, Khartoum, Sudan
[8]Systems & Data Engineering Team, National School of Applied Sciences of Tangier, Morocco
[9]Laboratory of Bioinformatics, Biomathematics & Biostatistics LR 16 IPT 09, Institute Pasteur de Tunis, Tunisia
[10]Laboratory of Venoms & Therapeutic Molecules, Pasteur Institute of Tunis, 13 Place Pasteur, BP74, Tunis Belvedere- University of Tunis El Manar, Tunisia
[11]H3ABioNet, Bioinformatics Department, Centre for Proteomic & Genomic Research, Cape Town, South Africa
[12]DMPK Department, African Institute of Biomedical Science & Technology, Harare, Zimbabwe
[13]Département des Sciences de la Vie, Faculté des Sciences de Bizerte, Université Carthage, 7021 Jarzouna, BP 21, Tunisie
[14]Biotechnology Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohammed Ben Abdellah University, Fez 30000, Morocco
[15]Department of Biology, University of Mohammed Premier, Oujda, Morocco
[16]Department of Biology Faculty of Sciences, University of Sidi Mohamed Ben Abdellah, Fez, Morocco
[17]Department of Neurosurgery, National Center For Neurological Sciences, Khartoum, Sudan
*Author for correspondence: nicola.mulder@uct.ac.za
‡ Joint senior authors

Effective interventions and treatments for complex diseases have been implemented globally, however, coverage in Africa has been comparatively lower due to lack of capacity, clinical applicability and knowledge on the genetic contribution to disease and treatment. Currently, there is a scarcity of genetic data on African populations, which have enormous genetic diversity. Pharmacogenomics studies have the potential to revolutionise treatment of diseases, therefore, African populations are likely to benefit from these approaches to identify likely responders, reduce adverse side effects and optimise drug dosing. This review discusses clinical pharmacogenetics studies conducted in African populations, focusing on studies that examined drug response in complex diseases relevant to healthcare. Several pharmacogenetics associations have emerged from African studies, as have gaps in knowledge.

First draft submitted: 21 September 2019; Accepted for publication: 24 January 2020; Published online: 3 March 2020

Keywords: Africa • communicable diseases • genetic variation • noncommunicable diseases • pharmacogenetics • pharmacogenomics • precision medicine

There is huge genetic, cultural and lifestyle diversity among African populations that influence susceptibility to disease, disease progression and response to medical treatment employed against disease [1]. This applies to communicable and noncommunicable diseases, of which African countries face a huge burden of both. The challenges of the modern day patient compounded by the complexities associated with the poly-pharmacy requires careful management of dosages for the most optimal patient benefits while at the same time reducing the incidence

Future Medicine

Journal of Biomedical Informatics 122 (2021) 103900

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Special communication

Data Management Plans in the genomics research revolution of Africa: Challenges and recommendations

Faisal M. Fadlelmola [a,*], Lyndon Zass [b,1], Melek Chaouch [c,1], Chaimae Samtal [d,1], Verena Ras [b,1], Judit Kumuthini [e], Sumir Panji [b], Nicola Mulder [b]

[a] Centre for Bioinformatics and Systems Biology, Faculty of Science, University of Khartoum, Al-Gamaa Ave, Khartoum 11115, Sudan
[b] Computational Biology Division, Department of Integrative Biomedical Sciences, IDM, CIDRI Africa Wellcome Trust Centre, University of Cape Town, South Africa
[c] Laboratory of Bioinformatics Biomathematics and Biostatistics (LR16IPT09), Institut Pasteur de Tunis, 13 Place Pasteur, B.P. 74 1002 Tunis, Belvédère, Tunisia
[d] Laboratory of Biotechnology, Environment, Agri-food and Health, Faculty of Sciences Dhar El Mahraz–Sidi Mohammed Ben Abdellah University, Fez 30000, Morocco
[e] South African Bioinformatics Institute (SANBI), University of Western Cape (UWC), Life Sciences Building, Bellville, Cape Town, South Africa

ARTICLE INFO

Keyword
Data Management Plan
Africa
FAIR
Funders
Data sharing
Genomics data management

ABSTRACT

Drafting and writing a data management plan (DMP) is increasingly seen as a key part of the academic research process. A DMP is a document that describes how a researcher will collect, document, describe, share, and preserve the data that will be generated as part of a research project. The DMP illustrates the importance of utilizing best practices through all stages of working with data while ensuring accessibility, quality, and longevity of the data. The benefits of writing a DMP include compliance with funder and institutional mandates; making research more transparent (for reproduction and validation purposes); and FAIR (findable, accessible, interoperable, reusable); protecting data subjects and compliance with the General Data Protection Regulation (GDPR) and/or local data protection policies. In this review, we highlight the importance of a DMP in modern biomedical research, explaining both the rationale and current best practices associated with DMPs. In addition, we outline various funders' requirements concerning DMPs and discuss open-source tools that facilitate the development and implementation of a DMP. Finally, we discuss DMPs in the context of African research, and the considerations that need to be made in this regard.

1. Introduction

The African continent faces a large burden of both communicable and non-communicable diseases [1]. Individuals of African descent are known to possess high genetic diversity, which may influence susceptibility and resistance to, as well as treatment of these diseases [2,3]. To investigate the link between genetics and disease, African researchers are increasingly performing genome-sequencing studies that generate large datasets. The Human Heredity and Health in Africa (H3Africa) consortium, funded by the National Institutes of Health (NIH) and the Wellcome Trust (WT), through AESA (Alliance for Accelerating Excellence in Science in Africa), support several such studies and have thus far generated genome sequence and/or variation data for over 50,000 African individuals [4]. The H3Africa bioinformatics network (H3ABioNet) aims to develop bioinformatics capacity and provide infrastructure to support genomic analysis and data storage throughout the continent [5,6]. The increasing number of genomic studies in Africa, have also led to an increasing number of training opportunities, which facilitate genomics data analysis, computational skills, and data management.

Data management is an important part of any large-scale research process, particularly one that involves genomics- and related platforms. To keep track of data generated in such studies, as well as effectively organize and maintain it, a Data Management Plan (DMP) is essential. A DMP is a document that is developed before the start of a research project and details how research data will be collected, generated, and processed during the project, as well as stored and shared thereafter. DMPs have become a core component of modern academia due to the increasing amount of Big Data studies being conducted worldwide. The benefits of writing a DMP are multifold, including: 1) allowing compliance with funder mandates; 2) enabling transparent research; 3) enhancing data FAIRness (findability, accessibility, interoperability, and reusability); 4) protecting data subjects; and 5) allowing compliance

* Corresponding author.
E-mail address: faisalfadl@gmail.com (F.M. Fadlelmola).
[1] These authors contributed equally to this manuscript.

https://doi.org/10.1016/j.jbi.2021.103900
Received 12 March 2021; Received in revised form 24 August 2021; Accepted 25 August 2021
Available online 8 September 2021
1532-0464/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

DATABASE

Database, 2021, 1–13
doi:10.1093/database/baab016
Database tool

Database tool

H3ABioNet genomic medicine and microbiome data portals hackathon proceedings

Faisal M. Fadlelmola[1,*,†], Kais Ghedira[2,*,†], Yosr Hamdi[3,‡], Mariem Hanachi[2,4,‡], Fouzia Radouani[5,‡], Imane Allali[6,7], Anmol Kiran[8], Lyndon Zass[9], Nihad Alsayed[1], Meriem Fassatoui[3], Chaimae Samtal[10], Samah Ahmed[1], Jorge Da Rocha[11], Souad Chaqsare[12], Reem M. Sallam[13,14], Melek Chaouch[2], Mohammed Farahat[15], Alfred Ssekagiri[16], Ziyaad Parker[9], Mai Adil[1], Michael Turkson[17], Aymen Benchaalia[2], Alia Benkahla[2], Sumir Panji[9], Samar Kassim[13], Oussema Souiai[2] and Nicola Mulder[9]

[1]Centre for Bioinformatics and Systems Biology, Faculty of Science, University of Khartoum, Al-Gamaa Ave, Khartoum 11115, Sudan, [2]Laboratory of Bioinformatics, Biomathematics and Biostatistics (BIMS), Institut Pasteur de Tunis (IPT), 13, Place Pasteur BP 74, Tunis 1002, Tunisia, [3]Laboratory of Biomedical Genomics & Oncogenetics, Institut Pasteur de Tunis, Université Tunis El Manar, 13, Place Pasteur BP 74, Tunis 1002, Tunisia, [4]Faculty of Science of Bizerte, University of Carthage, Zarzouna, Bizerte 7021, Tunisia, [5]Research Department, Chlamydiae and Mycoplasmas Laboratory, Institut Pasteur du Maroc, 1, Place Louis Pasteur, Casablanca 20360, Morocco, [6]Laboratory of Human Pathologies Biology, Department of Biology, Faculty of Sciences, Mohammed V University, 4 Ibn Battouta Avenue, Rabat, BP 1014 RP, Morocco, [7]Genomic Center of Human Pathologies, Faculty of Medicine and Pharmacy, Mohammed V University, Rabat, Morocco, [8]Malawi-Liverpool Wellcome Trust, Clinical Research Programme, PO Box 30096, Chichiri, Blantyre 3, Blantyre, Malawi, [9]Computational Biology Division, N1.05 Werner Beit North, Department of Integrative Biomedical Sciences, Faculty of Health Sciences, Anzio Road, Observatory, Cape Town 7925, South Africa, [10]Faculty of Sciences Dhar El Mahraz, Department of Biology, Genetics Unit, Atlas-Fez 1796, Morocco, [11]Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, 9 Jubilee Road, Parktown, Johannesburg 2193, South Africa, [12]National Institute of Health, Informatics Unit, 27 Ibn Batouta Avenue, Agdal, Rabat BP 769, Morocco, [13]Medical Biochemistry & Molecular Biology, Faculty of Medicine, Ain Shams University, Abassia, Cairo 11381, Egypt, [14]Department of Basic Medical Sciences, Faculty of Medicine, Galala University, Galala City, Suez 43511, Egypt, [15]Information Systems Department, Faculty of Computers and Artificial Intelligence, Helwan University, Ain Helwan, PO Box 11795, Cairo, Egypt, [16]Division of Entomology and core molecular Biology/Bioinformatics facility, Uganda Virus Research Institute, 51/59, Nakiwogo Road, Entebbe 31301, Uganda and [17]National Institute for Mathematical Sciences, PMB Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana

*Corresponding author: Tel: +249911556077; Email: faisalfadl@gmail.com
Correspondence may also be addressed to Kais Ghedira. Tel: +2167843755; Email: ghedirakais@gmail.com
†These authors contributed equally to this work.
‡These authors contributed equally to this work as second authors.

Citation details: Fadlelmola, F.M., Ghedira, K., Hamdi, Y. *et al.* H3ABioNet genomic medicine and microbiome data portals hackathon proceedings. *Database* (2021) Vol. 2021: article ID baab016; doi:10.1093/database/baab016

Received 9 November 2020; Revised 12 February 2021; Accepted 29 March 2021

© The Author(s) 2021. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/database/article/doi/10.1093/database/baab016/6232122 by guest on 06 October 2022

F1000Research F1000Research 2021, 10:1002 Last updated: 15 FEB 2022

Check for updates

METHOD ARTICLE

Performing post-genome-wide association study analysis: overview, challenges and recommendations [version 1; peer review: 2 approved]

Yagoub Adam [1], Chaimae Samtal [2], Jean-tristan Brandenburg [3], Oluwadamilare Falola [2], Ezekiel Adebiyi [1,4-6]

[1]Covenant University Bioinformatics Research (CUBRe), Covenant University, Ota, Ogun, 112233, Nigeria
[2]Laboratory of Biotechnology, Environment, Agri-food and Health, Sidi Mohammed Ben Abdellah University, Fez, Fez-Meknes, 30000, Morocco
[3]Sydney Brenner Institute for Molecular Bioscience (SBIMB), University of the Witwatersrand, Johannesburg, South Africa
[4]Computer & Information Sciences, Covenant University, Ota, Ogun, 112233, Nigeria
[5]Covenant Applied Informatics and Communication Africa Centre of Excellence, Covenant University, Ota, Ogun, 112233, Nigeria
[6]Applied Bioinformatics Division, German Cancer Center DKFZ - Heidelberg University, Heidelberg, Baden-Württemberg, 69120, Germany

v1 **First published:** 04 Oct 2021, **10**:1002
https://doi.org/10.12688/f1000research.53962.1
Latest published: 04 Oct 2021, **10**:1002
https://doi.org/10.12688/f1000research.53962.1

Abstract

Genome-wide association studies (GWAS) provide huge information on statistically significant single-nucleotide polymorphisms (SNPs) associated with various human complex traits and diseases. By performing GWAS studies, scientists have successfully identified the association of hundreds of thousands to millions of SNPs to a single phenotype. Moreover, the association of some SNPs with rare diseases has been intensively tested. However, classic GWAS studies have not yet provided solid, knowledgeable insight into functional and biological mechanisms underlying phenotypes or mechanisms of diseases. Therefore, several post-GWAS (pGWAS) methods have been recommended. Currently, there is no simple scientific document to provide a quick guide for performing pGWAS analysis. pGWAS is a crucial step for a better understanding of the biological machinery beyond the SNPs. Here, we provide an overview to performing pGWAS analysis and demonstrate the challenges behind each method. Furthermore, we direct readers to key articles for each pGWAS method and present the overall issues in pGWAS analysis. Finally, we include a custom pGWAS pipeline to guide new users when performing their research.

Keywords

PostGWAS, pGWAS, GWAS, Meta-analysis

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 04 Oct 2021	✓ view	✓ view

1. **Zhe Zhang**, South China Agricultural University, Guangzhou, China

2. **Yasuhiro Sato**, University of Zurich, Zürich, Switzerland

Any reports and responses or comments on the article can be found at the end of the article.

Journal of *Personalized Medicine*

MDPI

Article

African Genomic Medicine Portal: A Web Portal for Biomedical Applications

Houcemeddine Othman [1], Lyndon Zass [2], Jorge E. B. da Rocha [1,3], Fouzia Radouani [4], Chaimae Samtal [5], Ichrak Benamri [4,6], Judit Kumuthini [7], Yasmina J. Fakim [8], Yosr Hamdi [9,10], Nessrine Mezzi [9,11], Maroua Boujemaa [9], Chiamaka Jessica Okeke [12], Maureen B. Tendwa [12], Kholoud Sanak [4], Melek Chaouch [13], Sumir Panji [2], Rym Kefi [9], Reem M. Sallam [14,15], Anisah W. Ghoorah [16], Lilia Romdhane [9,11], Anmol Kiran [17,18], Ayton P. Meintjes [2], Perceval Maturure [2], Haifa Jmel [9], Ayoub Ksouri [19], Maryame Azzouzi [4], Mohammed A. Farahat [20], Samah Ahmed [21], Rania Sibira [21], Michael E. E. Turkson [22], Alfred Ssekagiri [23], Ziyaad Parker [2], Faisal M. Fadlelmola [21], Kais Ghedira [13], Nicola Mulder [2,†] and Samar Kamal Kassim [15,24,*,†]

1 Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, 9 jubilee Road, Parktown, Johannesburg 2193, South Africa; houcemeddine.othman@wits.ac.za (H.O.); jdarocha1@gmail.com (J.E.B.d.R.)
2 Computational Biology Division, Department of Integrative Biomedical Sciences, IDM, CIDRI Africa Wellcome Trust Centre, University of Cape Town, Rondebosch, Cape Town 7701, South Africa; lyndon.zass@uct.ac.za (L.Z.); sumir.panji@uct.ac.za (S.P.); aytonm@gmail.com (A.P.M.); perceval.maturure@uct.ac.za (P.M.); ziyaad.parker@uct.ac.za (Z.P.); nicola.mulder@uct.ac.za (N.M.)
3 Division of Human Genetics, National Health Laboratory Service, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2001, South Africa
4 Chlamydiae and Mycoplasma Laboratory, Research Department, Institut Pasteur du Maroc, Casablanca 20360, Morocco; radouani@gmail.com (F.R.); ichrakbenamri@gmail.com (I.B.); kholoud.sanak@gmail.com (K.S.); maryame.azz@gmail.com (M.A.)
5 Laboratory of Biotechnology, Environment, Agri-Food and Health, Faculty of Sciences Dhar El Mahraz, Sidi Mohammed Ben Abdellah University, Fez 30000, Morocco; samtalchaimae@gmail.com
6 Systems and Data Engineering Team, National School of Applied Sciences, Abdelmalek Essaâdi University, Tangier 93000, Morocco
7 South African Bioinformatics Institute, University of Western Cape, Cape Town 7535, South Africa; jkumuthini@gmail.com
8 Biotechnology Unit, Department of Agricultural and Food Science, Faculty of Agriculture, University of Mauritius, Reduit 80837, Mauritius; yasmina@uom.ac.mu
9 Laboratory of Biomedical Genomics and Oncogenetics, LR20IPT05, Institut Pasteur de Tunis, University of Tunis El Manar, Tunis 1002, Tunisia; yosr.hamdi@pasteur.utm.tn (Y.H.); mezzinessrine1@gmail.com (N.M.); maroua.boujemaa@gmail.com (M.B.); rym.kefi@pasteur.tn (R.K.); lilia.romdhane@pasteur.utm.tn (L.R.); jmelhaifa@gmail.com (H.J.)
10 Laboratory of Human and Experimental Pathology, Institut Pasteur de Tunis, Tunis 1002, Tunisia
11 Department of Biology, Faculty of Science of Bizerte, Jarzouna, Bizerte 7021, Tunisia
12 Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Makhanda 6139, South Africa; okekechiamaka39@gmail.com (C.J.O.); bilingatendwa@gmail.com (M.B.T.)
13 Laboratory of Bioinformatics, Biomathematics and Biostatistics, Pasteur Institute of Tunisia, University of Tunis El Manar, Tunis 1068, Tunisia; mcmelek@msn.com (M.C.); kais.ghedira@pasteur.tn (K.G.)
14 Basic Medical Science Department, Faculty of Medicine, Galala University, Suez 43727, Egypt; reemsallam@med.asu.edu.eg
15 Medical Biochemistry and Molecular Biology Department, Faculty of Medicine, Ain Shams University, Cairo 11381, Egypt
16 Department of Digital Technologies, Faculty of Information, Communication and Digital Technologies, University of Mauritius, Reduit 80837, Mauritius; a.ghoorah@uom.ac.mu
17 Malawi-Liverpool-Wellcome Trust, Queen Elizabeth Central Hospital, College of Medicine, Blantyre P.O. Box 30096, Malawi; anmol.kiran@gmail.com
18 Department of Clinical Infection, Microbiology and Immunology, Institute of Infection and Global Health, University of Liverpool, Liverpool L69 3BX, UK
19 Laboratory of Venoms and Therapeutic Molecules (LR16IPT08), 1" Place Pasteur, BP74, Pasteur Institute of Tunis, University of Tunis El Manar, Tunis 1002, Tunisia; ayoub.ksouri@pasteur.utm.tn
20 Compute Science Department, The Higher Future Institute for Specialized Technological Studies, Cairo 11381, Egypt; Mohammed_Farahat19@yahoo.com

check for updates

Citation: Othman, H.; Zass, L.; da Rocha, J.E.B.; Radouani, F.; Samtal, C.; Benamri, I.; Kumuthini, J.; Fakim, Y.J.; Hamdi, Y.; Mezzi, N.; et al. African Genomic Medicine Portal: A Web Portal for Biomedical Applications. *J. Pers. Med.* **2022**, *12*, 268. https://doi.org/10.3390/jpm12020265

Academic Editors: Ali Salehzadeh-Yazdi and Mohieddin Jafari

Received: 15 December 2021
Accepted: 26 January 2022
Published: 11 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Moroccan Journal Of Public Health Vol. 1, N° 1 (2020), 15-23 ISSN: 2658-8099

Review Article

SARS-CoV-2 virus genome structure and evolution

Chaimae SAMTAL[a], Nihal HABIB[b], Abdellah IDRISSI AZAMI[b], Sofia SEHLI[b], Bouchra CHAOUNI[c], Mohamed BEKKALI[d], Hicham BEKKARI[b], Salsabil HAMDI[e], Lahcen WAKRIM[f], Najib Al IDRISSI[g], Hassan GHAZAL[h*]

[a] *Faculty of Sciences Dhar El Mehraz, USMBA, Fes, Morocco*

[b] *Faculty of Medicine and Pharmacy of Rabat, Rabat, Morocco*

[c] *Faculty of Sciences of Rabat, Rabat, Morocco*

[d] *Department of Genetics, Faculty of Sciences, University of Granada, Spain.*

[e] *Environmental Health Laboratory Institut Pasteur du Maroc, Casablanca, Morocco*

[f] *Virology Unit, Immunovirology Laboratory, Institut Pasteur du Maroc, Casablanca, Morocco.*

[g] *School of Medicine, University Mohamed VI for Health Sciences, Casablanca, Morocco*

[h] Centre National de la Recherche Scientifique et Technique, Rabat, Morocco

*Corresponding Author E-mail: [a] hassan.ghazal@fullbrightmail.org

Article info

Received	*April 2020*
Accepted	*August 2020*
Online	*September 2020*

keywords

SARS-Cov-2,coronavirus

phylogeny , alignment , ACE2

SUMMARY

Late December 2019, a new virus outbreak was detected in Wuhan, Hubei province China and later spread all over the world. The virus is a new strain of the β-Coronavirus (SARS-CoV-2) with a single strand positive sense RNA, and a genome size of 29,890 kb that codes for 9744 amino acids.

Not many mutations are depicted in the SARS-CoV-2 genome, However, most of them are detected in the ORF8 region and the spike. The latter exhibits higher affinity with the Human ACE2 receptor than the SARS-CoV.

Alignment of SARS-CoV-2 genome with other Coronaviruses shows a high similarity to the Bat-CoV, but this latter was not enough to conclude that the Bat is the origin of the Human SARS-CoV-2. This review aimed to provide an overview of the viral genome structure and its origin.

1. Introduction

Coronaviruses (CoVs), a large family of single-stranded RNA viruses, can infect animals and humans, causing respiratory, gastrointestinal, hepatic, and neurologic diseases (Weiss and Leibowitz, 2011). As the largest known RNA viruses, CoVs are further divided into four genera: α-Coronaviruses, β-Coronavirus, γ-Coronaviruses a

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 11 October 2022
DOI 10.3389/fgene.2022.911101

Check for updates

OPEN ACCESS

EDITED BY
Qiuyin Cai,
Vanderbilt University, United States

REVIEWED BY
Uzoamaka Adaobi Okoli,
University College London,
United Kingdom
Adam de Smith,
University of Southern California,
United States

*CORRESPONDENCE
Hassan Ghazal,
hassan.ghazal@fulbrightmail.org
Hicham Bekkari,
hichambekkari@yahoo.fr

SPECIALTY SECTION
This article was submitted to Applied Genetic Epidemiology, a section of the journal Frontiers in Genetics

RECEIVED 04 April 2022
ACCEPTED 28 September 2022
PUBLISHED 11 October 2022

CITATION
Samtal C, El Jaddaoui I, Hamdi S, Bouguenouch L, Ouldim K, Nejjari C, Ghazal H and Bekkari H (2022). Review of prostate cancer genomic studies in Africa.
Front. Genet. 13:911101.
doi: 10.3389/fgene.2022.911101

COPYRIGHT
© 2022 Samtal, El Jaddaoui, Hamdi, Bouguenouch, Ouldim, Nejjari, Ghazal and Bekkari. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Review of prostate cancer genomic studies in Africa

Chaimae Samtal[1], Islam El Jaddaoui[2], Salsabil Hamdi[3], Laila Bouguenouch[4], Karim Ouldim[4], Chakib Nejjari[5,6], Hassan Ghazal[1,7,8]* and Hicham Bekkari[1]*

[1]Laboratory of Biotechnology, Environment, Agri-food and Health, Faculty of Sciences Dhar El Mahraz–Sidi Mohammed Ben Abdellah University, Fez, Morocco, [2]Laboratory of Human Pathologies Biology, Department of Biology, Faculty of Sciences, and Genomic Center of Human Pathologies, Faculty of Medicine and Pharmacy, University Mohammed V, Rabat, Morocco, [3]Laboratory of Environmental Health, Institut Pasteur Maroc, Casablanca, Morocco, [4]Medical Genetics and Oncogenetics Unit, Faculty of Medicine and Pharmacy, University Hospital Hassan II, Fez, Morocco, [5]Department of Medicine, School of Medicine, Mohammed VI University of Health Sciences, Casablanca, Morocco, [6]School of Medicine and Pharmacy, Fes, Morocco, [7]Laboratory of Genomics and Bioinformatics, School of Pharmacy, Mohammed VI University of Health Sciences, Casablanca, Morocco, [8]National Center for Scientific and Technical Research, Rabat, Morocco

Prostate cancer (PCa) is the second most commonly diagnosed in men worldwide and one of the most frequent cancers in men in Africa. The heterogeneity of this cancer fosters the need to identify potential genetic risk factors/biomarkers. Omics variations may significantly contribute to early diagnosis and personalized treatment. However, there are few genomic studies of this disease in African populations. This review sheds light on the status of genomics research on PCa in Africa and outlines the common variants identified thus far. The allele frequencies of the most significant SNPs in Afro-native, Afro-descendants, and European populations were compared. We advocate how these few but promising data will aid in understanding, better diagnosing, and precisely treating this cancer and the need for further collaborative research on the genomics of PCa in the African continent.

KEYWORDS
prostate cancer, omics, genomics, variants, GWAS, Africa

Introduction

Prostate cancer (PCa) is currently the third most aggressive neoplasm globally. PCa has the second highest incidence woldwide in men (Jemal et al., 2008). PCa is the second most common malignancy diagnosed (after lung cancer) in men and the fifth leading cause of cancer death globally (Pernar et al., 2018). In 2018, an estimated 1,276,106 new cases of PCa were reported worldwide based on GLOBOCAN (Rawla, 2019). In 2020, according to the World Health Organization (WHO), 1,414,259 new cases were recorded, which is 7.3% of the total number of new cancer cases recorded (Giona et al., 2021). PCa is particularly widespread in developed countries. In the United States, 180,890 new cases were reported in 2016, with 26,120 deaths (Pernar et al., 2018). The PCa mortality rates