USING FAMILIES TO UNDERSTAND THE IMPACT OF GENETIC VARIATION ON PROSTATE CANCER

by

Kelsie Raspin

BBiomedSc, BMedSc (Hons)

Menzies Institute for Medical Research

College of Health and Medicine

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy (Medical Studies)

University of Tasmania, December, 2019

DECLARATION OF ORIGINALITY

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of this thesis, nor does the thesis contain any material that infringes copyright.

Kelsie Raspin

27/11/2019

AUTHORITY OF ACCESS STATEMENT

The publishers of the paper comprising Chapter 5 hold the copyright for that content, and access to the material should be sought from the journal. The remaining non-published content of the thesis may be made available for loan and limited copying and communication in accordance with the *Copyright Act 1968*.

Kelsie Raspin

27/11/2019

STATEMENT OF ETHICAL CONDUCT

The research associated with this thesis abides by the international and Australian codes on human and animal experimentation, the guidelines by the Australian Government's Office of the Gene Technology Regulator and the rulings of the Safety, Ethics and Institutional Biosafety Committees of the University.

Kelsie Raspin

27/11/2019

STATEMENT OF CO-AUTHORSHIP

Kelsie Raspin has incorporated a version of a co-first-author paper, FitzGerald & Raspin *et al.* (2017) [1], into Chapter 5.

FitzGerald LM*, **Raspin K***, Marthick JR, *et al*. **Impact of the G84E variant on HOXB13 gene and protein expression in formalin-fixed, paraffin embedded prostate tumours**. *Sci Rep* 2017; 7:17778. *Joint first authors.

The co-authors and each authors contribution are outlined as follows:

FitzGerald, LM. and **Raspin, K.** contributed equally to this work.

FitzGerald, LM. and **Raspin, K.** wrote the main manuscript and prepared the tables. **Raspin, K.** prepared the figures. **Raspin, K.** and Marthick, JR. performed the laboratory work. **Raspin, K.,** Field, MA., Thomson, RJ. performed the analyses. Banks, A. traced the participants and Malley, R. and Donovan, S. sourced and reviewed the pathology material. Blackburn, NB. and Charlesworth, JC. were involved in the generation and analysis of the sequencing data. Dickinson, JL. and FitzGerald, LM. designed and directed the study and all authors reviewed the manuscript.

Prof. Joanne Dickinson
Primary Supervisor
Menzies Institute for Medical Research
University of Tasmania
19/12/2019

Prof. Alison Venn
Director
Menzies Institute for Medical Research
University of Tasmania
28/11/2019

STATEMENT OF CO-AUTHORSHIP

Data arising from this thesis has also been presented at the following scientific meetings:

Raspin K, FitzGerald LM, Marthick JR, Field MA, Malley RC, Thomson RJ, Banks A, Donovan S, Stanford J, Dickinson JL. A rare variant in the histone methyltransferase, *EZH2* is associated with prostate cancer risk in our Tasmanian resource. American Society of Human Genetics, San Diego, CA, USA, 2018 (Poster Presentation).

Raspin K, FitzGerald LM, Marthick JR, Field MA, Malley RC, Thomson RJ, Banks A, Donovan S, Stanford J, Dickinson JL. A rare variant in the histone methyltransferase, *EZH2* is associated with prostate cancer risk in our Tasmanian resource. GeneMappers, Noosa, QLD, 2018 (Poster Presentation).

Raspin K, FitzGerald LM, Marthick JR, Field MA, Thomson RJ, Banks A, Stanford J, Dickinson JL. The identification of a rare variant in the histone methyltransferase, *EZH2*, in Tasmanian prostate cancer pedigrees, using whole-genome sequencing. Lorne Genome Conference, Lorne, VIC, 2018 (Poster Presentation).

Raspin K, FitzGerald LM, Marthick JR, Field MA, Thomson RJ, Banks A, Stanford J, Dickinson JL. The identification of a rare variant in the histone methyltransferase, *EZH2*, in Tasmanian prostate cancer pedigrees, using whole-genome sequencing. Lorne Cancer Conference, Lorne, VIC, 2018 (Oral/Poster Presentation).

Raspin K, FitzGerald LM, Marthick JR, Field MA, Thomson RJ, Banks A, Stanford J, Dickinson JL. The identification of a rare variant in the histone methyltransferase, *EZH2*, in Tasmanian prostate cancer pedigrees, using whole-genome sequencing. Australasian Genomic Technologies Association Conference, Hobart, TAS, 2017 (Poster Presentation).

Raspin K, FitzGerald LM, Marthick JR, Field MA, Malley R, Thomson RJ, Blackburn N, Banks A, Charlesworth JC, Donovan S, Stanford JL, Dickinson JL. Comparison of HOXB13 gene and protein expression in prostate tumours of G84E variant carriers and non-carriers. GeneMappers, Geelong, VIC, 2017 (Oral/Poster Presentation).

Raspin K, FitzGerald LM, Marthick JR, Field MA, Malley R, Thomson RJ, Blackburn N, Banks A, Charlesworth JC, Donovan S, Stanford JL, Dickinson JL. Comparison of HOXB13 gene and protein expression in prostate tumours of G84E variant carriers and non-carriers. UTAS Graduate Research Conference, Hobart, TAS, 2016 (Poster Presentation).

ACKNOWLEDGEMENTS

There is one major difference between the beginning and end of your PhD and that is, in the beginning you have no idea where to start and in the end you have no idea where to finish. I guess they are actually very similar, they are the 'unknown' and bring about a great sense of doubt. There a number of people who I would like to thank for guiding me through this journey.

Firstly, a PhD shapes the researcher that you aim to be and I'll be forever in debt to my supervisors, Prof. Jo Dickinson and Dr Liesel FitzGerald for their knowledge, guidance, banter and trust. I feel privileged that you have both entrusted me with this wonderful resource and sincerely thank you for the light hearted humour you have both provided. I am overjoyed that my research journey continues here and I look forward to the progress we can make in the cancer genetics field. A very big thankyou to JM for keeping me sane over the last five years. James, you seriously have taught me all that you know and I'll be forever thankful for your lab skills and dad jokes. Thank you to my other co-supervisors, Dr Adele Holloway and Prof. Kathryn Burdon for your support and knowledge. A big thank you also to our collaborators, Dr Blackburn NB., Dr Thomson RJ. and Dr Blizzard L. for your statistical knowledge; to Dr Field MA., Dr Blackburn NB., Ms Polanowski A. and Dr Charlesworth JC. for your sequencing expertise; Dr Donovan S. and Dr Malley R for sourcing and reviewing all of the pathology material used in this study and an extra special mention to Ms. Narelle Phillips for her histology knowledge and helpful tips.

Secondly, a PhD shapes the person you want to be. The past four years have changed me for the better and I am truly thankful for all of my past and present friendship and support networks. To my fellow PhD friends, particularly those who have come and gone in 502, you guys are what makes the PhD journey easier. Thank you for the coffee dates, the Friday lunches and the supply of birthday cakes and treats. Thank you to my family for your never-ending support. You have always been big fans of whatever I choose to do and I thank you for shaping me to be the ambitious person I am today. To Josh, we have only known each other for a short time, but I know you'll be there throughout the most important years of my career. Thank you for pushing me to be my best, for supporting all of my endeavors, for providing me with all of the wine and for giving me all of the belly laughs and love any person would be envious of.

ABSTRACT

Prostate cancer (PCa) is the most common, non-cutaneous malignancy in men in the developed world. It is highly heritable, with twin studies suggesting that as much as 58% of disease risk can be explained by genetics. While more than 170 common genetic risk variants have been identified, these variants still only explain a minor portion of heritability, are largely of low to moderate effect size, and for many their function remains unclear. There has recently been significant success in the discovery of rare genetic variants contributing to complex disease through next-generation sequencing studies of large families. Mancuso and colleagues (2016) have estimated that as much as 42% of PCa risk is due to rare variants, but to date only 6% of this risk has been elucidated. With two-thirds of PCa heritability still unexplained, including the contribution of rare variants, we hypothesise that the utilisation of PCa families will aid in the identification of these rare variants.

Germline risk variants and somatic tumour alterations have traditionally been regarded as unrelated events in cancer. However, there is now increasing evidence to suggest that specific germline variants may predispose some somatic tumour events, including copy number changes and gene fusions. Of particular interest in PCa, is the fact that germline variants have been reported to be significantly associated with the *TMPRSS2:ERG* fusion. Given the high frequency of these fusion events and accumulating evidence from previous studies, we also hypothesise that there are inherited determinants of somatic tumour variation, and this will be the second focus of this thesis.

Family studies are proving highly valuable in the study of complex disease and here I will explore these hypotheses using the *Tasmanian Familial Prostate Cancer Study* cohorts, comprising genetic material from large families with multiple PCa cases and their relatives (*Tasmanian Familial Prostate Cancer Cohort*), as well as the *Tasmanian Prostate Cancer Case-Control Study*.

To address the first hypothesis, whole-genome sequencing (WGS) was undertaken in five large Tasmanian PCa pedigrees to identify rare genetic variants contributing to disease risk. Variants were prioritised on a per-family basis by minor allele frequency, segregation with disease, mutation type and predicted functional consequence. Of the 20 prioritised rare variants, four

were determined to be significantly associated with PCa risk in the Tasmanian population. This included rare variants in the genes *RND1, WNT1, EZH2* and the known G84E *HOXB13* variant. Both *RND1* and *WNT1* have been found to promote the growth and migration of cancer cells and, notably, in our study the variants appeared to be co-inherited.

The *EZH2* variant is a rare, intronic variant (rs78589034) present within a 3' splice consensus sequence. *EZH2* encodes the *histone methyltransferase* enzyme and is constitutively overexpressed in a range of cancers, including PCa. *EZH2* is a highly variable gene and multiple transcripts have been identified. In fact, Chen *et al* (2017) observed that alternative splicing involving the inclusion of exon 14 plays a major role in the tumourigenesis of renal cancer. While this variant was significantly associated with PCa risk in the Tasmanian population (OR=3.27, p=0.001), functional assays were unable to determine the potential impact of this variant on the splicing mechanisms of *EZH2*.

The G84E *HOXB13* variant (rs138213197) was initially observed in the WGS data and follow-up genotyping found a significant association with PCa risk in the larger *Tasmanian Familial Prostate Cancer Study* cohorts (OR=6.59, p=4.22×10^{-5}). Although multiple studies have demonstrated an association of the G84E variant with PCa risk, no study has assessed the functional impact of the variant on *HOXB13* gene and protein expression. Here, no difference in *HOXB13* gene or protein expression was observed between prostate tumours from G84E carriers and non-carriers, but interestingly, the variant allele was rarely transcribed in carriers. The unbalanced allele transcription did not appear to be caused by methylation differences and, thus, other mechanisms, such as DNA copy number variation at the *HOXB13* site or rapid targeted degradation of the variant mRNA transcript, may underpin the observed allelic imbalance. Hence, questions remain regarding how this variant influences tumour development. Given the rarity of the G84E variant, achieving a sufficient sample size for analyses is challenging, therefore, through collaboration with members of the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium, we aim to further explore the function of this variant.

To address the second hypothesis, germline and tumour samples from PCa cases were utilised to explore inherited determinants of somatic tumour variation. Tumours from 14 PcTas9 cases were analysed using the TruSight RNA Fusion Panel (Illumina), identifying seven tumours as *TMPRSS2:ERG* fusion positive. Subsequently, analysis of the entire *Tasmanian Prostate*

Tissue Pathology Resource showed that 31.5% of tumours were fusion positive. This event was more frequent in tumours from two families, PcTas2 and PcTas9 and, interestingly, was not identified in any of the eight sporadic tumours examined. These results suggest that there may be an underlying inherited genetic variant(s) predisposing to this fusion event. Subsequent work is focusing on screening for germline risk variants previously found to be associated with fusion positive tumours, including rare variants in *POLI* and *ESCO1*.

Somatic copy number changes, including amplifications and deletions, are also common events in tumours, leading to the suggestion that they may also arise due to germline genetic variation. To explore this hypothesis, array comparative genomic hybridisation was applied to 12 PcTas9 prostate tumours to determine shared altered chromosomal regions. The most consistent alteration involved amplification of the *EEF2* gene, which is a novel finding. EEF2 is highly expressed in human carcinoma tissue and has been suggested as a potential PCa biomarker. Immunohistochemistry of the *Tasmanian Prostate Tissue Pathology Resource* found that the EEF2 protein was overexpressed in 49% of malignant compared to matched benign tissue, but no difference was observed between tumours from PcTas9 cases and non-PcTas9 cases. However, gene expression assays found malignant cells from PcTas9 tumours had significantly higher *EEF2* 5'UTR/exon 2 expression compared to malignant cells isolated from non-PcTas9 tumours. Thus, these results suggest that the *EEF2* amplification may be specific to PcTas9 and due to an inherited predisposition variant(s). To test this hypothesis, recent WGS data generated for this family will be utilised in linkage analysis based on *EEF2* amplification status.

Establishing rare variants as disease-causing requires analysis of large cohorts and secondly, comprehensive functional analyses. This study has identified four rare germline variants significantly associated with PCa risk in the Tasmanian population. Variant screening in larger cohorts of PCa cases and controls is required to determine their contribution to other populations. Moreover, the functional impact of the *EZH2* and *HOXB13* variants on gene and protein expression remains unclear and requires more comprehensive functional analyses. This study also identified recurrent somatic variations in the tumour genomes of Tasmanian PCa cases. The *TMPRSS2:ERG* fusion and amplification of the *EEF2* gene is more apparent in tumours from the PcTas9 family, suggesting that these somatic tumour events could be underpinned by inherited predisposition.

There is currently a strong push to implement polygenic risk scores based on common variants in the clinical setting, yet with only one-third of genetic predisposition explained, clinical implementation may be premature. Studies such as the one described here, aim to directly explore genetic contribution to PCa. Rare germline variants and somatic tumour variation are of great interest as potential screening biomarkers and therapeutic targets, and if we are to understand the genetic determinants of PCa development, a strong focus on fully characterising these factors is essential.

TABLE OF CONTENTS

TABLES

FIGURES

ABBREVIATIONS

22Rv1	Human prostate carcinoma epithelial cell line
5'UTR	five prime untranslated region
95% CI	95% confidence interval
ABL1	abelson murine leukemia viral oncogene homolog 1 [*gene*]
aCGH	array Comparative Genomic Hybridisation
ADT	Androgen deprivation therapy
AMACR	alpha-methylacyl-CoA racemase [*gene*]
AR	androgen receptor [*gene*]
ASR	age-standardised rate
ATBF1	zinc finger homeobox protein 3 [gene-*ZFHX3*]
ATM	ataxia telangiectasia mutated [*gene*]
AUS	Australia
*B*actin	beta actin [*gene*]
BCF	BIM Collaboration Format
BCR	breakpoint cluster region protein [*gene*]
BDT	BigDye® Terminator (ThermoFisher Scientific)
bp	base pair
BP	boiling point
BPH	benign prostatic hyperplasia
BRCA1	breast cancer 1 [*gene*]
BRCA2	breast cancer 2 [*gene*]
BROCA	Breast and ovarian cancer associated (gene panel)
BTNL2	butyrophilin like 2 [*gene*]
C19orf48	chromosome 19 open reading frame 48 [*gene*]
CADD	Combined Annotation Dependent Depletion (*In silico* functional prediction tool)
CCL26	C-C motif chemokine ligand 26 [*gene*]
CDH1	cadherin-1 [*gene*]
cDNA	complementary DNA
cells/mL	cells per millilitre
CGH	Comparative Genomic Hybridisation
CHAD	chondroadherin [*gene*]
CHEK2	checkpoint kinase 2 [*gene*]

cis-eQTL	local quantitative trait loci (SNPs that reside within 1Mb of the transcription start site)
CLDN4	claudin 4 [*gene*]
CNKSR3	connector enhancer of kinase suppressor of ras 3 [*gene*]
CNV	copy number variation
CO_2	carbon dioxide
COSMIC	Catalogue of Somatic Mutations in Cancer (Welcome Trust Sanger Institute, UK)
CpG	A region of DNA where a cytosine nucleotide is followed by a guanine nucleotide
CRIP2	cysteine rich protein 2 [*gene*]
DAB+	3,3'-diaminobenzidine
DAPK3	death-associated protein kinase 3 [*gene*]
ddPCR	Droplet Digital™ PCR (BioRad Technologies)
df	degrees of freedom
dH_2O	distilled water
DLR	derivative log ratio
DNA	deoxyribonucleic acid
DNMT36	DNA methyltransferase 36
DPH1	diphthamide biosynthesis 1 [*gene*]
DRE	digital rectal exam
dsDNA	double stranded DNA
DVA	Unique identification number for individuals in the *Tasmanian Prostate Cancer Case Control Study* (described in Chapter 2.1.3)
EBRT	External beam radiation therapy
EEF2	eukaryotic elongation factor 2 [*gene*]
ELAC2	elac ribonuclease z 2 [*gene*]
EPS8	epidermal growth factor receptor pathway substrate 8 [*gene*]
ERBB2	erythroblastic oncogene B [*gene*]
ERG	ETS-related gene [*gene*]
ESCO1	establishment of sister chromatid cohesion N-acetyltransferase 1 [*gene*]
ETS	erythroblast transformation specific (family of transcription factors) [*gene*]
ETV1	ETS variant 1 [*gene*]
ETV4	ETS variant 4 [*gene*]
ETV5	ETS variant 5 [*gene*]

ExAC	Exome Aggregation Consortium, non-Finnish European, non-The Cancer Genome Atlas
EZH2	enhancer of zeste homolog 2 [*gene*]
FFPE	formalin fixed paraffin embedded
FHCRC	Fred Hutchinson Cancer Research Center (Seattle, USA)
FISH	Fluorescence *in-situ* hybridisation
FLI1	friend leukemia integration 1 [*gene*]
FOXA1	forkhead box protein A1 [*gene*]
FOXA2	forkhead box protein A2 [*gene*]
FOXP1	forkhead box protein P1 [*gene*]
FPKM	fragments per kilobase of exon model per million reads mapped
GAPDH	glyceraldehyde 3-phosphate dehydrogenase [*gene*]
GATK	The Genome Analysis Toolkit
gDNA	genomic DNA
GS	Gleason score
GSA	Global Screening Array
GTEx	Genotype-Tissue Expression Project (National Cancer Institute, USA)
GWAS	genome-wide association study
H&E	haematoxylin & eosin
H3K27me3	tri-methylation at the 27th lysine residue of the histone H3 protein
HDAC9	histone deacetylase 9 [*gene*]
HEK293	human embryonic kidney 293 (cell line)
HMTase	histone methyltransferase
HNRNPCL	heterogenous nuclear ribonucleoprotein L [*gene*] (family of genes)
HOXA9	homeobox A9 [*gene*]
HOXB13	homeobox B13 [*gene*]
HOXB13-AS1	HOXB13 antisense RNA 1 (long non-coding RNA)
HRP	horse radish peroxidase
HSD3B1	hydroxy-delta-5-steroid dehydrogenase, 3 beta-and steroid delta-isomerase 1 [*gene*]
ICA1	islet cell autoantigen 1 [*gene*]
ICPCG	International Consortium of Prostate Cancer Genetics
IgG$_1$	Immunoglobulin G subclass 1 (Isotype control, Dako)
IHC	immunohistochemistry
INDEL	insertion/deletion
IRS1	insulin receptor substrate 1 [*gene*]

ITGAD	integrin subunit alpha D [*gene*]	
JARID2	jumonji and AT-rich interacting domain containing 2 [*gene*]	
KMT2C	lysine methyltransferase 2C [gene]	
LB	lysogeny broth	
LD	linkage disequilibrium	
LNCaP	metastatic prostate cancer of the lymph node (cell line)	
M/PD	moderately-poorly differentiated (Gleason score)	
MAF	minor allele frequency	
Mb	mega base	
MD	moderately differentiated (Gleason score)	
MEIS	meis homeobox 1 [*gene*]	
MLE	maximum likelihood estimation	
MQLS	an association test that can combine familial and case-control genotyping data [2]	
MSMB	microseminoprotein beta [*gene*]	
MSR1	macrophage scavenger receptor 1 [*gene*]	
MYC	myelocytomatosis oncogene [*gene*]	
NAT10	N-acetyltransferase 10 [*gene*]	
NBN	nibrin [gene]	
NDE1	nude neurodevelopment protein 1 [*gene*]	
NEDD9	neural precursor cell expressed, developmentally downregulated 9 [*gene*]	
ng	nanogram	
ng/mL	nanograms per millilitre	
ng/mL/year	nanograms per millilitre per year	
ng/μL	nanograms per microlitre	
NGS	next-generation sequencing	
nM	nanomolar	
NOTCH3	notch receptor 3 [*gene*]	
NSB1	nibrin [gene-*NBN*]	
nt	nucleotide	
NTC	no template control	
OR	odds ratio	
OR5H14	olfactory receptor family 5 subfamily H member 14 [*gene*]	
p arm	short chromosomal arm	
P2RX7	purinergic receptor P2X 7 [*gene*]	
PALB2	partner and localiser of BRCA2 [*gene*]	

PARP	poly (ADP-ribose) polymerase [*gene*]
PC	unique identification number for individuals in the *Tasmanian Familial Prostate Cancer Cohort* (described in Chapter 2.1.2)
PC3	metastatic prostate cancer of the bone marrow (cell line)
PCa	prostate cancer
PcG	polycomb group (of proteins)
PCR	polymerase chain reaction
PcTas	family apart of the *Tasmanian Familial Prostate Cancer Cohort*
PD	poorly differentiated (Gleason score)
POLI	DNA polymerase iota [*gene*]
PolyPhen2	Polymorphism Phenotyping v2 (*In silico* functional prediction tool)
PRACTICAL	Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome consortium
PRAME	preferentially expressed antigen in melanoma [*gene*] (family of genes)
PRC1	polycomb repressive complex 1 (class of PcG proteins)
PRC2	polycomb repressive complex 2 (class of PcG proteins)
PSA	prostate-specific antigen
PT	unique identification number of the prostate needle biopsies analysed in Chapter 4 (described in Chapter 2.1.6)
PTEN	phosphatase and tensin homolog [*gene*]
q arm	long chromosomal arm
QC	quality control
r2	R squared
RAD51C	RAD51 paralog C [*gene*]
RB1	retinoblastoma protein 1 [*gene*]
RBFOX1	RNA binding FOX-1 homolog 1 [*gene*]
RHPN2	rhophilin rho GTPase binding protein 2 [*gene*]
RNA	ribonucleic acid
RNASEL (HPC1)	ribonuclease L (Hereditary prostate cancer 1) [gene]
RND1	rho family GTPase 1 [*gene*]
RP	radical prostatectomy
RR	relative risk
RSR	relative survival rate
RT-qPCR	quantitative reverse transcription PCR
RUNX1	runt-related transcription factor 1 [*gene*]

RUNX1T1	runt-related transcription factor 1 partner transcriptional co-repressor 1 [*gene*]
RYBP	RING1 and YY1 binding protein [*gene*]
SAM	Sequence alignment/map
SF3B1	splicing factor 3b subunit 1 [*gene*]
SF3B3	splicing factor 3b subunit 3 [*gene*]
SIFT	Sorting Intolerant from Tolerant (*In silico* functional prediction tool)
SLC30A4	solute carrier family 30 member 4 [*gene*]
SLC45A3	solute carrier family 45 member 3 [*gene*]
SNAIL	zinc finger protein [*gene*]
SNP	single nucleotide polymorphism
SNORD37	small nucleolar RNA, C/D box 37 [*gene*]
SORL1	sortilin-related receptor 1 [*gene*]
SOXA9	SRY-box transcription factor 9 [*gene*]
Sp6	sp6 transcription factor [*gene*]
SPINK1	serine peptidase inhibitor, kasal type 1 [*gene*]
SPOP	speckle type BTB/POZ Protein [*gene*]
SSH3	slingshot protein phosphatase 3 [*gene*]
T1E2	exon 1 of *TMRPSS2* fused to exon 2 of *ERG* (gene fusion)
T1E4	exon 1 of *TMRPSS2* fused to exon 4 of *ERG* (gene fusion)
TANGO2	transport and golgi organisation 2 [*gene*]
TAS	Tasmania
TCGA	The Cancer Genome Atlas (The National Cancer Institute, USA)
TCR	Tasmanian Cancer Registry
TIA1	TIA1 cytotoxic granule associated RNA binding protein [*gene*]
TINCR	TINCR ubiquitin domain containing [*gene*]
TMPRSS2	transmembrane serine protease 2 [*gene*]
TMPRSS2:ERG	*TMPRSS2* fused to *ERG* (fusion gene)
TP53	tumour protein 53 [*gene*]
TPM	transcripts per million
TREM2	triggering receptor expressed on myeloid cells 2 [*gene*]
TRUS	transrectal ultrasound biopsy
TURP	transrectal resection of the prostate
U2AF1	U2 small nuclear RNA auxiliary factor 1 [*gene*]
uF	microfarad
ug	microgram

uL	microlitre
uM	micromolar
USA	United States of America
UV	ultraviolet
VASP	Variant Analysis of Sequenced Pedigrees [3,4]
WD	well differentiated (Gleason score)
W/MD	well-moderately differentiated (Gleason score)
WES	whole-exome sequencing
WGS	whole-genome sequencing
WHSC1L1	wolf Hirschhorn syndrome candidate 1-like protein 1 [*gene*]
WNT1	WNT family member 1 [*gene*]
ZBTB7A	zinc finger and BTB domain containing 7A [*gene*]

CHAPTER 1 : **INTRODUCTION**

1.1   AN INTRODUCTION TO PROSTATE CANCER

Prostate cancer (PCa) is the most common, non-cutaneous malignancy in men in the developed world, with approximately 1.3 million men diagnosed in 2018, worldwide [5]. In Australia it is also the most common cancer in men, with 19,508 new cases expected to be diagnosed this year, and it is the second leading cause of male cancer-related deaths [6]. In Tasmania, an island state of Australia, PCa was the most commonly diagnosed cancer in males in 2016, however it was the third most common cause of male cancer-related deaths [7], suggesting a large proportion of men are diagnosed with indolent disease. Current diagnostic techniques, including the prostate-specific antigen (PSA) test cannot distinguish indolent *versus* aggressive PCa. Lack in specificity led to the over-diagnosis and treatment of indolent PCa in the early 1990's, which was associated with a spike in serious complications as a result of over-treatment. It is now known that PCa has a strong genetic component (58%) and over 170 common variants have been identified that explain approximately one third of this known genetic risk [8]. However, the underlying mechanism by which these common variants confer risk remains unclear [9]. Numerous genome-wide association studies (GWAS), comprising tens of thousands of individuals, have likely identified the majority of common risk variants. Therefore, it has been hypothesised that some of the 'missing' heritability is likely due to rare genetic variants [10]. The recent interest in rare genetic variants has led to a renewed focus on using family pedigrees for gene discovery. This is because reduced genetic complexity means rare variants are enriched in these families, which reduces the challenges normally associated with the search for disease-causing rare variants. The decrease in cost of next-generation sequencing (NGS) in recent years has also permitted whole-genome sequencing (WGS) to emerge as a useful tool in the identification of such rare variants [11-14].

Incidence rates of PCa vary greatly between populations worldwide. For example, Australia and New Zealand's age-standardised incidence rate (ASR) in 2018 was 85.6 per 100,000, compared to South-Central Asia with only approximately 5.0 PCa cases per 100,000 men [5]. These two regions represent the world's highest and lowest PCa incidence rates, respectively. More than 70% of cases recorded in 2018 (893,274) were in more developed regions of the world (extracted from GLOBOCAN 2018; the most recent comprehensive worldwide study of cancer in the adult population (Figure 1.1; [5]). The rates of PCa are highest in more developed

countries, such as Australia, New Zealand, Northern and Western Europe and North America [5,15], and these high incidence rates are likely due to readily available healthcare which drives the rate of diagnosis. This includes the practice of PSA testing and subsequent biopsy, which has resulted in the detection of clinically insignificant (indolent) disease [16]. Mortality rates are however reversed, with the number of estimated deaths from PCa in 2018 being greater in less developed regions (27.9 ASR per 100,00; South Africa) compared to more developed countries (10.0 ASR per 100,000; Australia) [5]. This is likely due to a lack of available health care; including screening and prevention strategies, as well as access to treatment [17]. In terms of survival rates, in developed countries these have increased in patients with localised PCa. For example, in Australia a 5-year relative-survival rate (RSR) of 59.2% was recorded between 1986-1990, compared to 95.2% between 2011-2015, according to the Australian Institute of Health and Welfare data [6]. Although, this is dramatically reduced in men with advanced metastatic PCa, with a 5-year RSR of only 29% [18]. Overall, there is great disparity in incidence, mortality and survival rates worldwide, due to access to healthcare, however within populations these rates can fluctuate too and they may be in part mainly attributed to genetic differences.
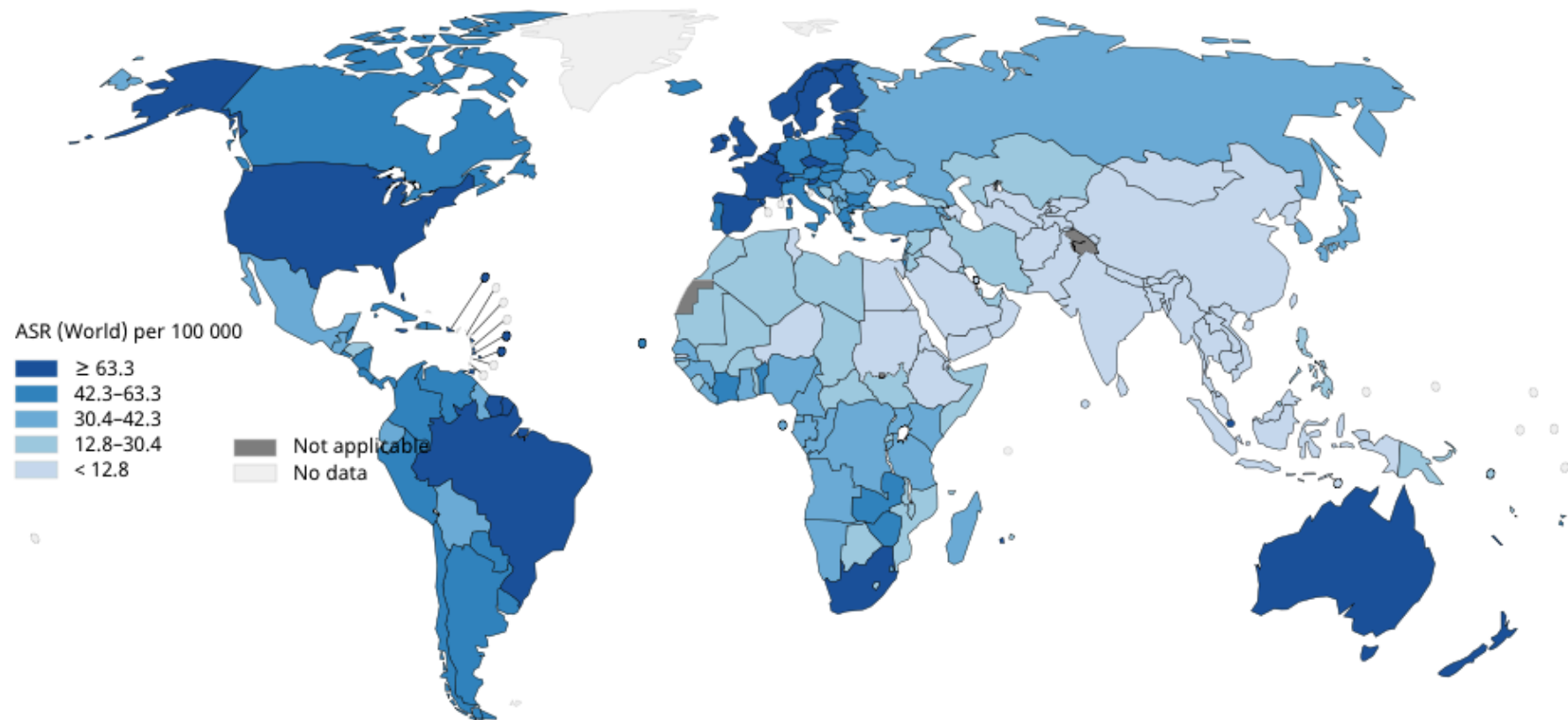
Figure 1.1 Prostate cancer incidence rates worldwide in 2018 (age-standardised incidence rate per 100,000 men).

Presented here is the age-standardised incidence rate per 100,000 men in each country. Higher incidence rates are more prominent in developed countries of the world, such as Australia, Northern and Western Europe and North America, where readily available healthcare drives the rate of diagnosis [5,19].

1.2   AN INTRODUCTION TO THE PROSTATE

The prostate, the largest male accessory gland, is located in front of the rectum, below the bladder, surrounding the urethra [20] (Figure 1.2). The prostate plays an important role in male reproduction; it is a small exocrine gland that produces a fluid containing enzymes, lipids, amines and metal ions that comprise part of the semen. This fluid is essential for the normal function of spermatozoa and is stored with the sperm in the seminal vesicles until ejaculation. As well as its role in the male reproductive system, the prostate participates in the control of urine output from the bladder [20].
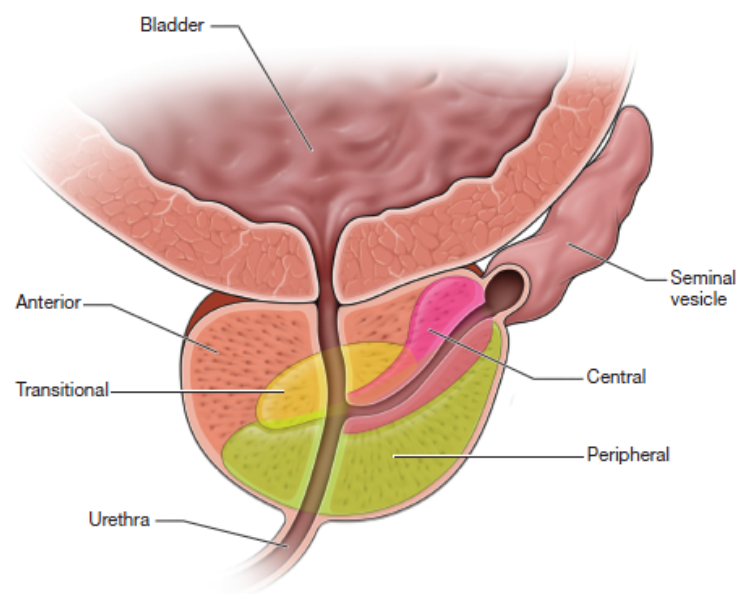
Figure 1.2 The anatomical location and zones of the prostate.

The prostate is located below the bladder and encompasses the urethra. It is composed of four zones, the peripheral (green), anterior (cream), central (pink) and transitional (yellow). Approximately 60-70% of all prostate cancer tumours arise in the peripheral zone and about 10-20% in the transitional zone [21].

1.2.1   The development of the normal prostate

The development of the male reproductive tract, including the prostate is dependent upon mesenchymal-epithelial interactions and fetal androgens [22]. Androgens, such as testosterone, dihydrotestosterone and androgen receptor (AR) play an important role in the development and maintenance of the prostate. The AR serves as an essential survival factor for prostate epithelial cells (reviewed in Davey *et al.* (2016) [23]). The normal adult prostate is composed of a glandular

epithelial and a fibromuscular stroma component. The glandular epithelium constitutes approximately 95% of the prostate and it is composed of a large peripheral zone and a small central zone. The remaining 5% is composed of the transitional zone and the peri-urethral glands [21] (Figure 1.2).

1.2.2   The development of prostate cancer

There are two different types of prostatic disease in adult males, benign prostatic hyperplasia (BPH) and PCa. BPH is a common, benign condition that involves the enlargement of the prostate, which can restrict the flow of urine and cause pain during urination [24]. BPH is non-life threatening and is neither a premalignant lesion nor a precursor of PCa. PCa is the uncontrolled division of prostate cells, with approximately 60-70% of all tumours arising in the peripheral zone and about 10-20% in the transitional zone [25] (Figure 1.2). PCa that arise in the peripheral zone retain some glandular structure, which classifies them as adenocarcinomas. The question remains how dysregulation of normal prostate development and maintenance leads to the initiation of cancer, however, it is believed that disruption of normal AR-regulated gene expression plays a vital role [26].

1.3   PROSTATE CANCER DIAGNOSIS

Currently screening guidelines for PCa recommend that men over 50 years of age, or men over 40 with a family history of PCa discuss testing with their doctor. There are usually no initial or early symptoms of PCa, but men experiencing symptoms, including a frequent or sudden urge to urinate, difficulty urinating (including discomfort and/or blood in their urine), lower back or pelvic pain and fatigue, require follow-up investigation. PCa is commonly diagnosed through a physical exam, such as a digital rectal exam (DRE) and a blood test to assess PSA level [27,28]. If both of these tests are indicative of PCa (enlarged prostate and an increased PSA level) an ultrasound-guided biopsy is undertaken.

1.3.1   Prostate-specific antigen testing

PSA is a glycoprotein produced by the epithelial cells of the prostate gland that can be detected in the blood (it is prostate-specific, but not prostate cancer-specific). There is no normal or abnormal PSA level for a male, but a higher reading may indicate the presence of cancer; as the prostate lumen and capillaries are disrupted and PSA is released into the serum [29]. Whilst most men who are disease-free have a PSA level under 4ng/mL, about 15% of men with a PSA

level under 4ng/mL will have disease on biopsy [30]. Conversely, a PSA level greater than 4ng/mL is not diagnostic for PCa, as common benign conditions, such as BPH also increase PSA levels [31]. Numerous strategies have been proposed to improve the diagnostic performance of PSA testing. This includes age- and race-specific reference ranges, and measuring PSA velocity, which is the rate of change of a man's PSA level [32]. The Baltimore Longitudinal Study of Aging found that men with a PSA velocity greater than 0.75ng/mL/year were at an increased risk of being diagnosed with PCa [33]. Carter *et al.* (2004) also concluded that PSA velocity was more specific than a 4ng/mL cutoff (90% *versus* 60% specificity), however, subsequent randomised trials suggested that PSA velocity adds little predictive information to total PSA. Thus, the 4ng/mL cutoff remains the gold standard for PCa screening because it balances the tradeoff between missing important cancers at a curable stage (about 15%) and avoiding both detection of clinically insignificant disease and subjecting men to unnecessary biopsies [27,32,34].

PSA testing was widely adopted for PCa screening in the early 1990s and subsequently, led to a dramatic increase in the incidence in developed countries. For example, following the over-implementation of PSA testing in Australia, the ASR peaked at 79.7 in 1994, compared to 42.0 just four years prior in 1990, a trend that was also evident in Tasmania [7,35] (Figure 1.3). A large proportion of these diagnoses included tumours that were insignificant and without PSA testing may not have presented clinically. Implementation of PSA testing saw many men with indolent PCa undergo invasive biopsies and radiotherapy, often with complications arising that were more severe than their original tumour [36,37]. This left clinicians and scientists questioning whether PSA was an appropriate tool for PCa diagnosis. As a result, clinical guidelines for the screening of PCa using PSA were revised in 2016 by a multidisciplinary expert advisory group under the leadership of the Prostate Cancer Foundation of Australia, and approved by the National Health and Medical Research Council [38]. Adoption of these new guidelines saw a 10% reduction in the number of PCa cases diagnosed in Australia in 2018 (18,274) compared to 2012 (20,065) and this reduction was also apparent in Tasmania [6,39].
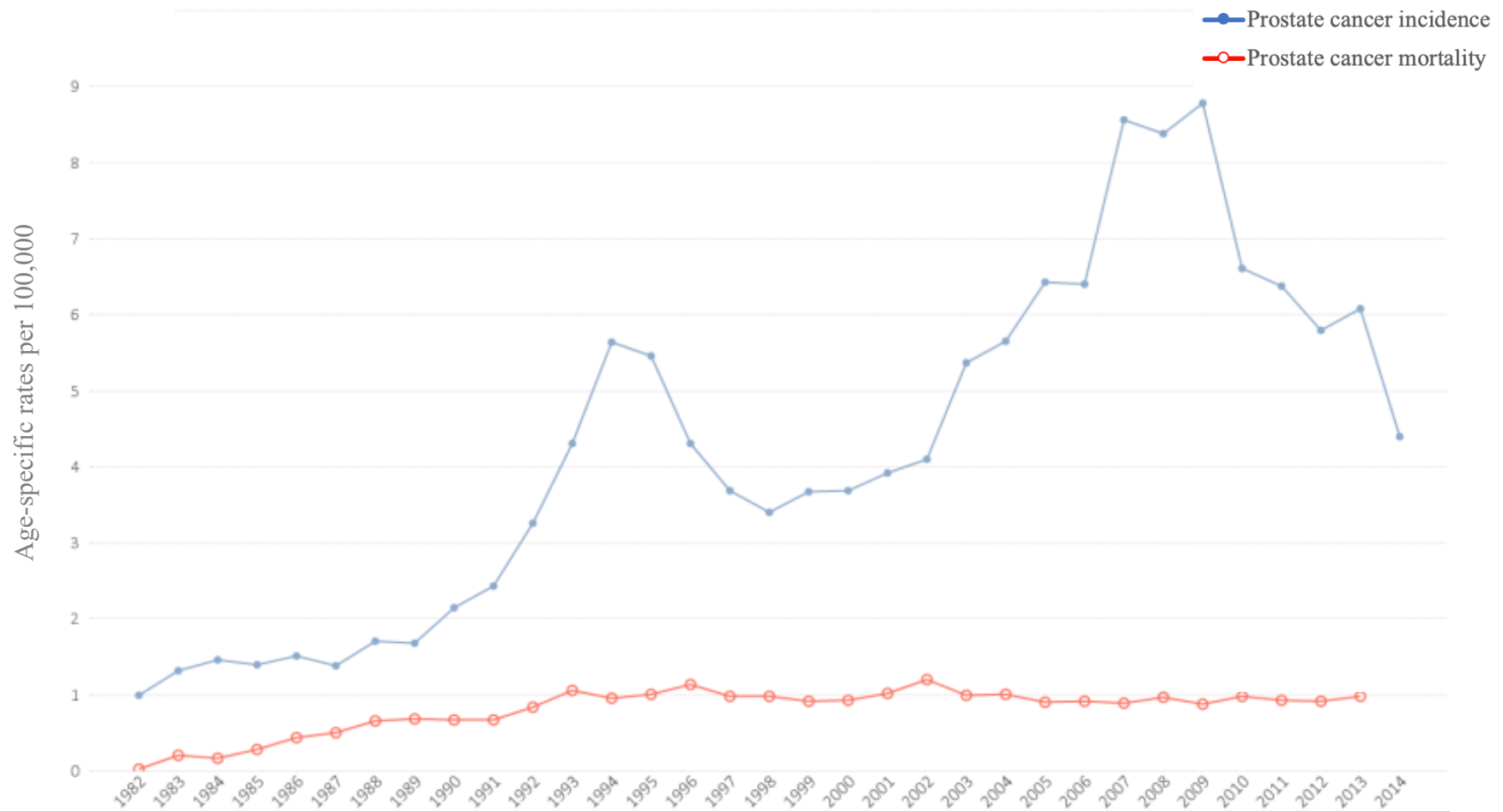
Figure 1.3 Trends in incidence and mortality of prostate cancer in Tasmania (age-specific rates per 100,000 men).

Presented here is the age-specific incidence and mortality rates per 100,000 men, in Tasmania between 1982 and 2014. In the early 1990's there was a sharp increase in PCa incidence following the over-implementation of PSA testing [7]. Age-specific rates are calculated by dividing the number of cases occurring in each specified age group by the corresponding population in the same age group.

1.3.2   Gleason scoring system

If a man has an enlarged prostate and an 'abnormal' PSA level, a urologist will perform an ultrasound-guided biopsy. This involves the removal of a small sample of prostate in a fine needle and it is the only definitive way to diagnose PCa. This sample is stained with haematoxylin and eosin (H&E) and is microscopically visualised by a histopathologist, who determines if there are any regions of malignancy. If malignant cells are present the histopathologist scores the tumour using the Gleason scoring system. A Gleason score (GS) is an evaluation of the ability of the prostate to form regular uniform glands and this score reflects the aggressiveness of a tumour, and often guides subsequent treatment. It is calculated by the addition of the most common and second most common pattern of cancer. Each pattern is graded as 1, 2, 3, 4 or 5; with 1 indicating small uniform glands (normal) and 5 representing occasional gland formation (cancerous) [40]. A GS ≤6 is considered to be an indicator of less aggressive disease with a good prognosis, as it is the most well-differentiated tumour grade. A GS of 7 (3+4) is similar, however the density of malignant glands and the tumours invasive potential is increased. Whereas, a GS of 7 (4+3) shows a clearly infiltrative tumour. Regardless, tumours with a GS ≤7 are, in most cases, curable. A GS of 9-10 indicates no glandular differentiation and these tumours tend to be advanced neoplasms, that are unlikely to be cured and have a high likelihood of metastasising [40].

1.3.3   Molecular subclassification to predict patient outcomes

Molecular profiling of prostate tumours has been undertaken with the aim of identifying early genomic alterations that may assist in the clinical setting. Prostate tumour samples in The Cancer Genome Atlas (TCGA) have been utilised to identify specific molecular subclasses of localised PCa, and these are largely mutually exclusive. Two major molecular subclasses of localised PCa are *ETS*-fusion positive and negative tumours, and both of these subclasses can be further subdivided as follows:

1. *ETS*-fusion positive (including overexpression of ERG, ETV1, ETV4, ETV5 and FLI1). Approximately 50% of all prostate tumours are *ETS*-fusion positive [41]. These tumours can be further subdivided into the following:
   - *ETS*-fusion positive with loss of PTEN. A study by Bismar *et al.* (2018) found 21.8% of 463 tumour samples had both loss of PTEN and gain of ERG

(p<0.001) [42]. PTEN loss is found in localised PCa but is much more common in advanced, metastatic disease [43,44].

- *ETS*-fusion positive with genetic alterations, such as RB1 loss (28%), amplification of MYC (10%) and mutations in *ATM* (19%) and *BRCA2* (~7%) [43,45].

- *ETS*-fusion positive with loss of function mutations in *TP53* occur in 40-60% of cases. Like PTEN loss, *TP53* mutations are found in localised PCa, but are more common in advanced, metastatic disease [43,44].

2. *ETS*-fusion negative tumours [41], which can be further subdivided into the following:
    - *ETS*-fusion negative with recurrent *SPOP* mutations. These mutations are the most common point mutations in PCa, occurring in 6-15% of cases [46].
        o *ETS*-fusion negative with homozygous deletion of CDH1 occurs most commonly in the *SPOP* mutant subclass. Overall, this subclass occurs in 5-10% of PCa cases; 80% of which belong to the mutant *SPOP* subclass. This subclass is more common in advanced, metastatic disease [46,47].
    - *ETS*-fusion negative with missense *FOXA1* mutations. This subtype has been identified in 4% of the TCGA cases, and are mostly mutually exclusive of *ETS*-fusion positive and mutant *SPOP* tumours [43,46].
    - *ETS*-fusion negative with SPINK1 overexpression. SPINK1 is overexpressed in 5-10% of PCa and is associated with aggressive disease [43,48,49].

These PCa subtypes remain under investigation, as the acquisition of these changes in tumour development and their predicted value for prognosis and treatment remains unclear. However, the potential clinical utility of such classification tools could prove invaluable to predict PCa progression, aggressiveness and response to treatment [50,51]. Overall, it is apparent that each subclass of PCa is predisposed to its own defined set of progression events. However, some of these later events, such as loss of PTEN, loss of function mutations in *TP53*, mutations in *ATM*, deletion of CDH1 and overexpression of SPINK1, co-occur in different subclasses throughout PCa progression to the metastatic stage [46]. There is evidence to suggest that acquisition of genetic changes in tumours is not random, and inherited genetic variants may predispose to some acquired changes, which will be discussed in Chapters 6 and 7 of this thesis.

1.4 PRIMARY PROSTATE CANCER TREATMENT

1.4.1 'Active Surveillance'

PCa is clinically and biologically heterogenous and may remain present as indolent disease for many years. Autopsy studies have shown a high prevalence of clinically undetected PCa at time of death, with as many as 87% of men over 80 years of age found to have indications of PCa at the time of autopsy, suggesting that many men can live with indolent PCa [52,53]. 'Active surveillance' programs, or a 'watch and wait' approach to treatment is recommended for those with low grade disease, GS ≤6 [54,55].

1.4.2 Prostatectomy

A transurethral resection of the prostate (TURP) is a surgery used to treat urinary problems due to an enlarged prostate. It involves a prostate resection to relieve blockages in the urinary tract and is a treatment option for BPH. If a biopsy or TURP is suggestive of advanced PCa, GS ≥7, a radical prostatectomy (RP) is undertaken; the surgical removal of all of the prostate, part of the urethra and the seminal vesicles. Surgical castration can often result in nerve damage, loss of bladder control, impotence and infertility [56]. The RP tissues are histologically reviewed by a pathologist and scored using the Gleason scoring system, as well as the stage of disease, i.e. is it localised or has it progressed beyond the prostate.

1.4.3 Radiotherapy

Radiotherapy may be offered to men with early-stage PCa, and/or where surgery may be contraindicated. It is delivered externally using external beam radiation therapy (EBRT) or internally using brachytherapy. EBRT uses targeted radiation in the form of x-ray beams whereas brachytherapy involves the placement of the radiation source directly within the prostate, which limits the effects on nearby organs, such as the rectum and bladder [57]. Both EBRT and brachytherapy have similar side effects including impotence, changes in ejaculation, pain when urinating, blood in the urine, poor urine flow and bladder irritation. Studies have shown that radiotherapy is often associated with an increase in overall and PCa-specific mortality compared with surgical interventions [58].

1.4.4 Androgen deprivation therapy

Aggressive PCa cells require testosterone to grow, therefore slowing the production may slow the growth of the cancer or shrink it temporarily. Androgen-deprivation therapy (ADT) works

by blocking the body's production of testosterone. ADT injections are often used before, during and after radiotherapy and can slow the growth of a localised tumour for many years. Side effects can include fatigue, erection problems, loss of muscle strength, loss of bone density and increased risk of other problems such as, obesity, diabetes and heart disease [59]. ADT is the main treatment for advanced PCa, and can reduce or eliminate symptoms for months to years (reviewed in Abrahamsson *et al.* (2010) [60]). Concomitant chemotherapy is often used in parallel with ADT for advanced PCa and is sometimes the last resort for advanced cancers where ADT hasn't slowed tumour growth or relieved symptoms.

1.5   PROSTATE CANCER RISK FACTORS

On average, one in eight Australian men will be diagnosed with PCa before the age of 85 years, however, some men have a higher risk than others. Age, race and family history are the few established risk factors of PCa development. Like most other cancers, it is more common in older men, with 63% of cases diagnosed in men over 65 years of age [6]. Race, another risk factor, may explain some of the differences in incidence rates worldwide. Figure 1.1 shows that Asian men typically have the lowest PCa incidence rate, followed by Caucasian and African American men, respectively [15]. Indeed, African American men have a 60% higher incidence rate of PCa (275.3 per 100,000 men) than age-matched Caucasian populations (172.9 per 100,000 men) [61]. The higher rate of disease incidence and mortality among men of African descent in the United States and the Caribbean reflects the ethnic contribution to PCa development [62,63]. These studies also show that genetic factors, which underpin race, are an important determinant of the variation in risk and thus incidence at the population level. In fact, family history is the most consistently identified risk factor of PCa.

1.5.1   Prostate Cancer Heritability

Population-based cohort studies have frequently demonstrated a strong genetic component to PCa. Such studies have estimated that the risk for men with an affected first-degree relative is 2-3-fold higher than those without. This risk has been shown to increase up to 18-fold as the number of affected relatives and the relatedness of the affected case increases [64-67]. Further evidence of a genetic effect is shown by the observation that the relative risk (RR) to relatives increases as the age of the proband decreases [64,68-71]. Thus, a brother of a proband diagnosed with PCa at the age of 50, has a 1.9-fold higher risk of developing PCa compared with a brother of a man diagnosed with the disease at the age of 70 [64]. A meta-analysis of 33 epidemiological case-control and cohort-based studies, including over 12 million individuals and 27,000 PCa

cases, found that PCa risk appeared to be far greater for men with affected brothers (RR 3.14; 95% CI: 2.37-4.15) than for men with affected fathers (RR 2.35; 95% CI: 2.02-2.72) [72]. In an Italian study of 1,294 cases of PCa, risk was higher for men when the proband was younger, when two or more relatives were affected, and when the affected relative was a brother [73]. The increased RR between brothers compared to fathers is too large to be accounted for solely by an environmental effect, and therefore, a significant genetic component is implicated. Researchers have consistently identified a strong genetic component of PCa [74,75].

A Scandinavian study by Lichtenstein and colleagues (2000), reported that as much as 42% (95% CI, 29%-50%) of PCa risk can be explained by genetics [75]. However, a more recent study by Hjelmborg and colleagues (2014) of 30,054 dizygotic and 16,680 monozygotic male twin pairs, within the population-based registers of Denmark, Finland, Norway and Sweden, found that up to 58% (95% CI, 52%-63%) of PCa risk is heritable [74]. Previous studies have also shown that monozygotic twins have a 3- to 6-fold increased RR of developing PCa compared with dizygotic twins [64,68,70]. This finding is supported by Hjelmborg *et al.* (2014) who concluded that monozygotic twins have a 75% higher concordance for PCa than dizygotic pairs. Indeed, PCa is reported to have the highest degree of genetic transmission of any cancer (58%), followed by breast (13.6%) and colorectal cancer (12.8%) [74,75].

1.6 EARLY APPROACHES TO IDENTIFYING PROSTATE CANCER SUSCEPTIBILITY GENES

For decades, researchers have utilised families with a strong inheritance pattern of PCa in an effort to identify genetic variants that explain this heritability. One of the earlier approaches was segregation analyses, which take into consideration disease clustering, mode of inheritance, penetrance and estimated allele frequency of potential disease associated variants [64,76-78]. The first segregation analysis was conducted in 1992, of 740 familial probands who underwent RP. This study suggested an inherited predisposition of PCa and concluded that familial clustering of disease was due to a rare, highly penetrant variant. Carriers of the variant were predicted to have a cumulative risk of PCa development of 88% by the age of 85 years compared with risk of 5% for variant non-carriers [64]. Cui and colleagues (2001) evaluated genetic models in Australian pedigrees and modelled a rare variant that had a larger effect at younger ages [79]. This was supported by a Finnish study of 1,546 PCa families, in which a particular variant had a larger effect on men younger than 66 years of age [80]. Other segregation

studies have reached similar conclusions however, the identified variants were more common and only moderately penetrant [76,77]. The difference in allele frequency between studies may be explained by the genetic heterogeneity of PCa, in which multiple genes and modes of inheritance can be responsible for risk even within the one family [64,76,77,79,81-83].

Candidate-gene association studies have also featured strongly in the search for PCa susceptibility genes. These studies look for variants in genes that are involved in normal prostate development and/or other cancers, and compare the frequency of genetic variants in patients with PCa to individuals without disease. Notably, both breast and PCa tend to cluster within families. Therefore, given the known effect of *BRCA1* and *BRCA2* mutations in breast cancer, variants in these genes have been investigated in PCa cohorts. A study by Leongamornlert *et al.* (2012) found that deleterious *BRCA1* mutations confer a RR of PCa of ~3.75-fold (95% CI: 1.02-9.6) translating to a 8.6% cumulative risk by age 65 in their cohort of 913 cases aged between 36 and 86 years [84]. Examination of 1,864 PCa cases identified 19 protein-truncating mutations, three in-frame deletions and 69 missense variants in *BRCA2* and all were significantly associated with disease risk [85]. It was estimated that germline mutations in the *BRCA2* gene confer an increased PCa RR of 8.6-fold by the age of 65 years (95% CI: 5.1-12.6; [85]). Candidate-gene association studies have yielded several other interesting candidate genes, including the AR. The role of the AR in PCa is well known; the AR helps regulate prostate cellular proliferation and differentiation (reviewed in Montgomery *et al.* (2001) [86]). Plus, mutations in the AR enable PCa cells to grow even more rapidly. In fact, sequencing of the transcriptional network of the AR in PCa has highlighted novel mechanistic and functional insights in to how AR mutated cells gain a growth advantage (reviewed in Chng *et al.* (2013) [87]). However, lack of replication of some candidate gene associations, including *NBS1* [88], *CHEK2* [89] and *PALB2* [90] has limited their utility and has meant that these findings are somewhat unreliable [88,89,91,92].

Linkage analysis has proven a successful approach to gene discovery and is based on co-segregation of variants with disease in families, comparing the genotypes between PCa affected individuals and their unaffected relatives. Linkage analysis is based on the premise that known genetic markers in close proximity to the disease variant are inherited together with the disease trait. Linkage studies typically search for mutations that are rare in the population, are moderately to highly penetrant and have a large effect size (RR >2.0) [93]. Thus far, several candidate genes have been identified by linkage analysis and these regions are shown in Figure

1.4. The *RNASEL* (*HPC1*) gene at chromosome 1q25 is one of the most extensively researched genes identified by linkage analysis and it has been found to be associated with disease in families with five or more affected relatives, father to son transmission, a younger age of diagnosis and a higher GS [94]. Other genome-wide scans for linkage in PCa families have implicated 5p13, the chromosomal region of *AMACR* [95,96]. Replication studies have proven that overexpression of *AMACR* is an important marker of PCa and Zheng *et al.* (2002) identified four missense changes (M9V, G1175D, S291L and K227E) that had significantly different genotype frequencies between PCa cases and unaffected controls [97]. The *AMACR* gene variants, M9V and D175G have been identified in the Tasmanian PCa resource used in this study. In fact, both were found to be significantly associated with PCa risk, and whilst this association remained significant, it was diminished when relatedness amongst familial PCa cases was considered [98]. Conversely, evidence suggests that many of the other PCa genes identified through linkage studies, including *ELAC2* at 17p11 and *MSR1* at 8p21-23 [99], account for disease in only a small subset of families, which is consistent with the concept that PCa exhibits locus heterogeneity and that the identified variants are rare [100].
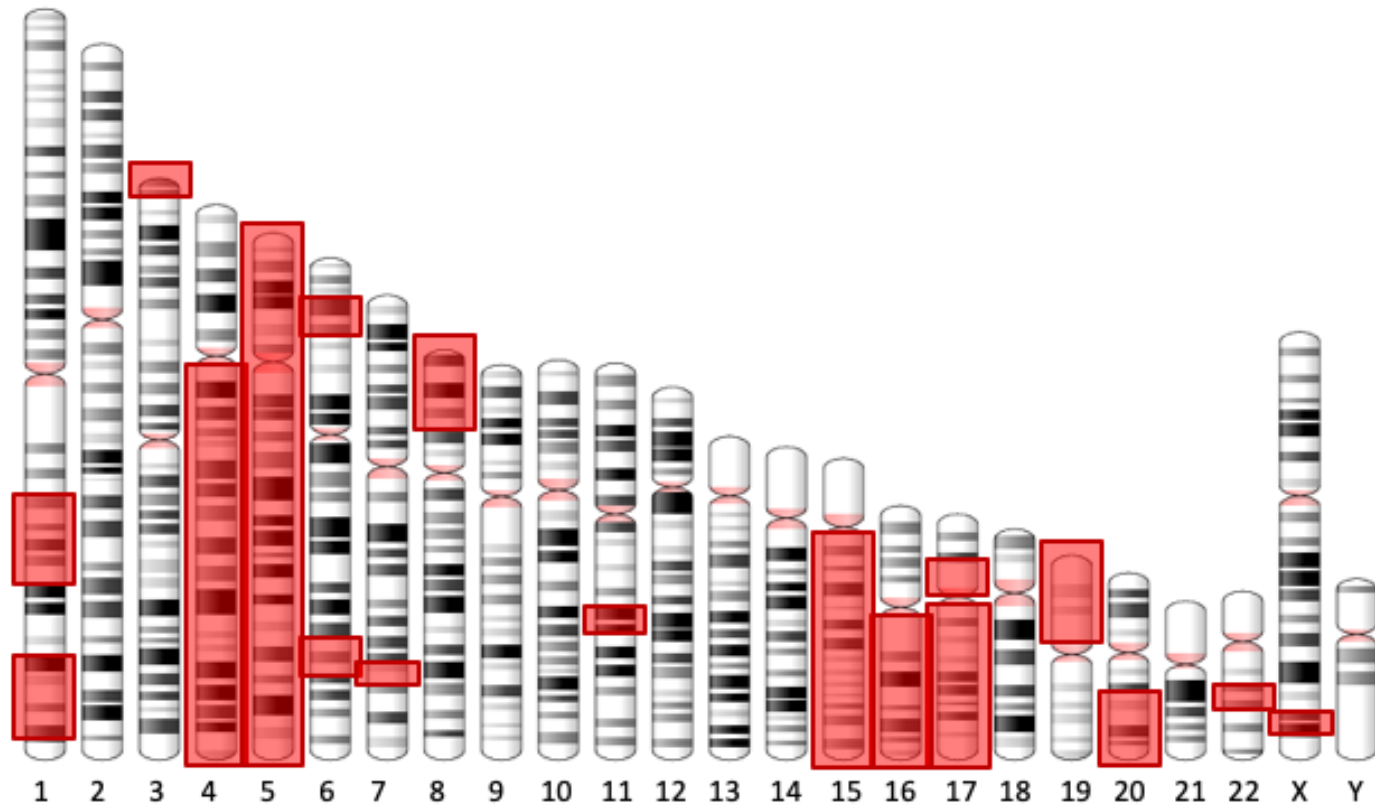
Figure 1.4 Chromosomal regions with evidence of linkage in prostate cancer pedigrees.

Schematic of the autosomes and sex chromosomes; the regions with suggestive evidence of linkage in PCa pedigrees are shown in red (reviewed in Karayi *et al.* (2000) [100]).

High-density screening panels of up to 5 million variants can now be assayed on a genome-wide scale and studies utilising these arrays in case-control populations are known as genome-wide association studies (GWAS) [101]. The variants identified by GWAS are common in the population; defined as having a minor allele frequency (MAF) of >2%, and have an overall small effect size. Therefore, a large sample size is required to identify them in a case-control cohort-based study. Until recently, GWAS and fine-mapping efforts have identified more than 100 common PCa risk variants across multi-ancestral populations, most of which were identified in populations of European ancestry [102-115]. Schumacher and colleagues (2018) developed a custom high-density genotyping array designed to tag most common genetic variants [9]. A meta-analysis combining these summary statistics and seven previous PCa GWAS or high-density single nucleotide polymorphism (SNP) panels (totaling 79,194 PCa cases and 61,112 controls) identified 62 novel loci with 38 variants found within gene-rich regions. Their findings included a missense variant, rs1800057 (odds ratio (OR) =1.16; p=8.2x10^{-9}) in *ATM* [9], a gene that plays a central role in cell division and DNA repair, and therefore is of great interest in cancer [116]. This latest meta-analysis brings the total number of identified common PCa-risk variants to over 170, which accounts for approximately 38.5% of known familial risk (Figure 1.5). Each common variant's contribution to PCa is only small, with an OR for disease risk of less than 1.3 [117]. In combination, common variants have a greater overall impact on disease risk than individually and as a result they are often associated with complex PCa phenotypes [118]. Overall, given that the majority of identified GWAS variants are not within genes and the functional role of those identified remain largely unknown, they have yet to be translated into useful clinical biomarkers.

Common, low penetrance variants also contribute to familial disease. The International Consortium of Prostate Cancer Genetics (ICPCG) demonstrated that 16 of 25 common variants identified by GWAS are also significantly associated with risk in men with a family history, in their study of 9,560 familial PCa cases [119]. A study by Teerlink *et al.* (2014) involved a larger analysis of the same 25 common variants in over 12,000 individuals, which also showed evidence that several common variants identified by GWAS contribute to both sporadic and familial disease [120]. This familial study also led to the discovery of rare genetic variants that underly these common disease loci, following imputation and additional targeted NGS of a number of GWAS regions. These underlying risk alleles were rarer and had larger effect sizes than the common variants [120]. Therefore, this study highlights the potential significance of rare variants in common diseases, such as PCa.
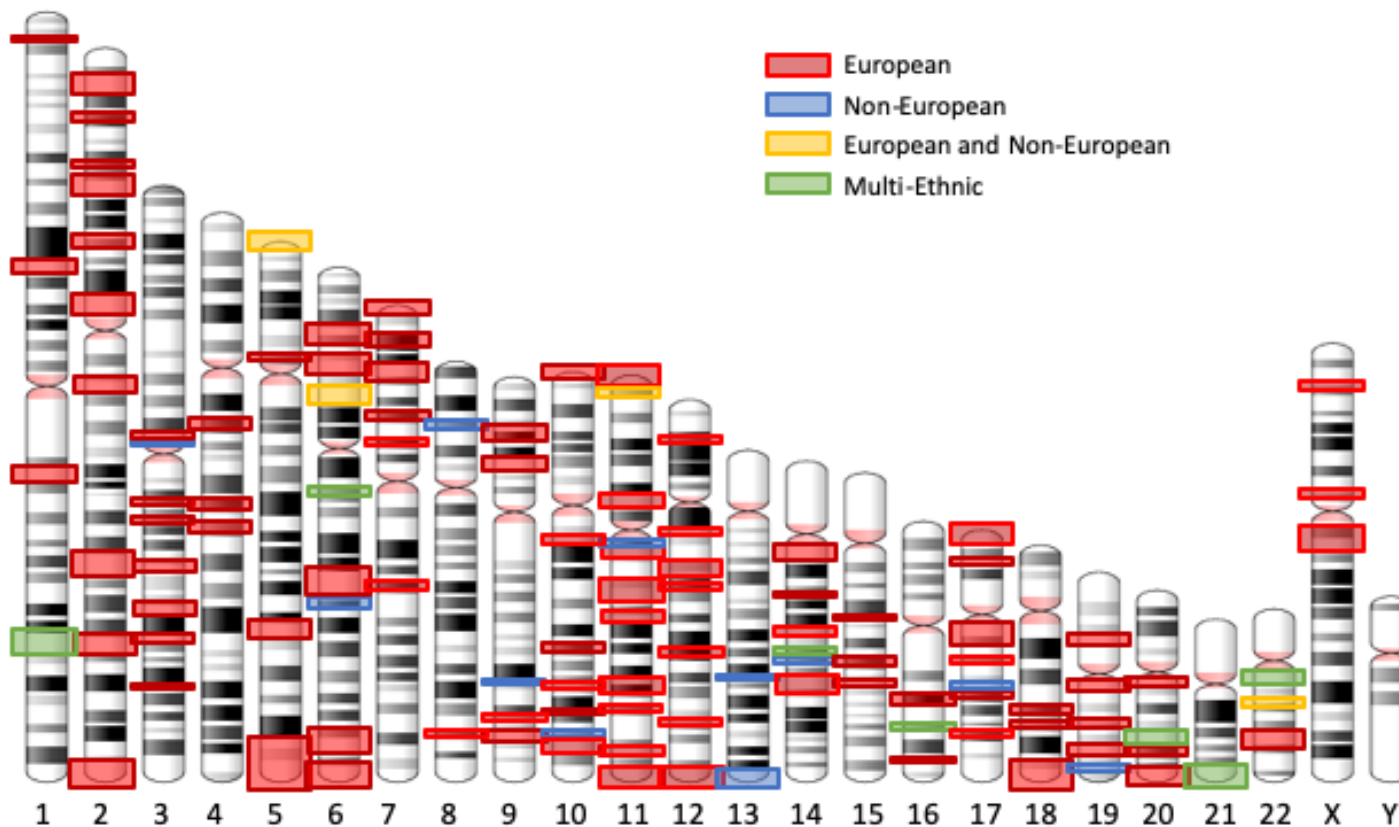
Figure 1.5 The chromosomal regions of the 170 common variants identified by genome wide association studies.

Schematic of all of the autosomes and sex chromosomes; regions harbouring common PCa risk variants are shown in red, blue, yellow and green. Thus far, 170 common genetic variants associated with PCa risk have been identified by 40 GWAS (as highlighted in Schumacher *et al.* (2018) [9]). These variants comprise 115 chromosomal regions and most have been identified in European populations however, some studies have included non-European individuals and those of multi-ethnicity.

1.7 EXAMINING THE CONTRIBUTION OF RARE VARIANTS TO PROSTATE CANCER RISK

Recent studies have suggested that rare variants may have a more apparent role in PCa risk than first thought. Rare disease variants often have a higher effect size compared to common variants, which predominantly have a lower impact on disease risk. This suggests that rare variants with high effect sizes are likely to have an overall greater contribution to disease risk than variants with low effect sizes, however, they are often hard to identify using standard genetic analysis methods, such as GWAS. (Figure 1.6). Mancuso and colleagues (2016) estimated that $\simeq$ 42% (95% CI: 21%-63%) of the genetic risk of PCa is due to rare (MAF <2%) or very rare variants (MAF <1%), and acknowledge that this may be an underestimate [10]. According to the 1000 Genomes project, rare variants are defined as having a MAF of less than 2% (though MAF labels are arbitrary) and it is estimated that there are 10 million in the general population [121]. Rare variants occur too infrequently in the population to be detected by GWAS designed studies, yet the recent GWAS meta-analysis by Schumacher *et al.* (2018) [9] was powered enough to detect rarer genetic variants (MAF 1-2%). They are more easily identified when studying families with a dense aggregation of disease, as there is reduced genetic complexity and rare disease-causing variants are enriched [122].
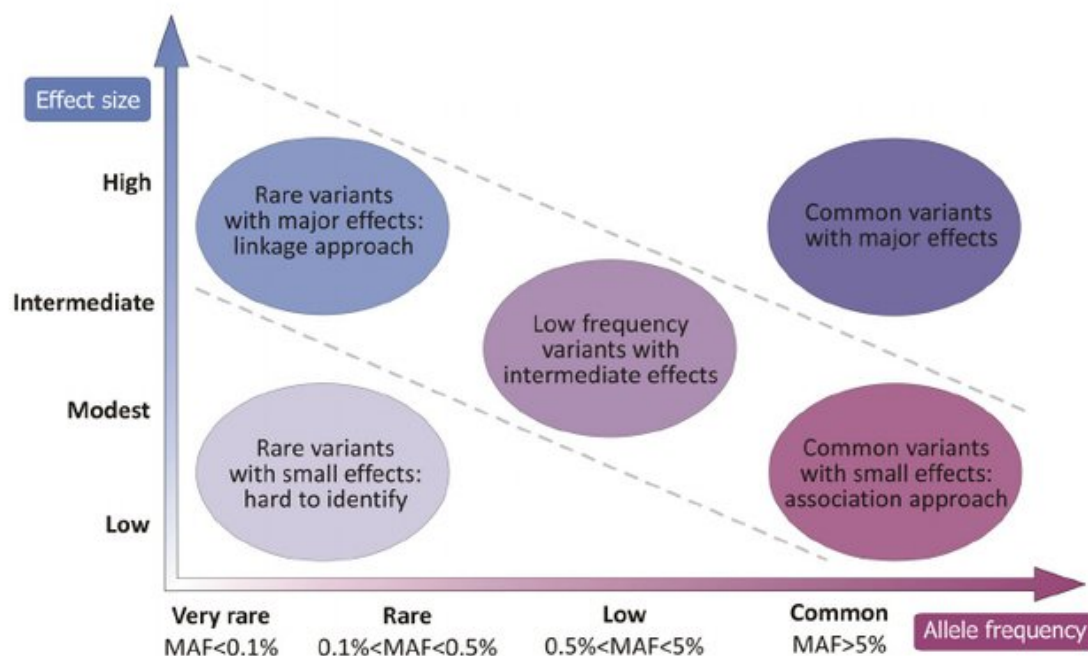
Figure 1.6 Relation of minor allele frequencies, effect sizes and feasibility of identifying disease-associated variants by common genetic tests.

Rare variants often have higher effect sizes compared with common variants thus, they have a greater contribution to disease risk. Adapted from McCarthy *et al.* (2008) and Manolio *et al.* (2009) [123,124].

Massive parallel sequencing encompasses whole-exome sequencing (WES) and WGS, and WGS provides a detailed map of inherited common and rare variants [125]. NGS facilitated the discovery of a rare PCa-associated variant in *HOXB13* (G84E, rs138213197), underpinning a previously established region of linkage at chromosome 17q21-22 [11]. More than 200 genes in the 17q21-22 region were screened by sequencing genomic DNA (gDNA) from 94 cases from PCa families sharing linkage to 17q21-22 (one case per family). These probands are from the University of Michigan Prostate Cancer Genetics Project (n=54) and John Hopkins University (n=40) cohorts. The variant was identified in probands from only four of the 94 families, and following additional sequencing of family members, all 18 men with PCa in these four families carried the G84E variant. Additional sequencing of a larger cohort of cases and controls identified a significantly higher carrier frequency in men with early-onset, familial PCa (3.1%) than in those with late-onset, non-familial PCa (0.6%; $p=2.0 \times 10^{-6}$) [11]. Overall, this study concluded that the novel *HOXB13* G84E variant is associated with a significantly increased risk of familial PCa. Another study by Zuhlke and colleagues (2012) performed targeted NGS of the *NBN* gene in the same 94 familial probands sequenced by Ewing *et al.* (2012). One proband was found to have a novel heterozygous variant in exon 14 of the gene (S706X) and additional sequencing of male relatives showed partial segregation of the variant with PCa [126]. However, this *NBN* variant was not observed among 1,859 PCa cases and 909 controls, all of whom were unrelated. Further to the study by Ewing *et al.* (2012), re-analysis of this NGS data also led to the identification of a novel *SPOP* missense variant (N296I) in a proband who had an early age of disease onset (43 years). Subsequent sequencing confirmed segregation with disease in the proband's family [14]. Interestingly, *SPOP* mutations are the most frequently acquired somatic mutations. Whilst the two later studies did not find an association between *NBN* and *SPOP* variants and disease in a larger PCa cohort, each study has shown the success of applying NGS to family pedigrees to identify rare PCa risk variants through segregation. Further studies using this methodology will be discussed in Chapter 3 of this thesis.

1.8   GERMLINE VARIANTS DRIVE SOMATIC TUMOUR EVENTS

As previously mentioned, several recent studies have revealed evidence to suggest that specific germline variants may increase the probability of a tumour acquiring a particular somatic mutation and together they may interact to drive carcinogenesis [127,128]. Carter *et al.* (2017) used publicly available data from TCGA to identify and validate 395 genetic interactions between germline variants and major somatic events [128]. For example, germline variants in *RBFOX1*

increased the incidence of *SF3B1* somatic mutations, while 19p13.3 germline variants were associated with somatic mutations in *PTEN*. This study concluded that common germline variants influence how and where (in the prostate) tumours develop [129]. A recent study by Mamidi and colleagues (2019) of 305 aggressive tumours and 52 control TCGA samples, observed that genes containing germline mutations also had somatic mutations which interact and cooperate with one another in molecular networks and biological pathways [130]. The interaction between germline variants and somatic tumour events will be discussed in further Chapters 6 and 7 of this dissertation.

1.9   INHERITED DETERMINANTS OF CLINICAL OUTCOMES

The identification and characterisation of rare or novel PCa risk variants will enable a better understanding of both familial and sporadic disease, in particular, the genes and pathways involved. At present there has been relatively little focus on elucidating the role of rare genetic variants contributing to PCa. As a result, efforts to develop tools to improve diagnosis, provide informed prognostic information, and broaden treatment options beyond the traditional therapies of prostatectomy and hormonal and radiation therapy is being hampered. Pharmacological targets of the identified genes and pathways associated with PCa may provide disease control in the advanced, metastatic setting. For example, *PARP* inhibitors are effective in the treatment of melanoma, breast and ovarian cancers and now metastatic PCa, in individuals who carry inherited or somatic rare variants in *BRCA1*, *BRCA2*, *CHEK2*, *PALB2* or *ATM* [131,132]. Plus, recently, preclinical studies have demonstrated an association between *ETS* gene fusions and the effectiveness of *PARP* inhibitors, in which the fusion may confer increased sensitivity to these DNA repair protein inhibitors [133]. Such advances in therapeutic options have been made possible as a result of understanding the genetic drivers of disease, including the functional role of identified genetic risk variants and somatic tumour alterations. Insight into PCa genetical aetiology is required to better understand causal pathways.

1.10   HYPOTHESIS AND AIMS OF THIS STUDY

Our understanding of PCa heritability has improved in recent years due to the identification of both common and rare variants, which explain a proportion of this risk. Rare variants are likely to significantly contribute to PCa heritability and Mancuso and colleagues (2016) estimate that as much as 42% of disease risk may be explained by rare variants [10]. To date, only about 6% of disease risk is currently explained by known rare variants, as their identification in complex

disease has proven quite challenging due to their rarity in the general population. To maximise rare variant discovery, the study of families with an aggregation of disease is a valuable approach due to reduced genetic complexity and an enrichment of the rare-disease causing variant(s) [122]. In recent years the application of NGS to familial studies has also aided in their discovery.

Herein the following hypothesis is addressed:

Rare genetic variants contribute to prostate cancer development and they can be identified by whole-genome sequencing individuals from families with a dense aggregation of disease. The identification of rare prostate cancer risk variants will highlight genes and pathways involved in the malignancy and elucidate some of the currently unexplained heritability of prostate cancer.

This project will utilise the *Tasmanian Familial Prostate Cancer Cohort* and the *Tasmanian Prostate Cancer Case-Control Study* (detailed in Chapter 2). The familial cohort is a rare collection of 52 large Tasmanian PCa families with densely aggregated disease, and consists of genealogical information, clinical and tumour information from pathology reports, and blood and tissue samples from affected men and their unaffected relatives. The population-based case-control study consists of blood and saliva samples from PCa cases and age-matched controls, plus tumour samples from sporadic cases with clinical and tumour information from pathology reports.

This hypothesis will be tested as follows:

Aim 1: Identify rare genetic variants segregating with disease in selected Tasmanian prostate cancer families using whole-genome sequencing data from affected men and selected unaffected/unknown relatives.

Aim 2: Examine the contribution of the identified rare variants to prostate cancer risk in Tasmania, using the remaining families from the *Tasmanian Familial Prostate Cancer Cohort*, and the *Tasmanian Prostate Cancer Case-Control Study*.

Aim 3: Examine the functional effect of the prioritised rare variants using gene and protein expression analyses, as well as determine whether they are associated with particular clinical characteristics.

As briefly discussed above (Chapter 1.8), it is apparent that there are inherited germline variants underly a proportion of somatic tumour variation.

Therefore, this study specifically hypothesises that:

Germline variants predispose the development of some somatic tumour alterations.

This project will utilise the *Tasmanian Prostate Tissue Pathology Resource* (described in Chapter 2), comprising formalin-fixed paraffin embedded (FFPE) prostate tumour tissue samples from the *Tasmanian Familial Prostate Cancer Cohort* and the *Tasmanian Prostate Cancer Case-Control Study*.

To test this hypothesis, I aim to:

Aim 4: Examine the occurrence of shared somatic tumour alterations, including copy number changes and gene fusions, in Tasmanian prostate cancer families, and, where sample size permits, determine whether germline variants predispose to these alterations.

CHAPTER 2 : METHODS

2.1 THE *TASMANIAN FAMILIAL PROSTATE CANCER STUDY*

2.1.1 Ethics approval

Ethics approval was obtained from the Human Research Ethics Committee Tasmania, Australia (H0017040) and this study was conducted in accordance with the Australian National Statement Ethical Conduct in Human Research 2007 (updated 2018) and the Australian Code for the Responsible Conduct of Research (2018). Written informed consent was obtained from all participating individuals. For deceased cases a part of the *Tasmanian Familial Prostate Cancer Cohort*, a waiver of consent was obtained to collect prostate tissue specimens.

2.1.2 The *Tasmanian Familial Prostate Cancer Cohort*

This resource is a rare collection of 52 prostate cancer (PCa) families from the founder population of Tasmania. The number of affected men in these families ranges from five to over 140 and include up to five affected brothers and multiple father/son and uncle/nephew pairs. DNA samples from blood or saliva have been collected for 251 affected men and 463 unaffected/unknown male and female relatives. Families selected for whole-genome sequencing (WGS) analysis include; PcTas3, PcTas4, PcTas12, PcTas22 and PcTas72 (see Table 2.1). PcTas3, 4 and 22 will be discussed in more detail in Chapter 3, PcTas12 in Chapter 4 and PcTas72 in Chapter 5. The PcTas9 family was chosen for assessment of somatic tumour variation and is presented in Chapters 6 and 7 (Table 2.1). Herein, these families are referred to as PcTas families, with each family assigned a number (PC1; prostate cancer family 1) and each individual a unique identification number (i.e. PC1-1; individual 1 from prostate cancer family 1). All familial PCa cases are confirmed by the Tasmanian Cancer Registry (TCR) and cases of other cancer types are self-reported.

Table 2.1 Summary of the *Tasmanian Familial Prostate Cancer Cohort* families utilised in this study.

Family Identification*	Known PCa cases	Generations with PCa	PCa cases with DNA	Unaffected relatives with DNA
PcTas3[3]	14	2	8	14
PcTas4[3]	25	4	9	45
PcTas12[4]	35	4	11	36
PcTas22[3]	89	5	27	70
PcTas72[5]	23	4	12	52
PcTas9[6,7]	58	4	30	75
*The extended pedigrees of the Tasmanian PCa families studied in this thesis are shown in the chapters stated[3-7].				

2.1.3 The *Tasmanian Prostate Cancer Case-Control* Study

The *Tasmanian Prostate Cancer Case-Control Study* is a population-based resource, which includes DNA from blood or saliva samples from 498 PCa cases and 355 age-matched controls. Cases were identified from the TCR and were recruited if they were diagnosed under the age of 75 between the years of 1996 and 2005. Controls were selected at random from the Tasmanian electoral roll and matched by five-year age groups to the cases. Controls are annually checked against the TCR for subsequent PCa diagnosis, hence the number of PCa cases have increased and controls decreased. Herein, each sample has its own unique identification number (i.e. DVA1; individual 1 in the case-control resource) and are often referred to as Tasmanian sporadic cases and controls where required. Following initial variant prioritisation, 94 of these controls were randomly chosen to screen for prioritised rare variants to ensure they were not enriched in the Tasmanian population.

2.1.4 Extraction of germline DNA from blood and saliva

For participants in the *Tasmanian Familial Prostate Cancer Study* cohorts, including the familial and case-control resources, genetic material for DNA was extracted from blood using the Nucleon BACC3 Kit (GE Healthcare) and from saliva using the Oragene DNA Kit (DNA Genotek), according to the manufacturers' directions. Quality and quantification of DNA was performed using the Nanodrop® ND-1000 UV-vis spectrophotometer (Nanodrop® Technologies).

2.1.5 The *Tasmanian Prostate Tissue Pathology Resource*

The *Tasmanian Prostate Tissue Pathology Resource* comprises 76 familial (PC) and 22 sporadic (DVA) formalin-fixed paraffin embedded (FFPE) prostate tumours. Clinical information including Gleason score (GS), age at diagnosis, and diagnoses and treatment history were obtained from pathology reports corresponding to the FFPE tumour blocks retrieved for the functional analyses of this study. If reports were vast or unattainable, Dr Shaun Donovan (Pathologist, Hobart Pathology, AUS) re-graded the tumour blocks using the contemporary Gleason scoring system, as described in Chapter 1.3.2.

2.1.6 The *Tasmanian Prostate Tissue Needle Biopsy Resource*

Several prostate needle biopsies were also available for use in this study. These samples were collected by a urologist whilst patients underwent a prostate resection. The radical prostatectomy was sent to pathology for diagnosis and the biopsies for research purposes (stored in RNAlater). These biopsies consist of cores from the right and left lobe of the prostate, and are herein referred to as PT samples. Ethics approval was obtained from the Human Research Ethics Committee Tasmania, Australia (H0011544) for use of these biopsies in this study.

2.1.7 Extraction of genetic material from prostate tumour samples

FFPE prostate tissue blocks were sectioned to 8μm, dewaxed and rehydrated using a standard xylene-ethanol deparaffinisation protocol. Malignant and benign glands were marked on haematoxylin and eosin (H&E) stained tissue sections by a pathologist. Marked malignant and benign regions were macro-dissected separately for both DNA and RNA. DNA was extracted using the QIAamp DNA FFPE Tissue Kit (QIAGEN), according to the manufacturer's instructions and eluted in 50μL of ATE Buffer. DNA was quantified using the Nanodrop® ND-1000 UV-vis spectrophotometer (Nanodrop® Technologies). RNA was extracted using the RecoverAll Total Nucleic Acid Isolation Kit (ThermoFisher Scientific), according to the manufacturer's instructions and eluted in 30μL of dH$_2$O. RNA quality (% of sample >200nt in length) and quantity (ng/μL) was assessed using the 2100 Bioanalzyer (Agilent Technologies) and/or the 4200 Tapestation (Agilent Technologies), with their respective software. The SuperScript™ VILO™ cDNA Synthesis Kit (Invitrogen) was used for cDNA synthesis, according to the manufacturer's instructions, using the thermal cycling conditions in Appendix 1. For the needle biopsy samples in RNAlater, small sections of tissue were transferred to a

new tube for extraction of genetic material using the protocols described above. The right and left lobe biopsies were extracted separately.

2.2 PCR; PRIMER DESIGN, QUANTIFICATION AND VISUALISATION

All primers used for amplification of gDNA, and FFPE DNA and RNA were designed using Primer3 [134,135] or Primer-BLAST [136] and were synthesised by Sigma-Aldrich or Integrated DNA Technologies. Veriti 96 thermal cyclers from Life Technologies were used for all PCR amplifications, unless otherwise specified in Appendix 1. Primer pairs and their optimal annealing temperatures are shown in Appendix 2. PCR products were visualised on 2% agarose gel (80 volts for 30 minutes) for length and mass quantification. The agarose gels were visualised and photographed with the ChemiDoc XRS+ System (BioRad). PCR products for Sanger sequencing were quantified using the Nanodrop® ND-1000 UV-vis spectrophotometer (Nanodrop® Technologies).

2.3 QUANTIFICATION OF ABSOLUTE GENE EXPRESSION BY RT-QPCR

SYBR green real-time quantitative PCR (RT-qPCR) assays were used to determine expression of the genes of interest and two housekeeping genes, *β-Actin* and *GAPDH*. RT-qPCR primers were designed to the most commonly transcribed isoform in the prostate (as per GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2; https://gtexportal.org/home/)) [137] and are displayed in Appendix 3. Amplification was performed on 50ng FFPE cDNA, in triplicate, as per the conditions in Appendix 1. Quantitation and melt data was visualised using the Rotor Gene 6000 Series Software 1.7 or the QuantStudio™ Design and Analysis Software v1.5 and each RT-qPCR run was conducted with a DNA-free NTC.

Standard curves were generated for the genes of interest and the two housekeepers to determine PCR efficiency and normalise absolute gene expression of the genes of interest. PCR products were pooled and visualised by gel electrophoresis, as described in Chapter 3.2.1.2. Bands were excised (SafeImager, Invitrogen) and purified using the QIAquick Gel Extraction Kit (QIAGEN). Serial dilutions of this product were amplified by RT-qPCR and standard curves plotted (Appendix 4). The copy number of the gene of interest and the two housekeeping genes was determined using the log equation from the line of best fit. The absolute gene expression was determined by normalising the copy number of the genes of interest to the geometric mean of the copy number of *GAPDH* and *β-Actin*.

The paired Student's t-test was used to compare absolute gene expression between malignant and adjacent benign cells. The unpaired Student's t-test was used to compare absolute expression in the malignant glands of variant carriers *versus* non-carriers, and in the benign glands of variant carriers *versus* non-carriers. In Chapters 6 and 7, the unpaired Student's t-test was used to compare absolute gene expression in malignant glands of PcTas9 tumours *versus* non-PcTas9 tumours, and likewise in benign glands. P values <0.05 were considered to be statistically significant, with fold changes presented in box plot format using R studio, version 0.99.887.

2.4   QUANTIFICATION OF PROTEIN EXPRESSION BY IMMUNOHISTOCHEMISTRY

Following dewaxing, tissue sections (3.5μm) were pre-treated with Target Retrieval Solution (Dako), followed by inactivation of endogenous peroxidases using 3% hydrogen peroxidase (Sigma-Aldrich). Non-specific staining was blocked using Protein Block (Dako). Sections were incubated with primary antibody (Appendix 5) in a humidified chamber for one hour, followed by a 30-minute incubation with a HRP-Labelled Polymer (Dako). Protein staining was visualised with 3-3' diaminobenzidine (DAB$^+$) for 10 minutes, and the sections were counterstained using Mayer's haematoxylin, cleared and cover slipped using the Dako Automated Coverslipper.

The immuno-stained sections were scored by a pathologist (Drs Donovan and Malley; Hobart Pathology) blinded to variant carrier status. Staining was scored as none, weak, moderate or strong, depending on the most common staining intensity in the entire tissue section. Immunostaining was assessed using a quasi-continuous score, created by multiplying each intensity level (0 for no stain, 1 for weak stain, 2 for moderate stain, and 3 for strong stain) by the corresponding percentage of positive cells. As benign prostate tissue was also present in some sections, immunostaining was assessed for both malignant and benign glands separately [138].

The paired Student's t-test was used to compare protein expression between malignant and adjacent benign cells. Unpaired Student's t-tests were used to compare protein expression in the malignant cells of variant carriers *versus* non-carriers, and in the benign cells of carriers

versus non-carriers. P values <0.05 were considered to be statistically significant. Images were taken using the Leica DM2500 microscope with the Leica Applicate Suite software, version 3.4.1 or the Olympus BX53 microscope, using the DP73 camera and software (x100).

CHAPTER 3 : PRITORITISATION, VALIDATION, SEGREGATION AND ASSOCIATION ANALYSES OF RARE VARIANTS

3.1 INTRODUCTION

Gene discovery has proven useful for attaining a greater understanding of disease and aiding in the identification of new targets for therapy. Studies of families with familial hypercholesteraemia have not only identified genes and pathways associated with increased lipid levels in cardiovascular disease, but have also facilitated the development of statins [139,140]. The recent emergence of next-generation sequencing (NGS) has proven very successful, particularly the combined use of family cohorts in the common disease setting. Such studies have highlighted the contribution of rare variants to common disease [141]. NGS-based studies of families with early-onset Alzheimer's disease have each identified unique rare variants in *NOTCH3, SORL1* and *TREM2*, all associated with disease risk in their respective cohorts [142-144]. The proven role of rare variants in complex disease, including breast and ovarian cancers, suggest that such discoveries would also be highly valuable in prostate cancer (PCa).

Cirulli *et al.* (2010) highlighted that an agnostic NGS approach when applied to families can be more successful than a hypothesis driven, targeted sequencing approach, but there are very few studies published using this method [122]. To date, while not truly genome-wide, there have only been two whole-exome sequencing (WES) studies of familial PCa. One of the first studies was performed at the Fred Hutchinson Cancer Research Center (FHCRC), which included 91 individuals from 19 PCa families with an aggressive or early-onset phenotype [12]. A total of 130 rare variants identified from the WES data were then genotyped in an independent set of 270 PCa families, which included 819 cases and 496 unaffected relatives. Two missense variants in *BTNL2* (D336N, G454C) were identified in 1.5% (D336N; p=0.0032) and 1.2% (G454C; p=0.0070) of affected men, but no unaffected men were observed to carry either variant. Further genotyping of the variants in a population-based case-control cohort (n=1,155 PCa cases and 1,060 age-matched controls) suggested both variants were associated with an elevated risk of PCa (D336N: Odds ratio (OR)=2.7, p=0.010; G454C: OR=2.5, p=0.019) [12].

More recently, Karyadi and colleagues (2017) performed a second analysis of WES data generated from the FHCRC familial resource, including 160 PCa cases from 75 families. Analysis took into account the genetic heterogeneity and incomplete penetrance of PCa susceptibility alleles and identified 341 candidate risk variants [13]. Analysis of these variants in the FHCRC population-based, case-control resource identified nine variants significantly associated with an increased risk of PCa. In a second analysis of an independent case-control cohort (n=7,121), there was evidence for association with risk for a rare variant in *TANGO2* (S17X: OR=1.39, p=0.065) and the established *HOXB13* variant (G84E: OR=3.78, p=0.0003) [13]. A meta-analysis of the two case-control studies identified two additional variants with suggestive evidence for an association with PCa risk, *OR5H14* (M59V: OR=1.39, p=0.026) and *CHAD* (A342D: OR=1.53, p=0.046). Similar to the original *HOXB13* study, these WES studies highlighted novel rare variants that segregated with PCa in multiple high-risk families, but were also found to contribute to disease risk in the general population [13]. Furthermore, several studies have since replicated the *HOXB13* finding in Caucasian familial and case-control populations and estimate the variant to be associated with a 4- to 8-fold increase in PCa risk, as well as with early-onset disease [102,145-150]. Such studies highlight the success in combining familial datasets and NGS technologies to discover rare variants associated with PCa risk.

Although NGS studies of PCa families have revealed that rare PCa risk variants exist, studies are few and far between [11-13]. Studies by Ewing *et al.* (2012), FitzGerald *et al.* (2013) and Karyadi *et al.* (2017) assessed the contribution of these rare variants to other PCa families, as well as case-control cohorts and found significant associations with PCa risk in their cohorts. However, follow-up studies assessing the contribution of the rare variants to other populations is non-existent, with the exception of the *HOXB13* G84E variant [11,151-156].

Here, I sought to address the hypothesis that rare genetic variants contribute to PCa risk. This chapter will describe the application of whole-genome sequencing (WGS) to our rare *Tasmanian Familial Prostate Cancer Study* cohorts, with the aim of identifying rare PCa-risk variants. The identification of disease-associated rare variants should be facilitated by the fact that Tasmania has an isolated population with reduced genetic heterogeneity [157]. Thus, the anticipated enrichment of rare variants in our Tasmanian PCa families is likely to reduce genetic complexity and increase statistical power for the identification of risk genes [122]. Herein, five Tasmanian PCa families were selected for WGS based on dense disease aggregation and

availability of DNA samples; PcTas3, 4 and 22 are discussed in this chapter, whilst PcTas12 is discussed in Chapter 4 and PcTas72 in Chapter 5.

3.2   METHODS

3.2.1   Whole-genome sequencing analysis

Thirty-three individuals from five Tasmanian PCa families, PcTas3, 4, 12, 22 and 72 (described in Chapter 2.1.2), including 23 PCa cases and 10 unaffected relatives, were selected for WGS. Individuals were prioritised for WGS based on the following pedigree features; affected first-degree relatives from densely clustered affected regions of the pedigree; second-degree affected relatives; early-onset and/or aggressive disease; and, where possible, unaffected, older, first-degree male relatives as a potential comparative genome from the same family, and availability of funding. Distantly related, affected family members were also included, as these cases will share less of the main pedigrees' genome, perhaps revealing the shared disease-causing variants. WGS was performed for eight controls from the *Tasmanian Prostate Cancer Case-Control Study* (described in Chapter 2.1.3) to provide us with rare variant sequence data from unaffected age-matched members of the Tasmanian population. WGS was performed at the Kinghorn Centre for Clinical Genomics, Australia, on the Illumina HiSeq X^{TM} Ten platform, using the TruSeq Nano library preparation.

3.2.2   Whole-genome sequencing analysis pipeline

Sequence data analysis was undertaken using the Variant Analysis of Sequenced Pedigrees (VASP) analytical pipeline, developed specifically to detect disease causing variants in sequenced pedigrees [3,4]. VASP integrates information from each pedigree member, and therefore describes the likely inheritance pattern of shared variants, whilst incorporating external annotation of these variants, including population frequency information from Exome Aggregation Consortium (ExAC; non-Finnish European, non-TCGA (The Cancer Genome Atlas) population) [158], as well as SIFT [159], PolyPhen2 [160] and CADD (Combined Annotation Dependent Depletion; model v1.3) [161] scores for estimating the functional effect of missense mutations. Individual samples were analysed independently, followed by a pedigree-wide variation analysis, with all work run in parallel at the National Computational Infrastructure on the Raijin cluster. Sequence data were aligned to the human reference genome (hg19) using BWA, and BAM files and variants were called using either SAMtools/BCFtools or GATK best practices. Variants were annotated using Ensembl Variant Effect Predictor [162] and overlapped

with Ensembl canonical transcripts and splice site variants; defined as 10bp either side of a coding exon. VASP can accommodate pedigrees of any size and will report disease inheritance patterns and gene phasing information when an individual and at least one parent is sequenced. Consistent with our hypothesis each family was analysed separately, although cross referencing of prioritised rare variants was undertaken. The entirety of this work was performed by Dr Matt Field, James Cook University (AUS).

Variant reports for single nucleotide variants and insertions/deletions (indels) were generated when variants were detected in at least one pedigree member. Variants and indels were categorised as either novel, rare or common, or no frequency data available. Prioritisation was firstly guided by the frequency of the variant (minor allele frequency; MAF) in a publicly available population database; MAF <2% in ExAC [158] Secondly, whether the variant segregated with disease in the sequenced individuals, i.e. most, if not all PCa cases carried the variant. And thirdly, *in silico* functional prediction tools, such as SIFT, PolyPhen2 and CADD. SIFT predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids [159] (pipeline illustrated in Figure 3.1). Each variant is appraised qualitatively, as tolerated (score of 0.05-1.0) and deleterious (score of 0.0-0.05) [159]. Polyphen2 predicts the possible impact of an amino acid substitution on the structure and function of a human protein, with the prediction based on a number of features comprising the sequence, phylogenetic and structural information characterising the substitution [160]. Each variant is appraised qualitatively, as benign (score of 0.0-0.15), possibly damaging (score of 0.15-0.85) and probably damaging (score of 0.85-1.0) [160]. CADD predicts the deleteriousness of single nucleotide polymorphism (SNP) variants and insertion/deletion variants by integrating multiple annotations including conservation and functional information into one metric [161]. CADD provides a ranking rather than a prediction or default cut-off, with higher scores more likely to be deleterious. A CADD score above 30 ranks the variant in the top 0.1% of deleterious variants in the human genome; a CADD of 20-30 in the top 1% and 10-20 in the top 10% [161]. Finally, the carrier frequency of these prioritised variants were determined in the eight controls and a literature search was undertaken (Figure 3.1). ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) [163] and PubMed were used to determine if the variant has been associated with a particular disease, identify whether the gene/proteins function is biologically relevant to prostate or cancer biology and finally, whether the gene has been associated with any type of cancer. Throughout the relevant tables, ClinVar annotations are reported, including what condition the variant has been associated with, as well as

interpretation of the variant. The interpretation of the variant is based on aggregating data from submitters [163].

3.2.3 Validation and segregation of prioritised variants

Variants identified by WGS were validated in the original sequenced individuals by PCR and Sanger sequencing. Upon validation, close relatives were also genotyped by Sanger sequencing to determine segregation of the particular variant with PCa (Figure 3.1). If gDNA was unavailable, DNA from formalin-fixed paraffin embedded (FFPE) prostate tissue (where available) was sequenced to determine carrier status. 10ng/µL of genomic DNA (or FFPE DNA) was amplified, according to the conditions in Appendix 1. A no template control (NTC) was included with each PCR run. PCR products were visualised by gel electrophoresis and then purified prior to sequencing by paramagnetic bead purification, using AGENCOURT AMPure XP beads (Beckman Coulter), according to the manufacturer's instructions. The Big Dye Terminator (BDT) v3.1 Cycle Sequencing Kit (Life Technologies) was used to sequence the purified product, as per the conditions in Appendix 1. The BDT DNA fragments were purified using the AGENCOURT CleanSeq beads (Beckman Coulter), according to the manufacturer's instructions. Purified products were sequenced on the ABI 3500 Genetic Analyser (Applied Biosystems). Sanger sequencing results from the 3500 Series Data Collection Software 3 were analysed using the Sequencher software package, version 4.10.1 (Gene Codes Corporation).

A

| ExAC minor allele frequency <2% | + | Segregation with disease in the WGS individuals | + | *In silico* functional prediction SIFT, PolyPhen2 & CADD Score Nonsense/missense variants: CADD >15 Splice variants: CADD >10 | = | Prioritised |

B

| Variant not present in control WGS data 0 or 1 out of 8 controls carried the variant | + | Variant/gene associated with disease Known disease associated variant Association of the gene with cancer Gene/protein function biologically relevant | = | Prioritised for validation |

C

| Validation in WGS individuals Genotype of sequenced individuals confirmed by Sanger sequencing | + | Segregation with disease The genotype of each family member with DNA available was determined Variants that segregated with PCa in their founder family were prioritised | = | Prioritised for follow-up studies |

Figure 3.1 Pipeline for prioritisation of rare variants.

Following WGS, rare variants (ExAC MAF <2%) that segregated with PCa and were predicted to have a functional consequence by SIFT, PolyPhen2 and CADD were prioritised. Variants were screened in eight Tasmanian controls and variants in none or one of these controls were prioritised further. ClinVar [163] and PubMed were used to determine if the variant was associated with a particular disease, identify whether the gene/proteins function is biologically relevant to prostate or cancer biology and finally, whether the gene has been associated with any type of cancer. These prioritised variants were validated by Sanger sequencing of the individuals who were WGS and then determined if they segregated with disease in the founder families. Next, the rare segregating variants were screened in an additional 94 Tasmanian control samples, followed by the entire *Tasmanian Familial Prostate Cancer Cohort* and the *Tasmanian Prostate Cancer Case-Control Study* . ExAC: ExAC, non-Finnish European, non-The Cancer Genome Atlas database; WGS; Whole-genome sequenced; CADD: Combined Annotation Dependent Depletion; Control: Control from the *Tasmanian Prostate Cancer Case-Control Study.*

3.2.4 TaqMan genotyping of the *Tasmanian Familial Prostate Cancer Study* cohorts

Following validation and segregation analyses, rare variants were screened in an additional 94 Tasmanian control samples to ensure that they are not specifically enriched in Tasmania (described in Chapter 2.1.3). Variants with a carrier frequency in the 94 controls less than twice as high as the ExAC database MAF were considered not enriched. This cut-off is reasonably high given that these 94 controls are a small random representation of the larger control resource (n=355). If not enriched, custom TaqMan SNP (single nucleotide polymorphism) genotyping assays were used to genotype the remaining 51 PcTas families and the *Tasmanian Prostate Cancer Case-Control Study* for the prioritised variants (Appendix 6; Applied Biosystems). This was performed on all available gDNA samples, according to the conditions in Appendix 1. Analysis was conducted using the LightCycler® II 480 software, version 1.5.1.62 SP2, which was used to determine the genotype of each sample. Heterozygous individuals were confirmed by Sanger sequencing, as described above (Chapter 3.2.3).

3.2.5 Statistical analysis of genotyping data

Genotype data were analysed using M_{QLS} [2], an association analysis that maximises power by performing tests of association in the combined familial and case-control datasets, while taking into account relatedness of individuals. M_{QLS} can distinguish between unaffected controls and controls of unknown phenotype (unaffected male yet to reach average age of PCa diagnosis) and incorporates phenotype data about relatives who have missing genotype data for the particular variant being tested [2]. M_{QLS} uses variance components to examine the significance of association for related individuals, and when the disease status is known for first-degree relatives of cases, M_{QLS} obtains more power by giving increased weighting to those individuals with closely related disease-carrying relatives [2]. It is computationally feasible in large pedigrees and thus, here, a positive association (OR) with a p-value <0.05 was considered to be statistically significant, and therefore the variant strongly associated with PCa in the Tasmanian cohort. This analysis was performed by our collaborator, Dr Russell Thomson, Western Sydney University (AUS).

SOLAR Eclipse version 8.1.1 was also used to determine whether the variant of interest was enriched in our Tasmanian resource compared to the ExAC database, as well as comparing carrier status within the Tasmanian resource. This analysis was achieved by calculating a Maximum Likelihood Estimate (MLE), which is synonymous with allele frequency of each genotype in each group. These MLEs were then compared between groups using a Wald test,

generating a chi-square test with one degree of freedom, with a p-value <0.05 considered statistically significant. This analysis determines whether the variant of interest is enriched in Group A *versus* Group B, it does not weight by PCa case status. This analysis was performed by Dr Nicholas Blackburn, University of Texas Rio Grande Valley (USA).

3.3 RESULTS

3.3.1 Quality check and annotation of variants

All genomes passed standard quality and coverage assessment (minimum cut-off of 20X coverage). The total number of variants that passed quality assessment in each family are shown in Table 3.1. The total number of rare variants (MAF <2%), very rare variants (MAF <1%) and novel variants are also presented, with these cut-offs as per ExAC annotations [158]. Briefly, Illumina HiSeq Paired-End WGS data were aligned to the human reference genome (hg19) to identify the genomic variants that differed from the reference genome. Variants were called if >10% of the sequence reads at each base pair differed from the reference. The variants were then filtered under a set of pre-defined criteria to eliminate false-positives and were then annotated using VASP [3,4]. The total number of variants identified in all five Tasmanian families is presented below (Table 3.1). PcTas3, 4 and 22 analysis is presented in this chapter, whilst PcTas12 analysis is presented in Chapter 4 and PcTas72 in Chapter 5.

Table 3.1 The total number of variants that passed quality assessment in each of the families where whole-genome sequencing data was available.

Family Identification	Individuals WGS (affected/unaffected)	Total Variants	Rare Variants (MAF <2%)	Very Rare Variants (MAF <1%)	Novel Variants
PcTas3	5/0	178,311	103,675	73,315	52,159
PcTas4	4/1	167,774	147,455	107,195	14,554
PcTas12	2/1	116,381	66,744	47,952	35,722
PcTas22 Main*	5/1	332,047	191,172	109,480	70,497
PcTas22 Sub*	4/2	414,020	160,496	113,512	83,235
PcTas72	4/4	238,076	139,136	99,520	68,437
WGS; whole-genome sequenced: MAF; Minor allele frequency as per the ExAC, non-Finnish European, non-The Cancer Genome Atlas database: *Due to the number of individuals sequenced in PcTas22 and the magnitude of data available, two branches of the family were analysed separately as 'sub' and 'main' pedigree.					

3.3.2   Rare variant prioritisation in PcTas3

Tasmanian PCa family PcTas3 comprises 14 known cases across two generations (Figure 3.2). DNA was available for eight of these cases and five were successfully WGS (Figure 3.3). Men selected for sequencing represent three affected branches of the family and include two affected brother pairs (one brother with a relatively younger age of diagnosis (54 years)), plus a second/third cousin (Table 3.2). Variants in three, four or five out of the five PCa cases were prioritised. It is likely that such rare variants are not completely penetrant therefore, it is possible that not all PCa cases may carry the risk variant. Variants that were shared by the majority of the cases were prioritised for further study.
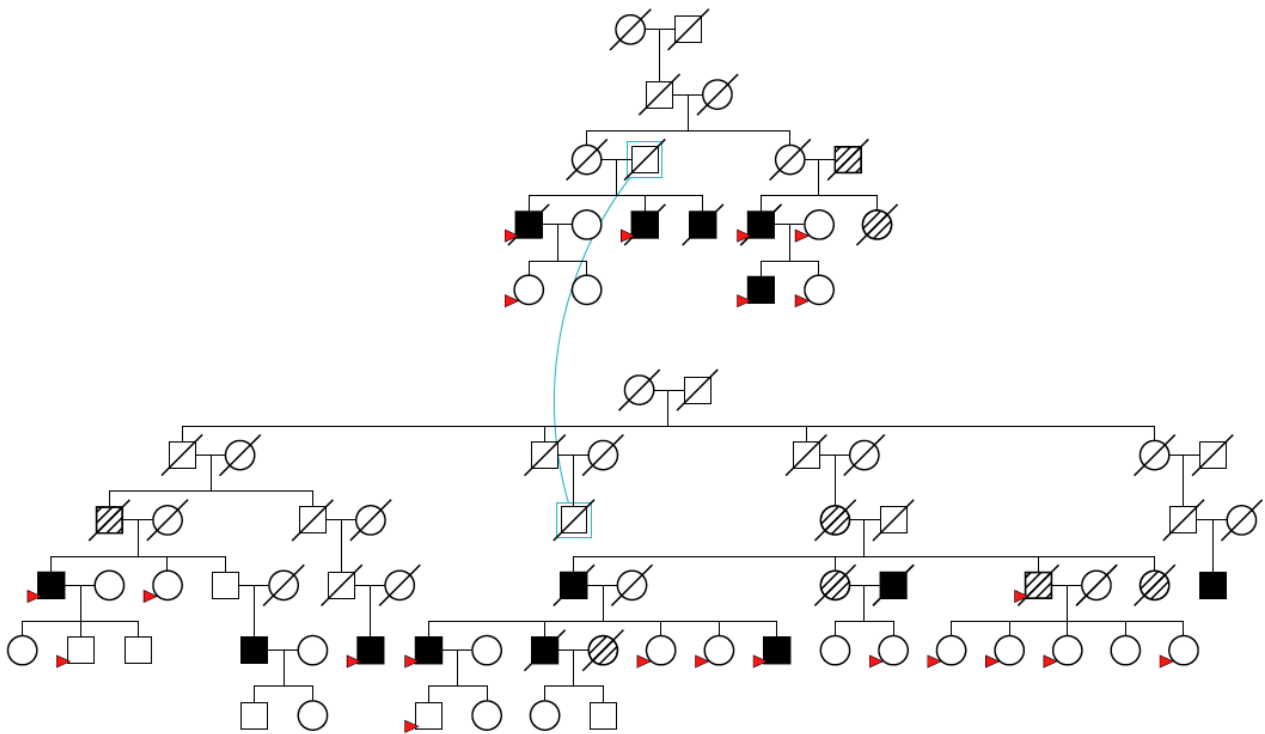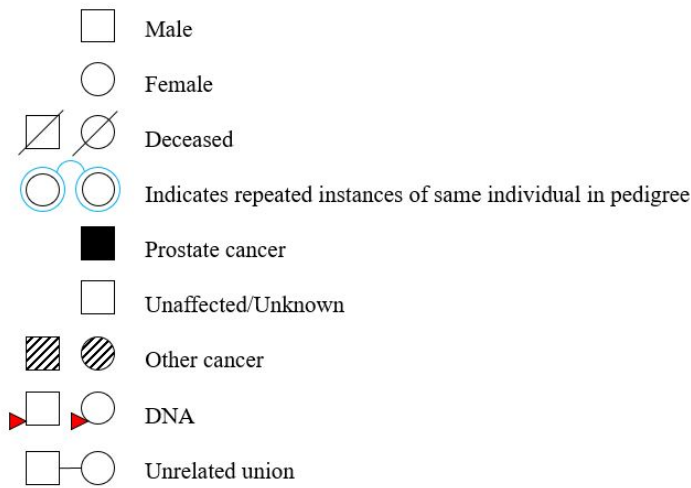
Figure 3.2 PcTas3 pedigree.

PcTas3 pedigree, depicting the number and relationships of PCa cases (shown in shaded squares), as well the availability of DNA from cases and their unaffected relatives, which is represented by red arrows. The disease status for earlier generations is generally unknown, unless this information was obtained from clinical records. And if so, these individuals have been marked as affected in the pedigrees. This pedigree is included to illustrate the size of the pedigree only, please refer to Figure 3.3-3.5 for individual annotations.

Table 3.2 Clinicopathological characteristics of individuals from PcTas3 chosen for whole-genome sequencing.

Sample Identification	Sex	Prostate Cancer Affection Status	Age at diagnosis	Tumour Grade[1]	Contemporary Gleason Score[2]
PC3-01	Male	Affected	79	MD	-
PC3-02	Male	Affected	75	WD	-
PC3-08	Male	Affected	69	MD	6 (3+3)
PC3-31	Male	Affected	54	-	5 (3+2)
PC3-44	Male	Affected	60	Unknown*	Unknown*

[1]Tumour grade obtained from pathology report; [2]Contemporary Gleason Score from FFPE tissue block chosen for macrodissection of nucleic acids and IHC; WD: well differentiated; MD: moderately differentiated; - : information not present in original pathology report; *Diagnosed interstate.
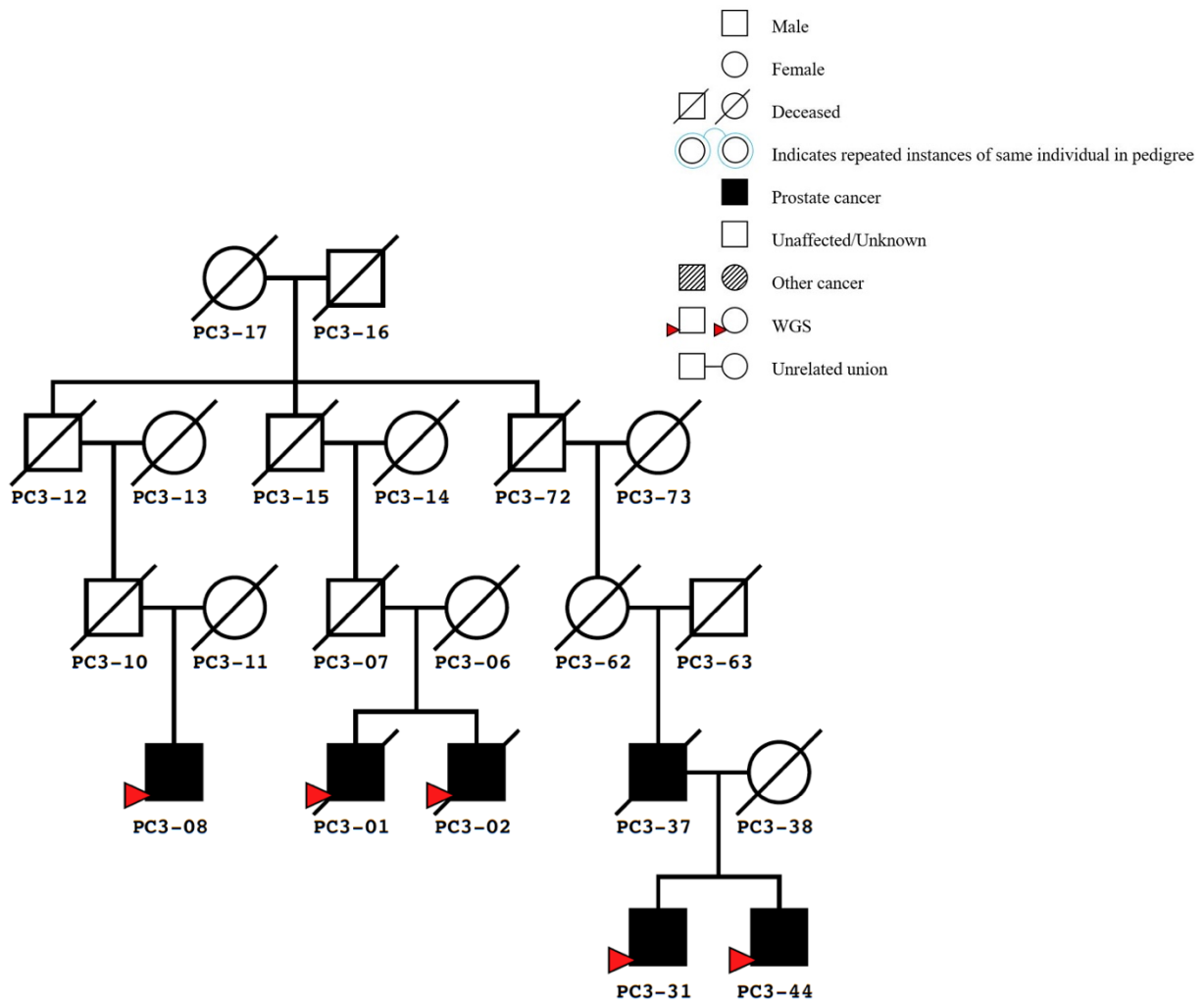
Figure 3.3 A condensed PcTas3 pedigree showing individuals chosen for whole-genome sequencing.

Individuals chosen for WGS are indicated by red arrows, in this case, five PCa cases were chosen.

Three rare variants and one novel variant were prioritised in this family (Table 3.3). These rare variants have not been previously reported as associated with cancer, however the genes they reside in are biologically relevant to cell development, growth and proliferation [116]. Each variant was validated by Sanger sequencing. Following Sanger sequencing of an additional two PCa cases and 13 relatives, only the *CCL26* and *P2RX7* variants were found to segregate with disease in the extended family members. Figure 3.4 shows the identification of four additional *CCL26* variant carriers in this family; an unaffected male and three females. Nine additional *P2RX7* carriers were identified in PcTas3, including a PCa case, two unaffected males and six females (Figure 3.5). The unaffected male, PC3-51, who carriers both of these variants died at age 90 and was affected with another cancer. At 50 years of age, PC3-48, an unaffected *P2RX7* carrier, is yet to reach the average age of PCa diagnosis (~65 years of age). The *NDE1* variant did not appear to segregate with disease. The variant in *CLDN4* validated, but only four individuals were identified as carriers, three of which were initially identified by WGS.

Table 3.3 Rare variants prioritised in the PcTas3 pedigree following whole-genome sequencing of five affected men.

Gene	rs number	Chromosome: base pair	ExAC[1] MAF (%)	Segregation in WGS Individuals (affected carriers)	CADD[2] Score	Allele Change; Amino Acid Change	Number of Control Carriers	ClinVar Search[3]	Validation in WGS Individuals	Segregation in Entire Family
CCL26	rs41463245	7:75,401,263	0.86	4 out of 5	34	C > T; W44X	0 out of 8	Not reported	Yes	Yes
P2RX7	rs28360447	12:121,600,238	1.27	4 out of 5	32	G > A; G150R	0 out of 8	Not reported	Yes	Yes
NDE1	rs113493697	16:15,785,049	0.88	5 out of 5	23.3	C > T; T191I	0 out of 8	Condition not specified: Benign	Yes	No
CLDN4	Novel	7:73,246,102	N/A	3 out of 5	20.8	A > G; K191E	0 out of 8	N/A	Yes	No

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; MAF: Minor allele frequency; N/A: Not found in ExAC or ClinVar; WGS: Whole-genome sequenced; [2]CADD: Combined Annotation Dependent Depletion [164]; Control: Control from the *Tasmanian Prostate Cancer Case-Control Study;* eight were WGS; [3]Associated condition: Interpretation of variant [163].
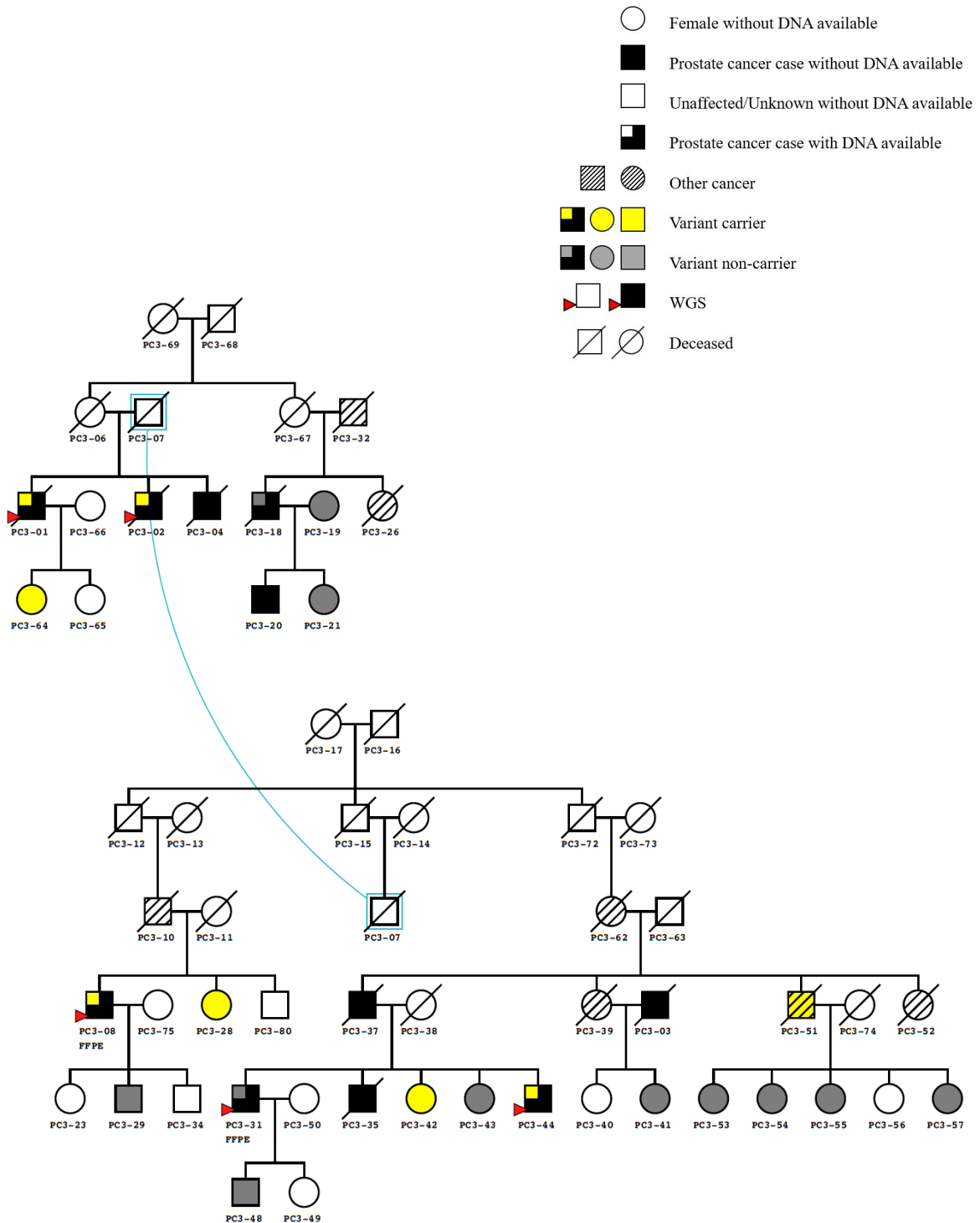
Figure 3.4 *CCL26* variant carriers in PcTas3.

This is a condensed pedigree of PcTas3 comprising all *CCL26* variant carriers (shown in yellow) and their relationship. Non-variant carrier family members are shown in grey and the five individuals who were WGS are indicated by red arrows.
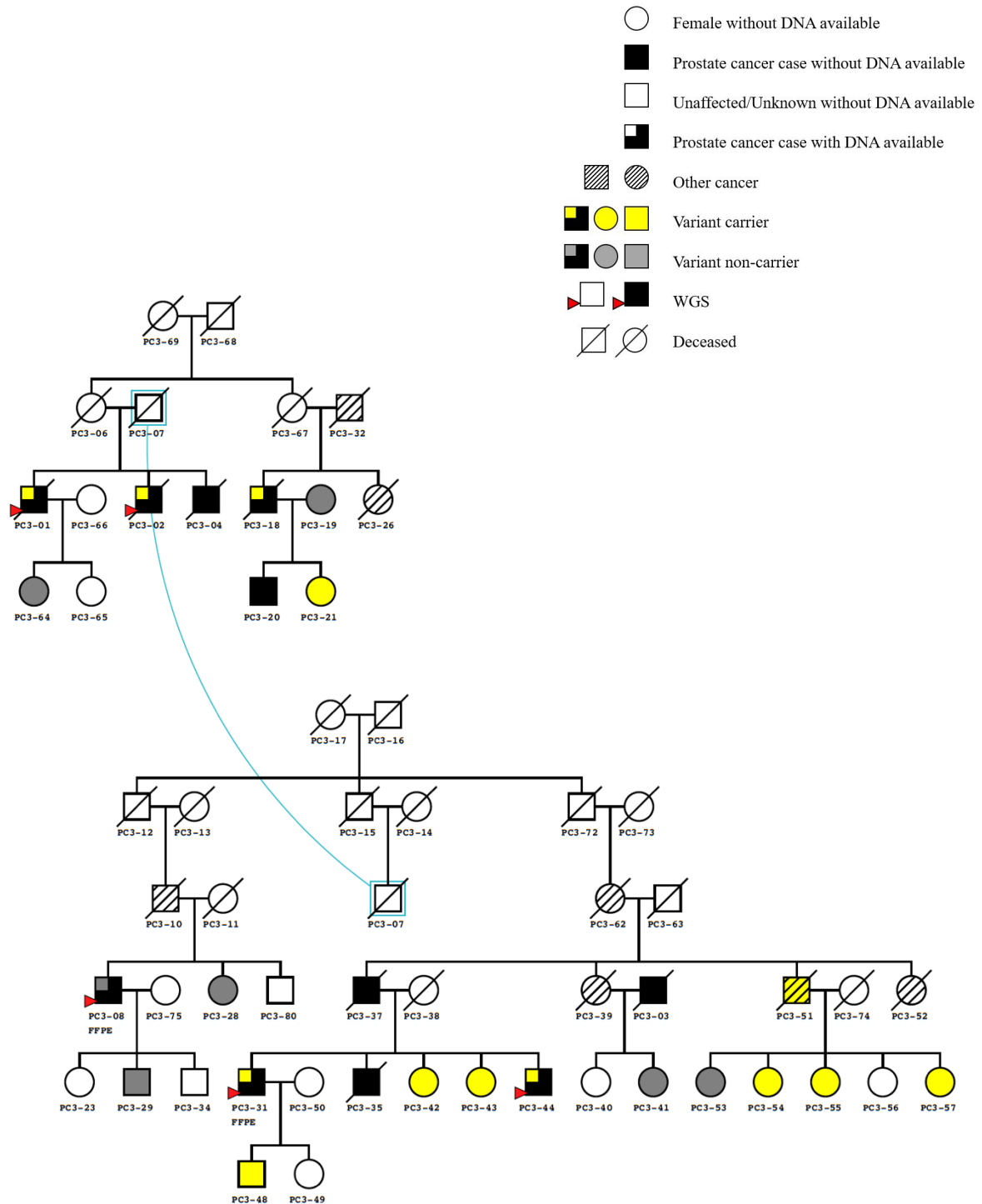
Figure 3.5 *P2RX7* variant carriers in PcTas3.

This is a condensed pedigree of PcTas3 comprising all *P2RX7* variant carriers (shown in yellow) and their relationship. Non-variant carrier family members are shown in grey and the individuals who were WGS are indicated by red arrows.

3.3.3   Rare variant prioritisation in PcTas4

PcTas4 is comprised of 25 PCa cases across four generations (Figure 3.6). A total of five individuals were successfully WGS, including an affected brother pair, an affected uncle/nephew pair (second cousins of the affected brother pair) and an unaffected cousin of these men (Table 3.4 and Figure 3.7). This older unaffected male (76 years of age) was chosen as a 'control' to enable higher prioritisation of variants only found in his affected relatives. However, given that disease-causing variants often exhibit incomplete penetrance, variants present in all sequenced individuals were not excluded completely.
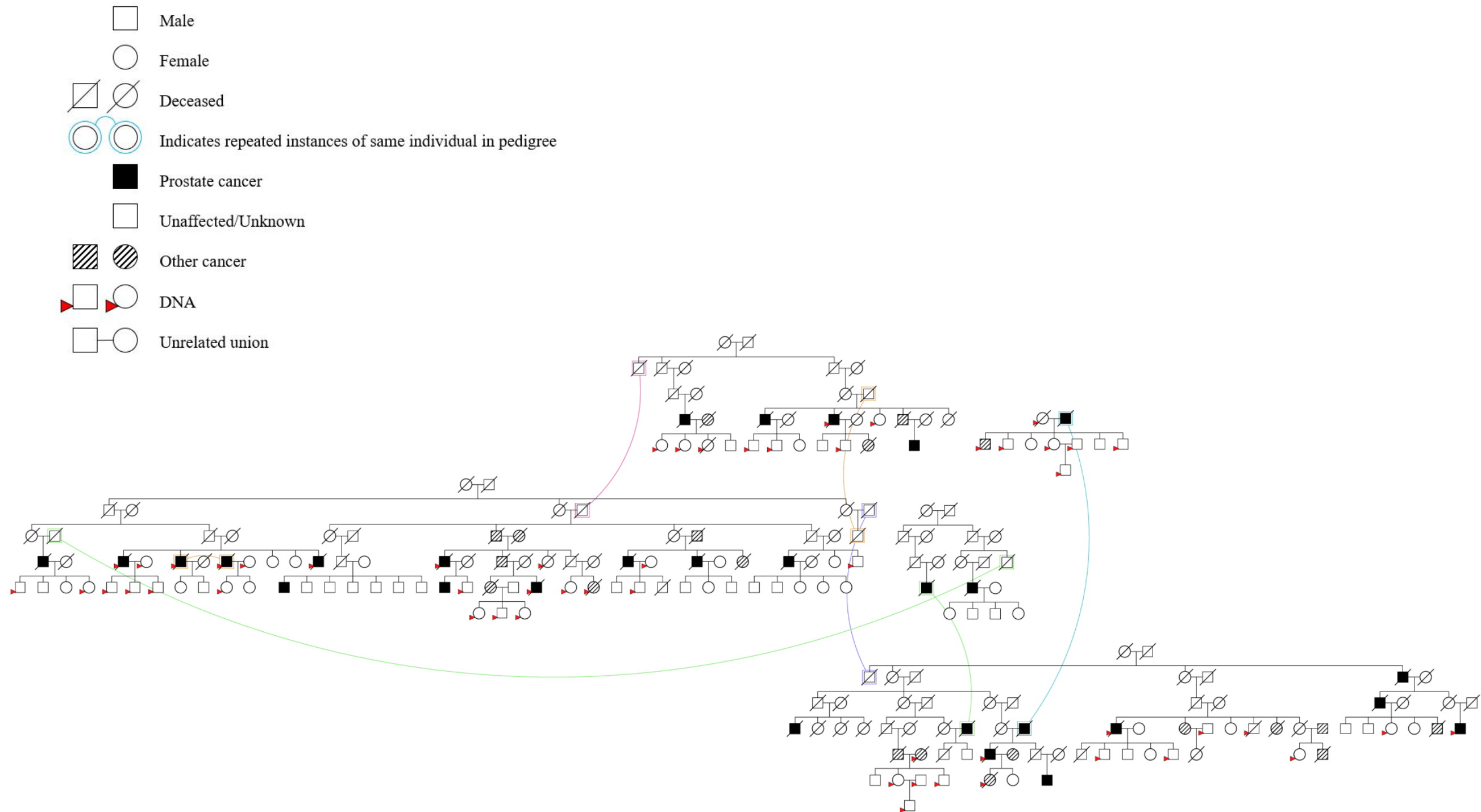
Figure 3.6 PcTas4 pedigree.

PcTas4 pedigree, depicting the number and relationships of PCa cases (shown in shaded squares), as well the availability of DNA from cases and their unaffected relatives, which is represented by red arrows. The disease status for earlier generations is generally unknown, unless this information was obtained from clinical records. And if so, these individuals have been marked as affected in the pedigrees. This pedigree is included to illustrate the size of the pedigree only, please refer to Figure 3.7 and 3.8 for individual annotations.

Table 3.4 Clinicopathological characteristics of individuals from PcTas4 chosen for whole-genome sequencing.

Sample Identification	Sex	Prostate Cancer Affection Status	Age at diagnosis	Tumour Grade[1]	Gleason Score[2]
PC4-01	Male	Affected	60	-	6 (3+3)
PC4-02	Male	Affected	73	-	6 (3+3)
PC4-03	Male	Affected	80	M/PD	7 (4+3)
PC4-95	Male	Affected	66	PD	9 (4+5)
PC4-161	Male	Unaffected	76*	N/A	N/A

*Unaffected, age at WGS; [1]Tumour grade obtained from pathology report; [2]Gleason Score obtained from pathology report; M/PD: moderately-poorly differentiated; PD: poorly differentiated; -: information not present in original pathology report; N/A: not applicable.
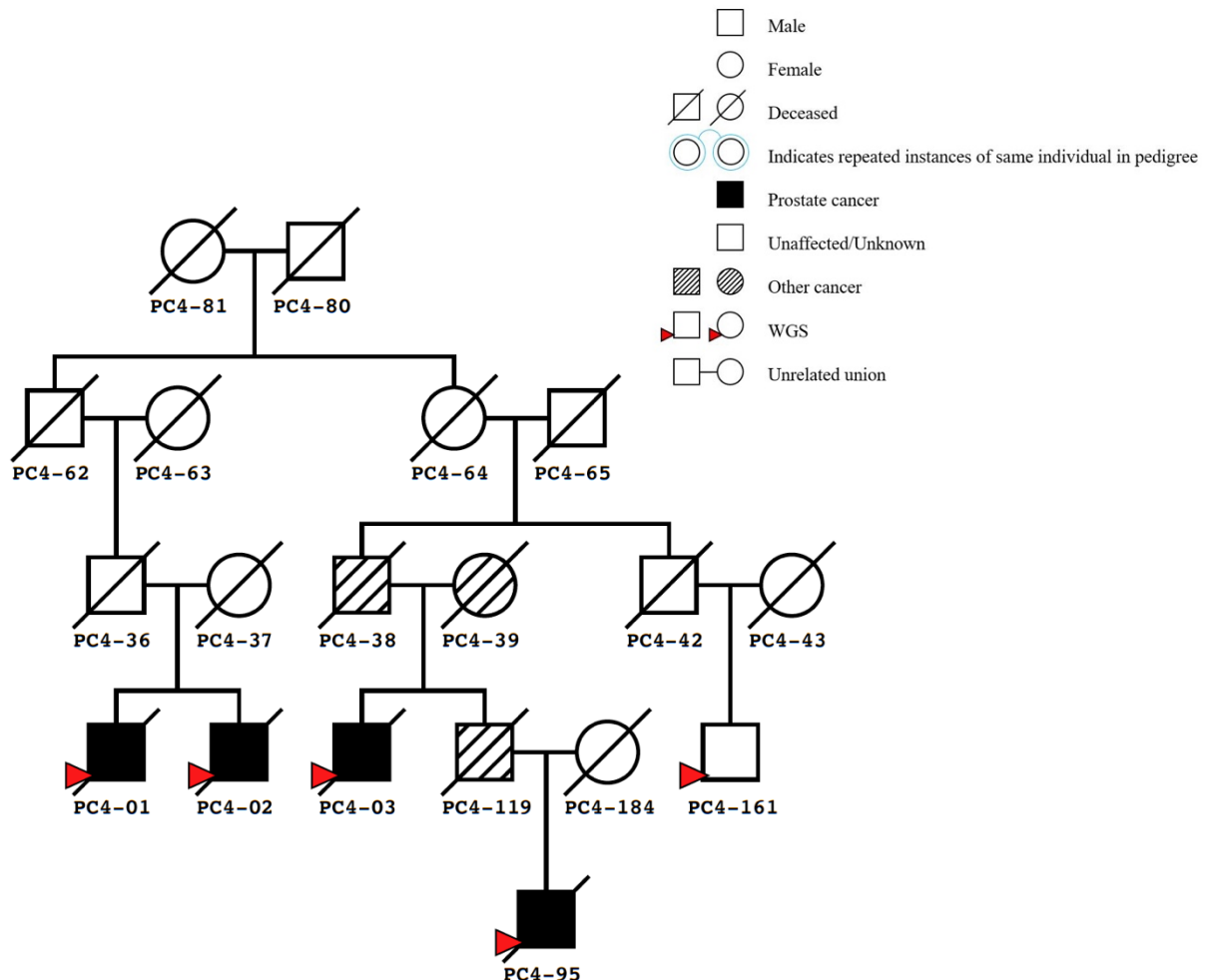
Figure 3.7 A condensed PcTas4 pedigree showing individuals chosen for whole-genome sequencing.

Individuals chosen for WGS are indicated by red arrows, in this case, four PCa cases and one unaffected male relative were chosen.

Five rare variants and one novel variant were prioritised in this family for follow-up studies (Table 3.5). Of these variants, variants in *KMT2C* and *RHPN2* did not validate by Sanger sequencing, indicating potential false positives. Subsequent genotyping of additional family members, including a further six PCa cases and 17 relatives, suggested that only the *ATM* variant segregated with PCa (Figure 3.8). An additional eight *ATM* variant carriers were identified, including three PCa cases. Unaffected men who carried the *ATM* variant are only now approaching the average age of PCa diagnosis. For example, the eldest unaffected man is 66 years of age (PC4-94) and the youngest just 40 years old (PC4-125). The variants in *IRS1, SSH3* and *CRIP2* did not segregate with disease. An additional four, two and three variant carriers were identified, respectively, however they were present more often in unaffected men, and PCa carriers were not within a tight pedigree cluster.

Table 3.5 Rare variants prioritised in the PcTas4 pedigree following whole-genome sequencing of four affected men and one older unaffected man.

Gene	rs number	Chromosome base pair	ExAC[1] MAF (%)	Segregation in WGS Individuals (affected carriers/unaffected carriers)	CADD[2] Score	Allele Change; Amino Acid Change	Number of Control Carriers	ClinVar Search[3]	Validation in WGS Individuals	Segregation in Entire Family
ATM	rs1800057	11:10,814,356	1.69	3 out of 4/ 0 out of 1	27.9	C > G; P1054R	0 out of 8	Hereditary cancer: Benign	Yes	Yes
SSH3	rs373641394	11:67,072,456	0.01	3 out of 4/ 0 out of 1	17.02	G > A; R106K	0 out of 8	Not reported	Yes	No
IRS1	rs41265094	2:227,661,003	0.82	2 out of 4/ 0 out of 1	21.6	C > G; G818R	1 out of 8	Diabetes mellitus type 2: Likely benign	Yes	No
CRIP2	rs375691223	14:105,945,992	0.01	2 out of 4/ 0 out of 1	8.02	C > T; Splice	0 out of 8	Not reported	Yes	No

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; MAF: Minor allele frequency; N/A: Not found in ExAC or ClinVar; WGS: Whole-genome sequenced; [2]CADD: Combined Annotation Dependent Depletion [164]; Control: Control from the *Tasmanian Prostate Cancer Case-Control Study;* eight were WGS; [3]Associated condition: Interpretation of variant [163].

Gene	rs number	Chromosome base pair	ExAC[1] MAF (%)	Segregation in WGS Individuals (affected carriers/unaffected carriers)	CADD[2] Score	Allele Change; Amino Acid Change	Number of Control Carriers	ClinVar Search[3]	Validation in WGS Individuals	Segregation in Entire Family
KMT2C	rs76844681	7:151,932,990	0.99	2 out of 4/ 0 out of 1	35	C > T; R894Q	0 out of 8	Not reported	No	N/A
RHPN2	Novel	19:15,564,233	N/A	2 out of 4/ 0 out of 1	28.8	A > G; V100A	0 out of 8	N/A	No	N/A

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; MAF: Minor allele frequency; N/A: Not found in ExAC or ClinVar; WGS: Whole-genome sequenced; [2]CADD: Combined Annotation Dependent Depletion [164]; Control: Control from the *Tasmanian Prostate Cancer Case-Control Study;* eight were WGS; [3]Associated condition: Interpretation of variant [163].
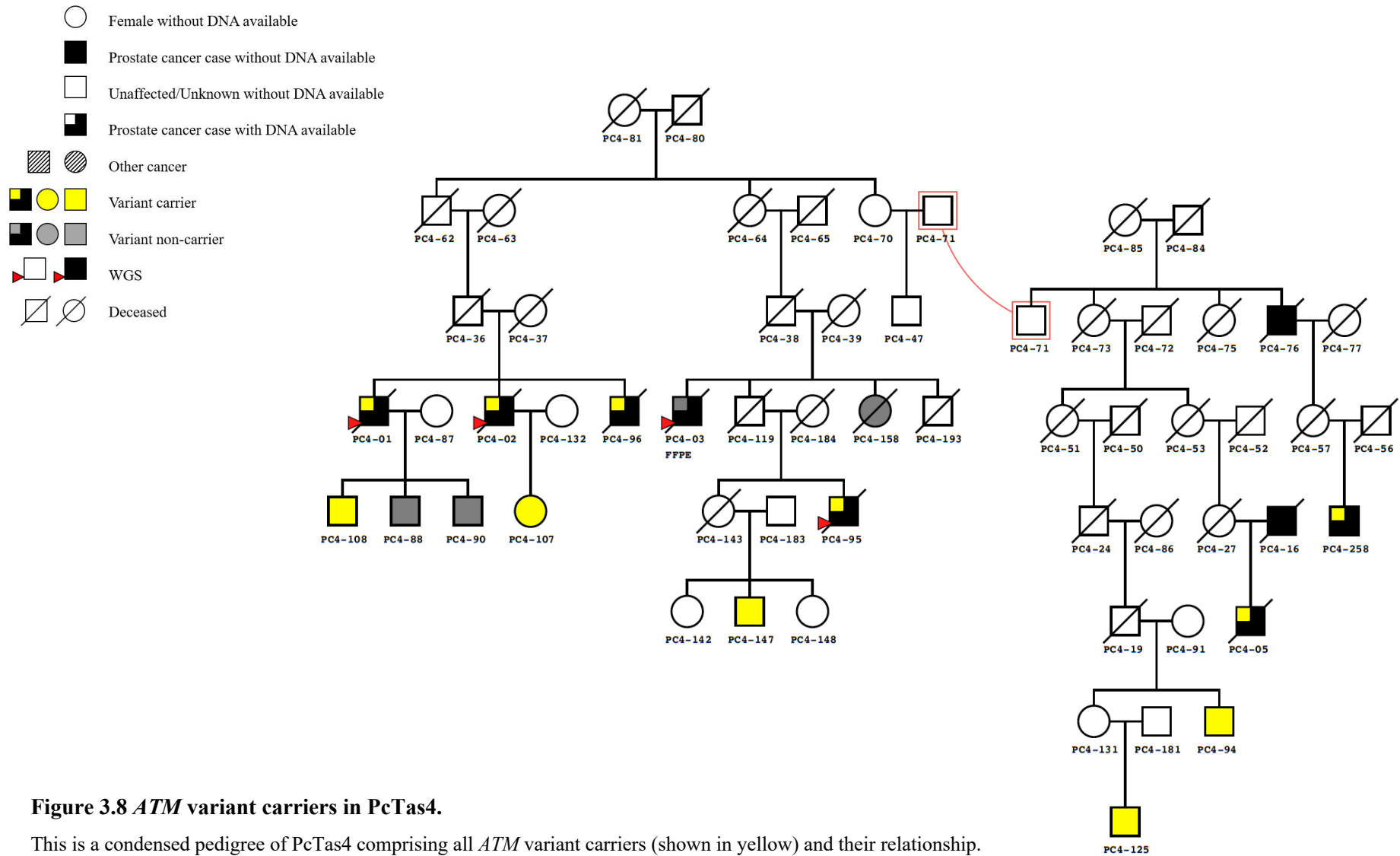
Figure 3.8 *ATM* variant carriers in PcTas4.

This is a condensed pedigree of PcTas4 comprising all *ATM* variant carriers (shown in yellow) and their relationship.

Non-variant carrier family members are shown in grey and the individuals who were WGS are indicated by red arrows.

3.3.4   Rare variant prioritisation in PcTas22

Family PcTas22 is the largest PCa family in the *Tasmanian Familial Prostate Cancer Cohort*, comprising a total of 89 cases of PCa spanning five generations (Figure 3.9). Eleven individuals were successfully WGS in this family, comprising two separate branches (Figure 3.10). Due to the number of individuals sequenced, and the magnitude of data available these two branches were analysed separately, as 'sub' and 'main' pedigree.
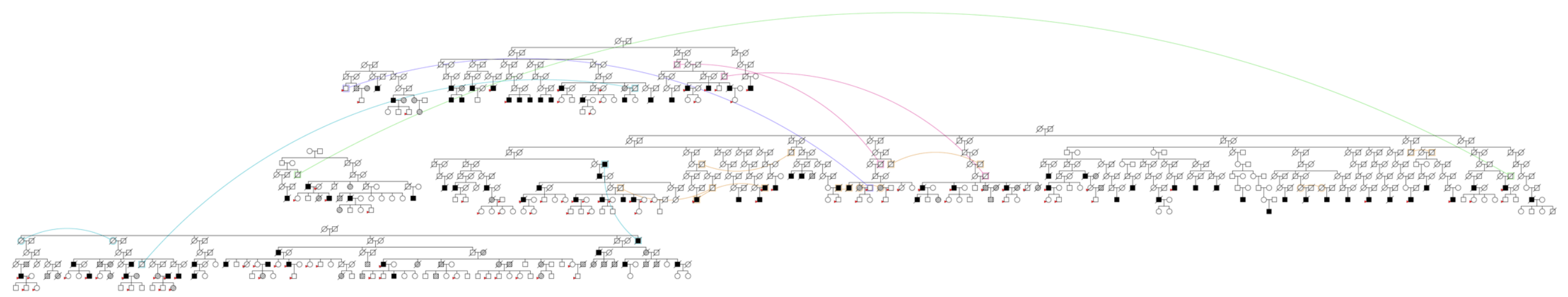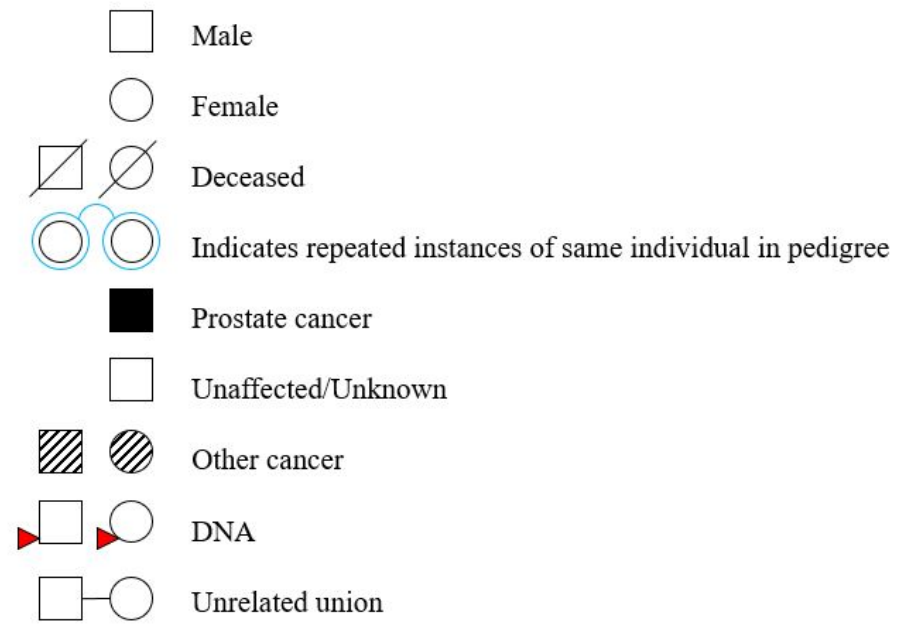
Figure 3.9 PcTas22 pedigree.

PcTas22 pedigree, depicting the number and relationships of PCa cases (shown in shaded squares), as well the availability of DNA from cases and their unaffected relatives, which is represented by red arrows. The disease status for earlier generations is generally unknown, unless this information was obtained from clinical records. And if so, these individuals have been marked as affected in the pedigrees. This pedigree is included to illustrate the size of the pedigree only, please refer to Figure 3.10 and 3.11 for individual annotations.
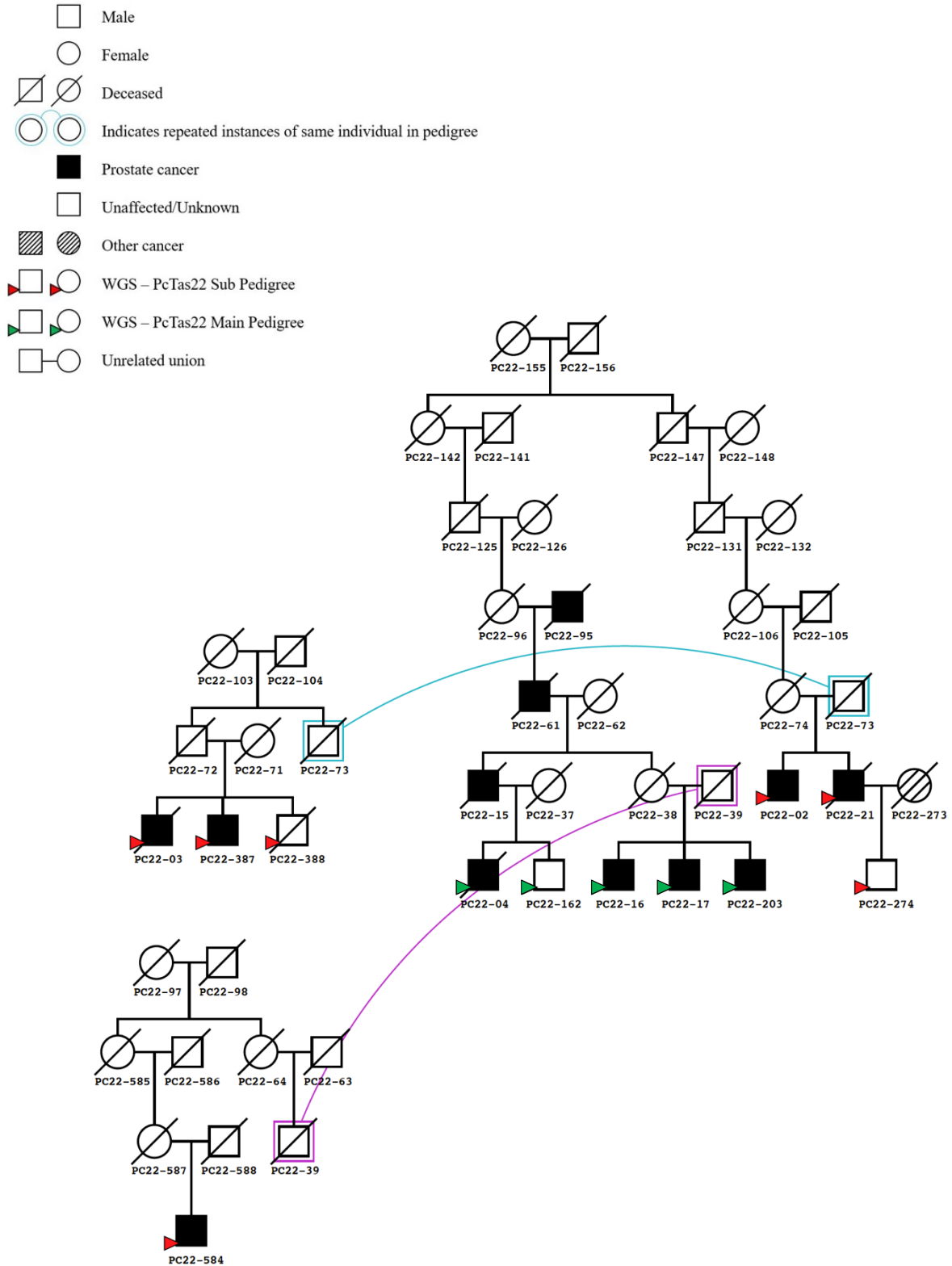
Figure 3.10 A condensed PcTas22 pedigree showing individuals from both branches of the family (sub and main pedigree) chosen for whole-genome sequencing.

Individuals chosen for WGS from the sub pedigree are indicated by red arrows, in this case, five PCa cases and two unaffected male relatives were chosen. Individuals chosen for WGS from the main pedigree are indicated by green arrows, in this case, four PCa cases and one unaffected male relative were chosen.

3.3.5   Rare variant prioritisation in the PcTas22 sub pedigree

Individuals sequenced from the PcTas22 sub pedigree included an affected brother pair and their unaffected older brother (died at 76 years of age), another affected brother pair (first cousins of the other brother pair) and an unaffected son (56 years of age) of one of these affected men (Table 3.6); these are indicated by red arrows in Figure 3.10. Rare variants present in all four affected men and not in the unaffected older brother were prioritised. Variants present in the four affected men and the unaffected son were also considered for further study, given the son is yet to reach the average age of PCa onset (~65 years of age).

Table 3.6 Clinicopathological characteristics of individuals from the PcTas22 sub pedigree chosen for whole-genome sequencing.

Sample Identification	Sex	Prostate Cancer Affection Status	Age at diagnosis	Tumour Grade[1]	Gleason Score[2]
PC22-02	Male	Affected	64	MD	6 (3+3)
PC22-03	Male	Affected	62	WD	-
PC22-21	Male	Affected	69	-	6 (3+3)
PC22-274	Male	Unaffected	56*	N/A	N/A
PC22-387	Male	Affected	83	-	8 (4+4)
PC22-388	Male	Unaffected	76*	N/A	N/A
*Unaffected, age at WGS; [1]Tumour grade obtained from pathology report; [2]Gleason Score obtained from pathology report; WD: well differentiated; MD: moderately differentiated; -: information not present in original pathology report; N/A: not applicable.					

Two rare variants were prioritised in the sub branch of the PcTas22 family (Table 3.7). Both variants were validated by Sanger sequencing, however, sequencing of additional family members, including 16 PCa cases and 17 unaffected relatives, revealed that neither of these segregated with disease. *HSD3B1* was only identified in one additional individual, an unaffected male, therefore with too few carriers it was not prioritised any further. An additional five carriers of the *NAT10* variant were identified, including two PCa cases however, the other three were all unaffected male relatives. Therefore, four out of the 10 carriers were unaffected men, thus the variant did not segregate with disease in this family.

Table 3.7 Rare variants prioritised in the PcTas22 sub pedigree following whole-genome sequencing of four affected men and two older unaffected men.

Gene	rs number	Chromosome: base pair	ExAC[1] MAF (%)	Segregation in WGS individuals (affected carriers/unaffected carriers)	CADD[2] Score	Allele Change; Amino Acid Change	Number of Control Carriers	ClinVar Search[3]	Validation in WGS Individuals	Segregation in Entire Family
HSD3B1	rs4986952	1:120,054,192	0.39	3 out of 4/ 0 out of 2	22.8	G > T; R71I	0 out of 8	Not reported	Yes	No
NAT10	rs72910804	11:34,165,079	1.97	4 out of 4/ 1 out of 2	16.14	A > G; Splice	1 out of 8	Not reported	Yes	No

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; MAF: Minor allele frequency; WGS: Whole-genome sequenced; [2]CADD: Combined Annotation Dependent Depletion [164]; Control: Control from the *Tasmanian Prostate Cancer Case-Control Study;* eight were WGS; [3]Associated condition: Interpretation of variant [163].

3.3.6 Rare variant prioritisation in the PcTas22 main pedigree

Individuals that were sequenced included four affected men and one unaffected older male relative (67 years of age), comprising an affected brother trio and an affected and unaffected brother pair (first cousins of the trio; Table 3.8). These individuals are indicated by green arrows in Figure 3.10. One of the men in each of the brother pair/trio had a relatively early age of disease onset (57 and 56 years, respectively). Rare variants were prioritised if they were present in all four affected men and not in the older unaffected man. However, variants in all five individuals with WGS were considered for follow-up studies, as reduced penetrance of such variants could explain why PC22-162 is also a variant carrier.

Table 3.8 Clinicopathological characteristics of individuals from the PcTas22 main pedigree chosen for whole-genome sequencing.

Sample Identification	Sex	Prostate Cancer Affection Status	Age at diagnosis	Tumour Grade[1]	Gleason Score[2]
PC22-04	Male	Affected	57	MD	6 (3+3)
PC22-16	Male	Affected	74	WD	-
PC22-17	Male	Affected	56	MD	6 (3+3)
PC22-162	Male	Unaffected	67*	N/A	N/A
PC22-203	Male	Affected	79	PD	8 (4+4)
PC22-584	Male	Affected	63	MD	7 (3+4)
*Unaffected, age at WGS; [1]Tumour grade obtained from pathology report; [2]Gleason Score obtained from pathology report; WD: well differentiated; MD: moderately differentiated; PD: poorly differentiated; -: information not present in original pathology report; N/A: not applicable.					

In total, three novel/rare variants were prioritised in the PcTas22 main pedigree WGS data. Subsequent genotyping of an additional 16 PCa cases and 18 relatives found that only the variants in *WNT1* and *RND1* segregated with PCa in PcTas22 (the *CHEK2* variant did not validate in the WGS individuals). Three additional carriers were identified and interestingly, every carrier of either variant, also carried the other. Therefore, Figure 3.11 shows carriers of both the *WNT1* and *RND1* variants. The average age of PCa diagnosis in this branch of the PcTas22 family is 68 years, therefore both PC22-162 and PC22-205 are yet to reach this age (67 and 58 years, respectively).

Table 3.9 Rare variants prioritised in the PcTas22 main pedigree following whole-genome sequencing of five affected men and one older unaffected man.

Gene	rs number	Chromosome: base pair	ExAC[1] MAF (%)	Segregation in WGS individuals (affected carriers/unaffected carriers)	CADD2 Score	Allele Change; Amino Acid Change	Number of Control Carriers	ClinVar Search[3]	Validation in WGS Individuals	Segregation in Entire Family
RND1	Novel	12:49,254,905	N/A	4 out of 5/ 1 out of 1	39	C > A; E110X	0 out of 8	N/A	Yes	Yes
WNT1	Novel	12:49,374,959	N/A	4 out of 5/ 1 out of 1	20.8	G > A; E217K	0 out of 8	N/A	Yes	Yes
CHEK2	rs200432447	22:29,083,962	0.002	5 out of 5/ 1 out of 1	24.4	G > C; R565G	0 out of 8	Hereditary breast and ovarian cancer; Pathogenic	No	N/A

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; N/A: Not found in ExAC or ClinVar, or did not validate therefore, segregation was not assessed; WGS: Whole-genome sequenced; [2]CADD: Combined Annotation Dependent Depletion; Control: Control from the *Tasmanian Prostate Cancer Case-Control Study;* eight were WGS; [3]Associated condition: Interpretation of variant [163].
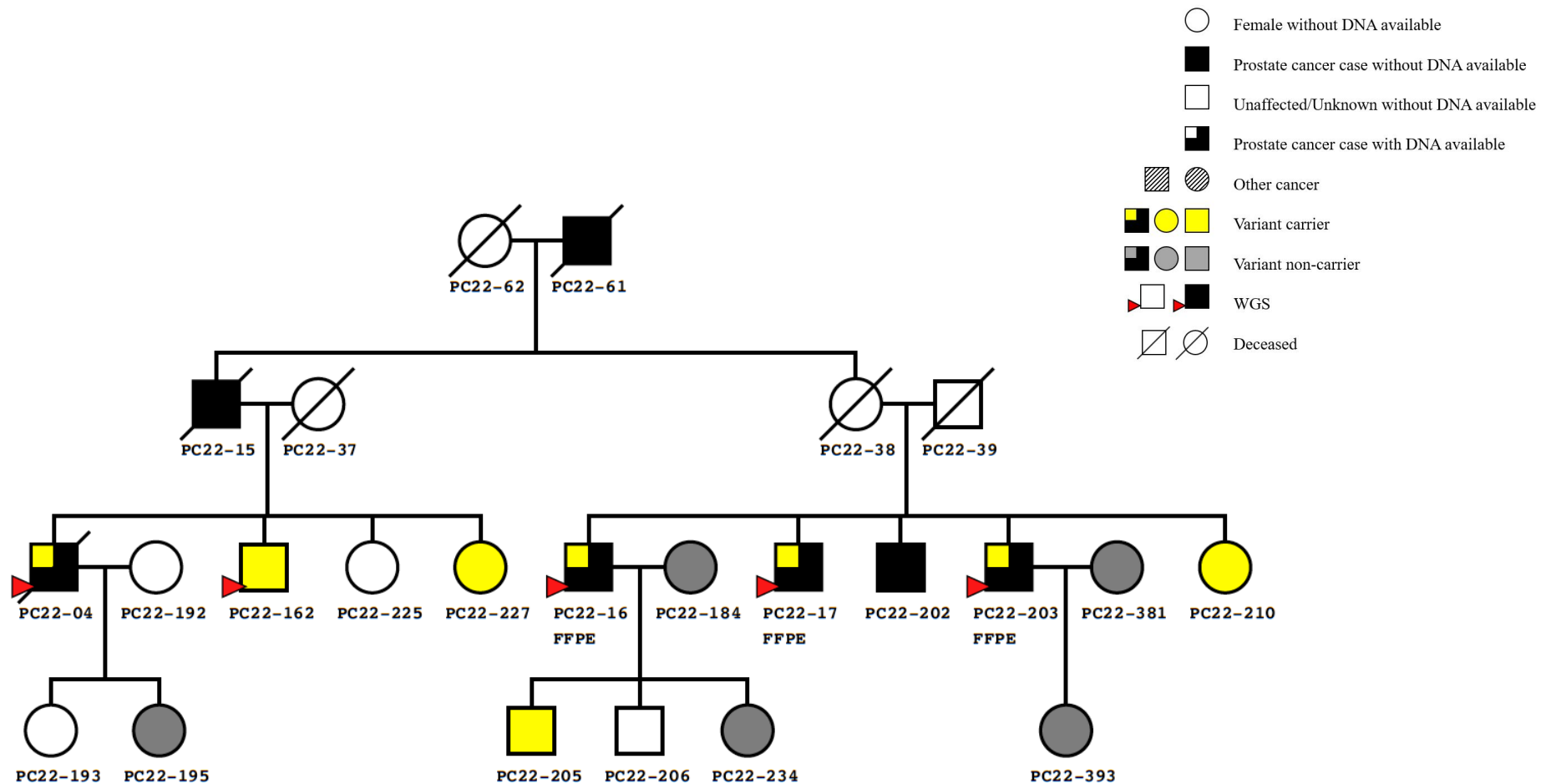
Figure 3.11 *WNT1* and *RND1* variant carriers in the PcTas22 main pedigree.

This is a condensed pedigree of the PcTas22 main pedigree comprising all *WNT1* and *RND1* variant carriers (shown in yellow) and their relationship. Notably, the *WNT1* and *RND1* variants co-segregated together and were not found to contribute to the sub pedigree of PcTas22. Non-variant carrier family members are shown in grey and the individuals who were WGS are indicated by red arrows.

3.3.7 Assessing the possibility of rare variant enrichment in the Tasmanian population

The rare variants which showed evidence of segregation in their founder family were screened in 94 control samples from the *Tasmanian Prostate Cancer Case-Control Study* (described in Chapter 2.1.3). Three out of five of the segregating variants were not found in any of the 94 controls (Table 3.10). The *ATM* variant, rs1800057 was found to have the highest carrier frequency in the controls (3 out of 94). All variants were considered rare enough in the representative Tasmanian population for them to be deemed not enriched. Therefore, the *Tasmanian Familial Prostate Cancer Study* cohorts were genotyped for all five rare segregating variants.

Table 3.10 Screening of the rare segregating variants in 94 controls from the *Tasmanian Prostate Cancer Case-Control Study*.

Family Identification	Gene	Variant	Number of Control Carriers
PcTas3	*CCL26*	rs41463245	0 out of 94
	P2RX7	rs28360447	2 out of 94
PcTas4	*ATM*	rs1800057	3 out of 94
PcTas22	*RND1*	Novel; E110X	0 out of 94
	WNT1	Novel; E217K	0 out of 94
Control: Control from the *Tasmanian Prostate Cancer Case-Control Study.* In total, 94 controls were genotyped for the prioritised rare variants by Sanger sequencing.			

3.3.8 Association of the prioritised rare variants with prostate cancer risk in Tasmania

Validated rare variants, which segregated with PCa in their founder family, and were considered to be not enriched in Tasmania, were chosen for high-throughput genotyping screens, using a TaqMan assay. The *Tasmanian Familial Prostate Cancer Cohort* (n=714) and the *Tasmanian Prostate Cancer Case-Control Study* (n=853) were screened for these variants. To increase the number of individuals available, pathology specimens for cases from variant carrier families, where germline DNA from blood or saliva was not available were also genotyped. The carrier frequency (%) of each variant was determined for the familial PCa cohort and the case-control study, and M_{QLS} analysis [2] of the two datasets was used to calculate

the OR and p-value, where <0.05 was considered statistically significant. The results of these analyses are shown in Table 3.11.

Following genotyping of all available DNA in our resource, M_{QLS} analysis [2] of the combined familial and case-control genotyping data demonstrated a significant association between PCa risk and two variants in the Tasmanian population (Table 3.11). This included the two co-segregating novel variants in *RND1* and *WNT1* identified in the main pedigree of PcTas22 (OR=6.21, p=0.0001; OR=7.81, p=5.01×10^{-6}, respectively). The variants in *CCL26, P2RX7* and *ATM* were not found to be statistically associated with PCa risk in our Tasmanian resource by M_{QLS} analysis.

Table 3.11 The association of the prioritised rare variants with prostate cancer risk in the *Tasmanian Familial Prostate Cancer Study* cohorts.

Gene	Variant	Founder Family	Other PcTas Families	Familial Case Carriers (n=251)[1]	Familial Unaffected Carriers (n=463)[1]	Sporadic Case Carriers (n=498)[1]	Control Carriers (n=341)[1]	ExAC[2] MAF (%)	Odds Ratio	p-value
CCL26	rs41463245	PcTas3	1, 9, 63, 72, 100	8 (3.21%)	11 (2.47%)	10 (2.02%)	7 (2.06%)	0.86	1.54	0.26
P2RX7	rs28360447	PcTas3	1, 9, 11, 12, 19, 22, 23, 63, 65, 837, 3255	14 (5.62%)	30 (6.77%)	18 (3.64%)	8 (2.35%)	1.27	1.84	0.22
ATM	rs1800057	PcTas4	1, 4, 9, 11, 12, 16, 22, 34, 38, 55,63, 65, 72, 100, 213	25 (9.96%)	31 (6.94%)	25 (5.02%)	18 (5.29%)	1.69	0	0.84
RND1	Novel; E110X	PcTas22	Nil	4 (1.66%)	4 (0.86%)	2 (0.40%)	0 (0%)	N/A	6.21	0.0001*
WNT1	Novel; E217K	PcTas22	Nil	4 (1.66%)	4 (0.86%)	2 (0.40%)	0 (0%)	N/A	7.81	5.01×10^{-6}*

Familial case and familial unaffected comprise the *Tasmanian Familial Prostate Cancer Cohort*; Sporadic case and control comprise the *Tasmanian Prostate Cancer Case-Control Study;* [1](n=total sample size); [2]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; MAF: minor allele frequency; N/A: Not found in ExAC; *Significant p-value.

Three of the five rare variants (other two are novel) were assessed for enrichment in groups within the *Tasmanian Familial Prostate Cancer Study* cohorts, as well as in comparison to the ExAC database and our Tasmanian controls. Table 3.12 shows that the *CCL26* variant was not enriched in any of our Tasmanian groups, or compared to ExAC. The *P2RX7* variant was enriched in the Tasmanian familial PCa cases *versus* ExAC (p=0.03), plus it was enriched in all PCa cases within our resource (familial & sporadic) compared to our population controls (p=0.02). This difference was still noticeable when comparing carrier status between just the Tasmanian familial cases and population controls (p=0.006), however not between the sporadic cases and population controls (p=0.28). The *ATM* variant appeared to be more frequent in the Tasmanian population compared to ExAC; all comparisons were significant, including the Tasmanian population control carrier frequency *versus* ExAC. The variant was also enriched in the Tasmanian familial cases compared to the Tasmanian controls (p=0.04), but not in the sporadic cases compared to the controls or the familial cases.

Table 3.12 Comparison of variant carrier status in the *Tasmanian Familial Prostate Cancer Study* cohorts compared to ExAC or Tasmanian controls.

Gene	Variant		Entire Resource *versus* ExAC[1]	Familial & Sporadic Cases *versus* ExAC[1]	Familial Cases *versus* ExAC[1]	Sporadic Cases *versus* ExAC[1]	Controls *versus* ExAC[1]	Familial & Sporadic Cases *versus* Controls	Familial Cases *versus* Controls	Sporadic Cases *versus* Controls
CCL26	rs41463245 (ExAC[1] MAF 1.34%)	Chi square;1df p-value	3.28 (-)[2] p=0.07	2.91 (-) p=0.09	2.02 (-) p=0.16	1.05 (-) p=0.31	0.60 (-) p=0.44	0.04 (-) p=0.85	0.10 (-) p=0.75	0.002 (-) p=0.96
		Number of carriers (n=total sample size)	36 (n=1,529) *versus* 364 (n=27,173)	18 (n=744) *versus* 364 (n=27,173)	8 (n=249) *versus* 364 (n=27,173)	10 (n=495) *versus* 364 (n=27,173)	7 (n=340) *versus* 364 (n=27,173)	18 (n=744) *versus* 7 (n=340)	8 (n=249) *versus* 7 (n=340)	10 (n=495) *versus* 7 (n=340)
P2RX7	rs28360447 (ExAC[1] MAF 1.85%)	Chi square; 1df p-value	0.65 (+)[2] p=0.42	2.50 (+) p=0.11	4.87 (+) p=0.03*	0.004 (-) p=0.95	2.54 (-) p=0.11	5.13 (+) p=0.02*	7.48 (+) p=0.006*	1.16 (+) p=0.28
		Number of carriers (n=total sample size)	70 (n=1,528) *versus* 500 (n=27,101)	32 (n=744) *versus* 500 (n=27,101)	14 (n=249) *versus* 500 (n=27,101)	18 (n=495) *versus* 500 (n=27,101)	8 (n=341) *versus* 500 (n=27,101)	32 (n=744) *versus* 8 (n=341)	14 (n=249) *versus* 8 (n=341)	18 (n=495) *versus* 8 (n=341)

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; Entire Resource includes the *Tasmanian Familial Prostate Cancer Cohort* and the *Tasmanian Prostate Cancer Case-Control Study*; Familial cases are a part of the *Tasmanian Familial Prostate Cancer Cohort*; Sporadic case and control comprise the *Tasmanian Prostate Cancer Case-Control Study*; [2]In the chi square test (+/-) indicates directionality, where (+) means the minor allele frequency is greater in the first named population *versus* the comparison dataset, whereas, (-) indicates it is more enriched in the second named population; *Significant p-value.

Gene	Variant		Entire Resource *versus* ExAC[1]	Familial & Sporadic Cases *versus* ExAC[1]	Familial Cases *versus* ExAC[1]	Sporadic Cases *versus* ExAC[1]	Controls *versus* ExAC[1]	Familial & Sporadic Cases *versus* Controls	Familial Cases *versus* Controls	Sporadic Cases *versus* Controls
ATM	rs1800057 (ExAC[1] MAF 1.32%)	Chi square; 1df p-value	26.90 (+)[2] p=2.14x10^{-7}*	23.13 (+) p=1.52x10^{-6}*	19.40 (+) p=1.06x10^{-5}*	5.63 (+) p=0.02*	4.62 (+) p=0.03*	1.13 (+) p=0.29	4.25 (+) p=0.04*	0.03 (-) p=0.87
		Number of carriers (n=total sample size)	99 (n=1,536) *versus* 356 (n=27,084)	50 (n=749) *versus* 356 (n=27,084)	25 (n=251) *versus* 356 (n=27,084)	25 (n=498) *versus* 356 (n=27,084)	18 (n=340) *versus* 356 (n=27,084)	50 (n=749) *versus* 18 (n=340)	25 (n=251) *versus* 18 (n=340)	25 (n=498) *versus* 18 (n=340)

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; Entire Resource includes the *Tasmanian Familial Prostate Cancer Cohort* and the *Tasmanian Prostate Cancer Case-Control Study*; Familial cases are a part of the *Tasmanian Familial Prostate Cancer Cohort*; Sporadic case and control comprise the *Tasmanian Prostate Cancer Case-Control Study;* [2]In the chi square test (+/-) indicates directionality, where (+) means the minor allele frequency is greater in the first named population *versus* the comparison dataset, whereas, (-) indicates it is more enriched in the second named population; *Significant p-value. The *WNT1* and *RND1* variants are novel therefore, a comparison with ExAC cannot be made.

3.4 DISCUSSION

3.4.1 Rare variants in *CCL26* and *P2RX7* as potential prostate cancer risk variants

Overall, three rare variants and one novel variant were identified in the PcTas3 pedigree following WGS of five affected men. The variants in *NDE1* and *CLDN4* did not segregate with disease, however the variants in *CCL26* and *P2RX7* did. These two variants were initially prioritised as four of the five affected men in PcTas3 were identified as carriers and, particularly interesting, they were both predicted to be in the top 0.1% of most deleterious to protein function variants in the human genome [161].

CCL26 participates in the promotion of cancer progression in liver and colorectal cancer [165,166], yet the *CCL26* variant has not previously been associated with any disease, as per ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) [163]. Wild-type CCL26 is a 94 amino acid protein, whereas the W44X variant causes a premature stop codon, which results in a small mutant protein of only 44 amino acids [167]. However, it would have to be speculated as to whether this mutant protein is actually functionally active. The variant lies within a chemokine domain that is important for receptor regulator activity and binding of other molecules [168], which may indicate that a premature stop codon could alter these interactions.

P2RX7 plays a role in infection and inflammation and is highly expressed in tumour cells [169,170]. The prioritised variant has not previously been associated with cancer, however has been found to be associated with primary gout and hyperuricemia susceptibility [171], yet is not reported in ClinVar [163]. The variant amino acid is larger and more basic compared to the small, neutral wild-type amino acid, which could cause the structure and function of the P2RX7 protein to be altered [168]. The variant residue is located in a domain that is responsible for ATP binding, ion channel activity and purinergic nucleotide receptor activity [168], thus the variant may affect these functions.

Overall, neither the *CCL26* or *P2RX7* variants were found to be associated with PCa risk in the Tasmanian population. In fact, the *CCL26* variant was not enriched in any of our patient groups; providing no evidence that the variant is associated with PCa. The carrier frequency of the *P2RX7* variant was significantly higher in Tasmanian PCa cases compared to controls. The variant was found to be enriched in the Tasmanian familial cases compared to our control population (p=0.006). However, this enrichment was not apparent when comparing Tasmanian

sporadic cases to controls (p=0.28). These analyses suggest that there may be a link between the *P2RX7* variant and inherited PCa predisposition. Overall, the familial unaffected individuals had a higher *P2RX7* carrier frequency compared to the familial PCa cases, which may underpin the lack of association with PCa risk, as per the M_{QLS} analysis. These individuals were only included in the enrichment analyses as part of the 'entire resource' group, and this type of analysis doesn't take into account the fact that related individuals are more likely to carry the variant. The enrichment analysis findings and high carrier frequency of the *P2RX7* variant in the *Tasmanian Familial Prostate Cancer Cohort* suggests further investigation in larger familial PCa cohorts is warranted, to establish whether this association can be replicated.

3.4.2   Prioritisation of a rare variant in *ATM*, a known prostate cancer predisposition gene

Six rare variants were prioritised in individuals from PcTas4, including three predicted to be in the top 1% of most deleterious coding variants in the genome, a splice variant and two variants identified in three out of four affected men. The variants in *KMT2C* and *RHPN2* did not validate and the variants in *IRS2, SSH3* and *CRIP2* did not segregate with disease. Thus, the highest prioritised variant in PcTas4 was rs1800057 in *ATM*, which was identified in three out of the four PCa cases and predicted to be deleterious to protein function, with a CADD score of 27.9 [161].

ATM is a DNA repair gene which is responsible for recognising damaged or broken DNA strands, but it also controls the rate at which cells grow and divide [116]. *ATM* is associated with an increased risk of familial breast, pancreatic and PCa, and is included on a number of gene screening panels, including the commonly used BROCA (breast and ovarian cancer associated) gene panel [172-176]. The *ATM* variant identified here was recently recognised as one of the latest PCa susceptibility loci, following a GWAS meta-analysis [9]. Notably, ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) reports the variant to be benign/likely benign [163]. The variant itself results in the substitution of a neutral amino acid with a larger, basic amino acid, which could affect the structure of ATM, potentially resulting in the inability to recognise damaged DNA [168]. In fact, the wild-type amino acid is a proline, which is known to have a very rigid structure, sometimes forcing the backbone into a specific conformation [168], thus, it is possible that the variant may disturb this local structure. It also lies within the serine/threonine-protein kinase domain, which is responsible for the main activity of the protein, including molecular function, and transferase and catalytic activity [168].

The *ATM* variant segregated with disease and was found to be present in a number of other families in our Tasmanian PCa resource. However, the M$_{QLS}$ analysis OR was undefined for this variant because the frequency of the variant allele was too common in the Tasmanian control population (OR=0). Enrichment analysis identified the variant to be enriched in all of our *Tasmanian Familial Prostate Cancer Study* resources compared to ExAC (including our Tasmanian control population), which is consistent with the observed higher frequency. There was also enrichment in the Tasmanian familial cases compared to controls; one would expect that a frequent rare variant in Tasmania would result in an enrichment in familial PCa cases compared to controls, given their relatedness, but the enrichment analysis does not take in to account the relatedness of family members. It is likely that the high frequency of the rare variant is due to the fact that Tasmania was established from a small founder population. Additional investigation of this variant and its possible association with PCa risk is warranted, as our findings suggest that the rs180057 variant plays a role in PCa risk. Given that this variant was also recently identified in a large GWAS meta-analysis [9] illustrates the utility of our family-based approach to rare variant prioritisation. Therefore, it would be worthwhile to look for other variants that have a similar pattern in our Tasmanian cohort.

3.4.3   The identification of novel, co-segregating variants in *RND1* and *WNT1*

A total of eleven individuals, encompassing two separate branches of PcTas22 were WGS. Two rare variants were identified in the PcTas22 sub pedigree, including *HSD3B1* and *NAT10*, yet neither segregated in the entire PcTas22 family. Three variants were prioritised in the PcTas22 main pedigree, including a variant in *CHEK2,* which did not validate, and novel variants in *RND1* and *WNT1*. Both novel variants were initially prioritised because they were carried by four out of five affected men. The *RND1* was predicted to have a deleterious effect on protein function; it is in the top 0.1% of all damaging variants in the genome (CADD=39), and the *WNT1* variant is in the top 1%.

RND1, a Rho GTPase, is known to promote the growth and migration of cancer cells. *RND1* expression is upregulated in oesophageal squamous cell carcinoma [177] and it is said to confer a malignant hepatocellular carcinoma phenotype with a poor prognosis [178]. Little is known about the role of *RND1* and its associated mutations in PCa development, however, increased expression is a prognostic signature in glioblastoma [179] and it promotes growth and migration of cancer cells [177,180]. The novel *RND1* variant causes a premature stop codon at position 110 of the protein, whereas wild-type RND1 has a stop codon at amino acid position 233. This

could result in the production of a truncated protein with increased activity, which could affect the function of RND1 [167]. The variant residue is located in a domain that is important for binding of other molecules including ions, nucleotides and nucleosides, which is important to sustain the proteins molecular, and catalytic and hydrolase activity. It is in contact with residues in other domains that are important for the activity of the protein and binding of other residues [168].

Wnt family member 1 (WNT1) is a Wnt signalling transduction pathway protein that is involved in the regulation of gene transcription, cytoskeleton formation and calcium levels within the cell [181]. This pathway is involved in embryonic development; controlling body axis patterning, cell fate specification, proliferation and migration [181]. Wnt signalling is also involved in carcinogenesis, with its clinical importance demonstrated by the identification of mutations that lead to various diseases, including breast and PCa [182,183]. Chen and colleagues (2004) also concluded that high levels of *WNT1* is associated with advanced, metastatic, hormone-refractory PCa, as they identified low levels in normal prostate cells compared to high levels in malignant cells [184]. The *WNT1* variant in this study causes the acidic wild-type glutamic acid residue at position 217 to be mutated to a basic, larger, lysine residue, which may affect protein folding, as the change in charge may cause repulsion with other residues in the protein or ligands [168]. The variant is located in the signalling receptor binding domain, which is important for binding of other molecules [168].

M_{QLS} analysis found that each of the variants were significantly associated with PCa in our Tasmanian resource (*RND1:* OR=6.21, p=0.0001; *WNT1:* OR=7.81, p=5.01×10^{-6}). Given these are previously undescribed variants, we were unable to test for enrichment of the *RND1* and *WNT1* variants within our *Tasmanian Familial Prostate Cancer Study* cohorts, or in comparison to ExAC. One interesting finding is that the *RND1* and *WNT1* variants were predominately identified in individuals from PcTas22 (with one additional sporadic carrier) and were co-inherited in every instance. Additionally, we tried to link the sporadic carrier in to this family, but to date, we cannot find a common ancestor. Co-inheritance of these novel variants suggest that they exist on a shared haplotype. A preliminary look at the variants that occur between these two genes, including nine common and one rare variant, revealed that variant carriers do have the same genotypes, suggesting a shared haplotype.

3.5 FUTURE DIRECTIONS

Currently the *RND1* and *WNT1* variants appear to be private to PcTas22, yet one sporadic case carrier was identified. WGS of this individual would enable us to genetically link this person to the family (if they are related). If they are, it would appear that the region of chromosome 12 between these two variants is linked to PCa risk in this family. If they are not, it is possible that there may be some linkage disequilibrium (LD) at the population level. This means that alleles at variants positioned close together on the same chromosome tend to occur together more often than is expected by chance [185]. The region between the two variants is large (~120kb) and therefore would represent an unusually large shared haplotype, however it is possible that LD may explain why these variants are co-inherited in these individuals. As these variants are previously undescribed, we aim to further explore the contribution of this variant to independent PCa populations through collaboration with members of the International Consortium of Prostate Cancer Genetics (ICPCG) and the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium. The functional impact of the *RND1* and *WNT1* variants should also be investigated by assessing gene and protein expression in FFPE samples from PcTas22 case carriers and a random selection of non-carriers. Whilst these novel variants appear to be private to a single Tasmanian family, given the function of these genes it would be prudent to screen the *Tasmanian Familial Prostate Cancer Study* resources for other possible disease-causing variants in these genes.

3.6 CONCLUSION

This chapter detailed the WGS of 18 familial PCa cases and four unaffected male relatives from three Tasmanian PCa families, PcTas3, 4 and 22. Altogether, 15 variants were prioritised, 12 validated and five segregated with disease in their founder families. Enrichment analysis suggested that the rare variants in *P2RX7* (rs28360447) and *ATM* (rs1800057) may be linked with inherited PCa predisposition, given the significantly higher carrier frequency in familial cases compared to sporadic cases. Yet, only the novel variants in *RND1* and *WNT1* were found to be significantly associated with PCa risk by M_{QLS} analysis (OR=6.21, p=0.0001; OR=7.81, p=5.01x10^{-6}, respectively). Neither of these variants have been previously described, however both genes have clear biological links to prostate biology and cancer development. Notably, the variants were only identified in one Tasmanian PCa family, and in every instance were co-inherited, suggesting a shared haplotype. Overall, this chapter highlights, firstly, that the study of families with a dense aggregation of disease can yield the identification of rare and novel

disease-associated variants by WGS. This is especially so, given Tasmania is a relatively homogenous population with reduced allelic variability and extended LD. Secondly, this study further supports the hypothesis that rare genetic variants do contribute to PCa risk and they do explain some of the 'missing' portion of known disease heritability.

CHAPTER 4 : IDENTIFICATION AND FUNCTIONAL ASSESSMENT OF A RARE PROSTATE CANCER RISK VARIANT IN *EZH2*

4.1   INTRODUCTION

Targeted and agnostic approaches to rare variant discovery has facilitated advances in recent years in the field prostate cancer (PCa) genetics. As discussed in Chapter 1.8, linkage analysis and targeted sequencing strategies were utilised to identify the rare PCa risk variant in *HOXB13* [11]. Replication studies have provided further evidence for an association with PCa risk, however functional studies are required to demonstrate how it plays a role in disease initiation. The importance of PCa families in identifying rare variants has been realised in recent studies by FitzGerald *et al.* (2013) and Karyadi *et al.* (2017), as discussed in Chapter 3.1, however such studies have been few. The premise on which these studies are based is that rare variants contribute to common disease, and they are enriched in families, which make the search for disease-causing variants easier, given there is reduced genetic complexity [122]. Using a similar approach, we selected a *Tasmanian Familial Prostate Cancer Cohort* family, PcTas12 for whole-genome sequencing (WGS), given that it comprises confirmed PCa cases across four generations and includes multiple father/son pairs, affected grandfather/father/son trios and affected brother pairs/trios (Figure 4.1).

4.2   METHODS

4.2.1   Whole-genome sequencing analysis

Three individuals from PcTas12 were WGS on the Illumina HiSeq X^{TM} Ten platform, as per Chapter 3.2.1. The data were analysed, annotated and variants called as previously described in Chapter 3.2.2.
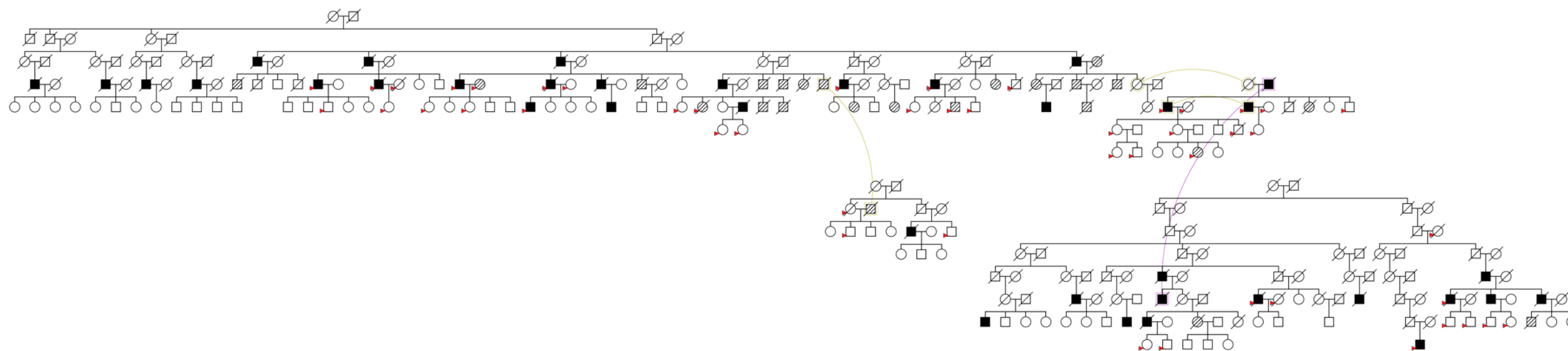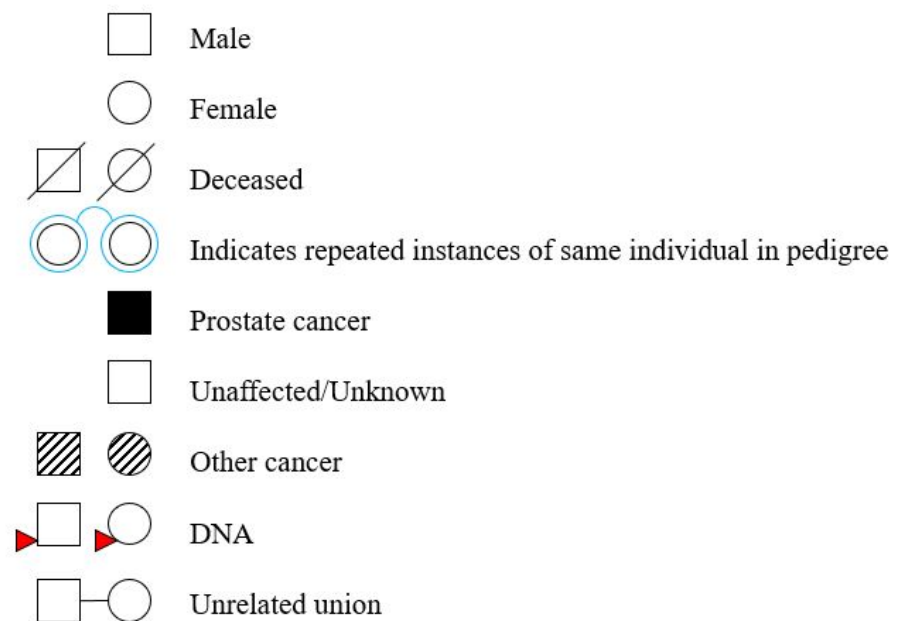
Figure 4.1 PcTas12 pedigree.

PcTas12 pedigree, depicting the number and relationships of PCa cases (shown in shaded squares), as well the availability of DNA from cases and their unaffected relatives, which is represented by red arrows. The disease status for earlier generations is generally unknown, unless this information was obtained from clinical records. And if so, these individuals have been marked as affected in the pedigrees. This pedigree is included to illustrate the size of the pedigree only, please refer to Figure 4.4 and 4.5 for individual annotations.

4.2.2 Validation, segregation and association analysis of prioritised rare variants

Variants prioritised from the WGS data were validated by Sanger sequencing. Sanger sequencing of additional family members was used to track segregation with disease, which is discussed in Chapter 3.2.3. For those cases without gDNA available, FFPE DNA was sequenced (Appendix 2). If found to segregate, rare variants were screened in the entire *Tasmanian Familial Prostate Cancer Cohort* and the *Tasmanian Prostate Cancer Case-Control Study*, as discussed in Chapter 3.2.4 (Appendix 6). M_{QLS} analysis [2] was used to determine if there was an association between the prioritised rare variants and PCa risk in the Tasmanian population, as discussed in Chapter 3.2.5.

4.2.3 Quantification of gene expression

EZH2 (ENST00000492143.1) gene expression in PCa cell lines, prostate needle biopsies and formalin-fixed paraffin embedded (FFPE) prostate tumour samples was initially assessed in exon 17 of the gene by RT-qPCR analysis, as per Chapter 2.3. Exon-level expression across a number of regions of *EZH2,* including exon 4/5, 8/9, 12/14, 14/16, 17/18 and 20/21 was also examined in these samples. Average *EZH2* expression was calculated by averaging the absolute expression of these six regions. The expression of *EZH2* target genes, *CDH1* (ENST00000261769.5)*, HOXA9* (ENST00000343483.6) and *MSMB* (ENST00000358559.2), as well as splicing factors, *SF3B1* (ENST00000424674.1)*, SF3B3* (ENST00000291552.4) and *U2AF1* (ENST00000291552.4) was also determined in these samples. RT-qPCR primers were designed to the most commonly transcribed isoform in the prostate (as per GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2; https://gtexportal.org/home/)) [137] and are displayed in Appendix 3.

4.2.4 Plasmid and transformation of the *EZH2* insert into competent prostate cancer cells

The pSpliceExpress plasmid was a gift from Stefan Stamm (Addgene plasmid #32485; https://www.addgene.org/32485/) and has been described previously by Kishore and colleagues (2008) [186] (Figure 4.2). Primers were designed to amplify a region of the *EZH2* gene surrounding the intronic splice variant, including exons 16-19 and up to 200bp of the surrounding introns (Figure 4.3A; Appendix 7). The insert was prepared from genomic DNA of an *EZH2* variant carrier (PC12-132) by standard PCR using a proofreading DNA Polymerase (Phusion® High-Fidelity with GC Buffer, New England Biolabs®; Appendix 1). A nested-PCR step was used to add the attB1 and attB2 attachment sites to the insert, as shown

in Figure 4.3A, and the product was purified by gel extraction (QIAGEN). The insert was mixed with 150ng/μL of the pSpliceExpress vector in a boiling point (BP) recombination reaction (Figure 4.3B; ThermoFisher Scientific). StrataClone Stratapack Competent cells (Agilent) were transformed and plated on Ampicillin-supplemented LB plates pre-warmed at 37°C. Single colonies were screened by restriction enzyme digest (*ApaI* and *XbaI*) and sequencing of isolated plasmid DNA (QIAprep Spin Miniprep Kit, QIAGEN). The pSpliceExpress plasmid contains two constitutively expressed rat insulin exons; such that no insertion of the *EZH2* fragment results in the two exons being splice together.
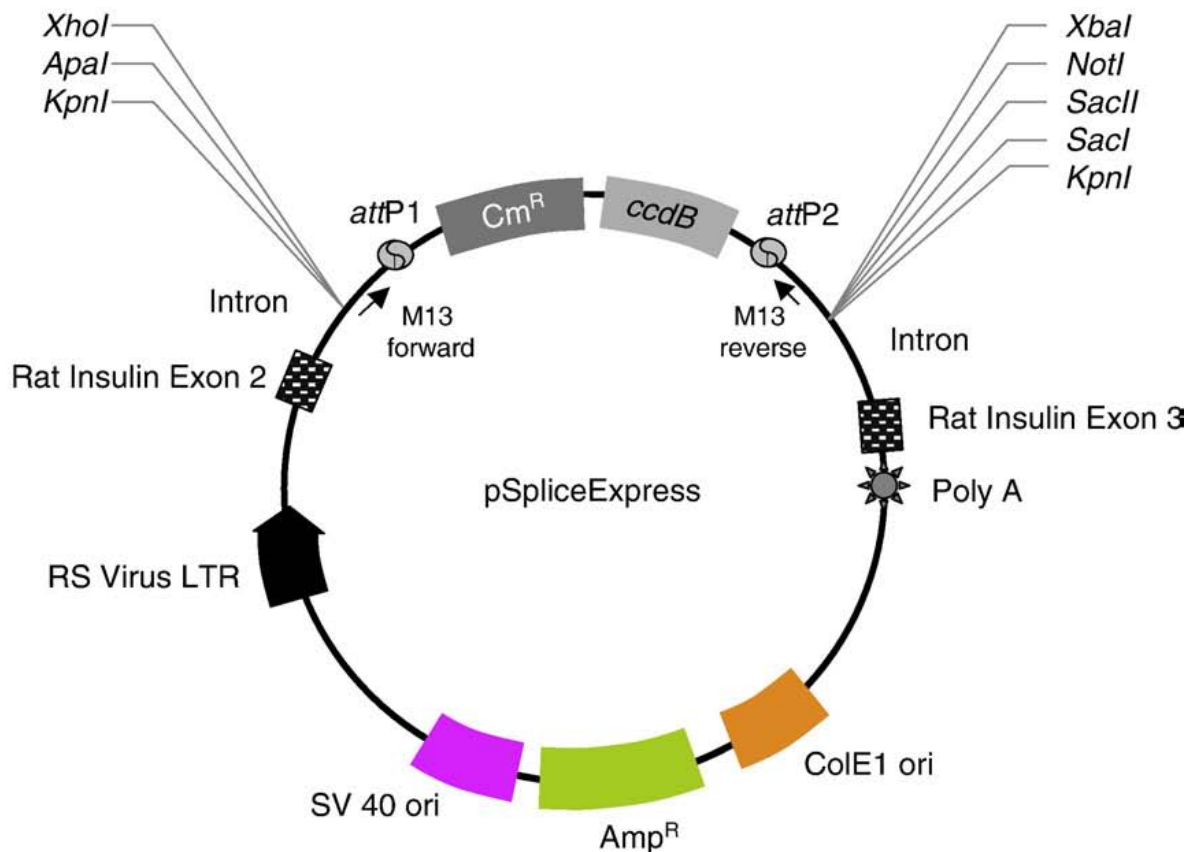
Figure 4.2 The structure of the pSpliceExpress plasmid.

This schematic details the structure of the pSpliceExpress plasmid, including the location of the rat insulin exons, where the restriction enzymes cut, as well as the location of where the *EZH2* insert was inserted in to the plasmid. During the boiling point reaction, the attP1/2 sites are replaced by the attB1/2 sites which were added to the *EZH2* insert. This was a gift from Stefan Stamm (Addgene plasmid #32485; https://www.addgene.org/32485/).
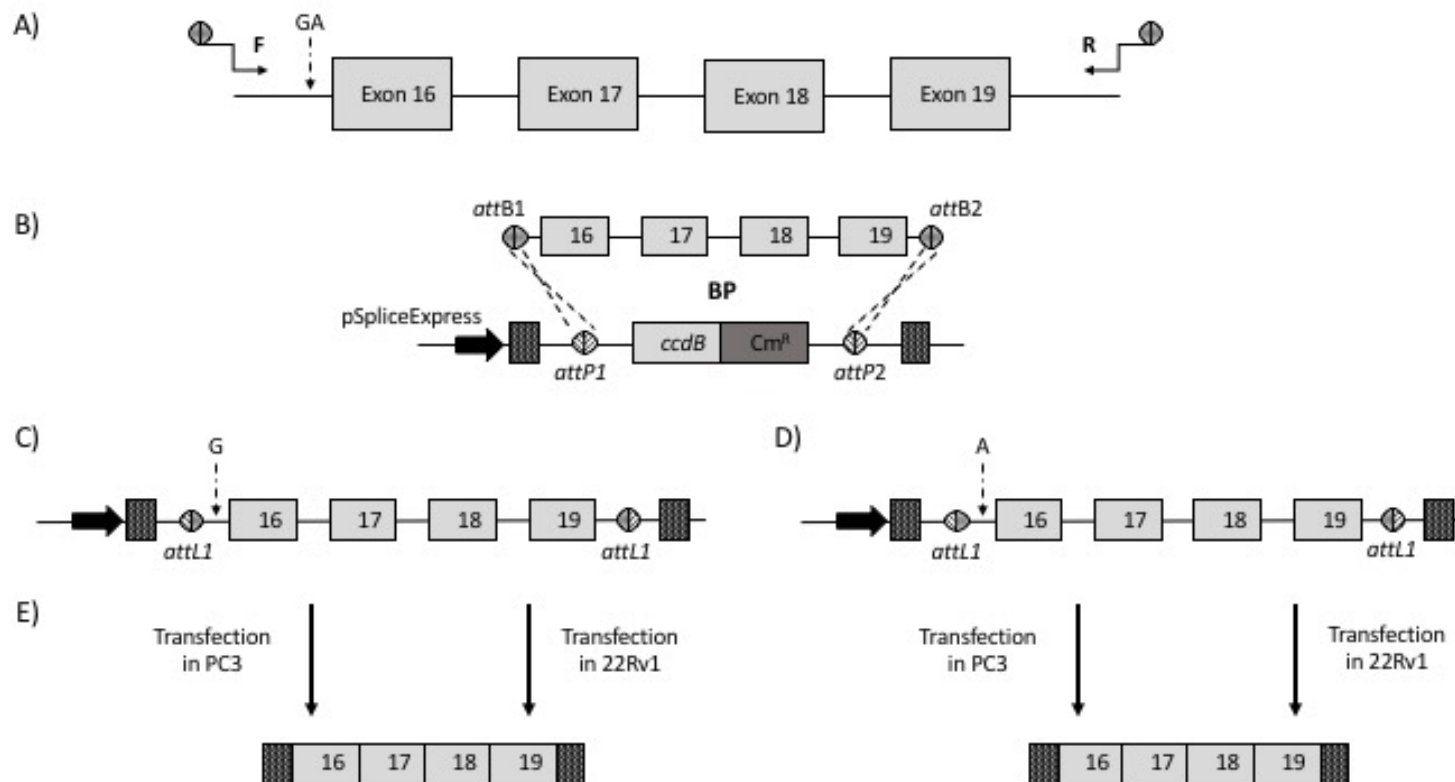
Figure 4.3 Overview of the *EZH2 in vitro* splicing assay.

A) Schematic of the amplified *EZH2* insert (mutation marked as GA), primers were designed in the intron before exon 16 and after exon 19. attB1 and attb2 attachment sites were added to the insert using forward and reverse primers with recombination sites (indicated by grey circles). **B)** The insert was recombined *in vitro* with the pSpliceExpress vector [186]. In this case, the attP1/2 sites are cut and the plasmid is recombined with the attB1/2 sites at the end of the *EZH2* insert. **C)** and **D)** Structure of the final wild-type (C) and variant (D) construct, with the specific allele indicated (G: variant; A: wild-type). The inserted region of *EZH2* is flanked by constitutive rat insulin exons, indicated by the dotted pattern. **E)** The wild-type and variant constructs were transiently transfected in to PC3 and 22Rv1 cells. The effect of the variant on splicing was determined by Sanger sequencing of cDNA, using primers in the rat insulin exons (Appendix 7). The plasmid without an insert was used as a positive control.

4.2.5 Cell culture

PC3 (ATCC® CRL-1435[TM]) and 22Rv1 (ATCC® CRL-2505[TM]) cells were obtained from American Type Culture Collection (Virginia, USA) and cultured in RPMI as previously described by Oakford *et al.* (2010) [187]. Cells were sub cultured every 3-4 days and were maintained between $1x10^5$ and $1x10^6$ cells/mL. All cells were cultured in a humidified incubator at 37°C and 5% CO_2.

4.2.6 Transfection and cDNA sequencing

PC3 and 22Rv1 cells (2×10^6) were transfected with 5µg of variant (Figure 4.2D) or wild-type (Figure 4.2C) plasmid at 300V and 500µF, using a Bio-Rad Gene Pulser X Cell as previously described {Holloway:2000eg}. At 24 hours post-transfection, total RNA was isolated using Tri reagent® (Sigma-Aldrich), and quantified using the Nanodrop® ND-1000 UV spectrophotometer (Nanodrop® Technologies). The Superscript[TM] VILO cDNA Synthesis Kit (Invitrogen) was used for cDNA synthesis, as per the manufacturer's instructions (Appendix 1). 50ng of cDNA was amplified using rat insulin exon 2 forward and exon 3 reverse primers (Appendix 7) [186]. Sanger sequencing was performed to determine the *EZH2* exons transcribed in both the rs78589034 variant and wild-type constructs (Figure 4.2E).

4.2.7 Quantification of EZH2 protein expression

Quantification of EZH2 protein expression in FFPE prostate tumours was assessed by immunohistochemistry (IHC), as discussed in Chapter 2.4 (Appendix 5). Cytospins of HEK293 cells and sections of human colon were used as positive EZH2 controls. Negative controls included primary antibody only, secondary antibody only and a mouse IgG$_1$ isotype control (Dako).

4.3   RESULTS

4.3.1   Rare variant prioritisation

A total of three individuals were successfully WGS in PcTas12, including an affected uncle/nephew pair and an older unaffected male cousin of the uncle (83 years of age; Table 4.1). This cluster of the family was chosen for WGS analysis as it comprises three generations affected with PCa, including an affected brother pair. Unfortunately, germline DNA was not available for one of the brothers, PC12-06 (Figure 4.4). According to the minor allele frequency (MAF) of identified variants in the publicly available database, Exome Aggregation Consortium (ExAC; non-Finnish European, non-TCGA (The Cancer Genome Atlas) population), a total of 66,744 rare variants (MAF) <2%), 47,952 very rare variants (MAF<1%) and 35,722 novel variants were identified in at least one individual from the three that were WGS. Rare variants shared by the affected uncle/nephew pair and not by the unaffected cousin were prioritised for validation and segregation analysis.

Table 4.1 Clinicopathological characteristics of individuals from PcTas12 chosen for whole-genome sequencing.

Sample Identification	Sex	Prostate Cancer Affection Status	Age at diagnosis	Tumour Grade[1]	Contemporary Gleason Score[2]
PC12-01	Male	Affected	63	MD	6 (3+3)
PC12-96	Male	Unaffected	83*	N/A	N/A
PC12-132	Male	Affected	61	-	8 (4+4)
*Unaffected, age at WGS; [1]Tumour grade obtained from pathology report; [2]Contemporary Gleason Score from FFPE tissue block chosen for macrodissection of nucleic acids and IHC; MD: moderately differentiated; -: information not present in original pathology report; N/A: not applicable.					
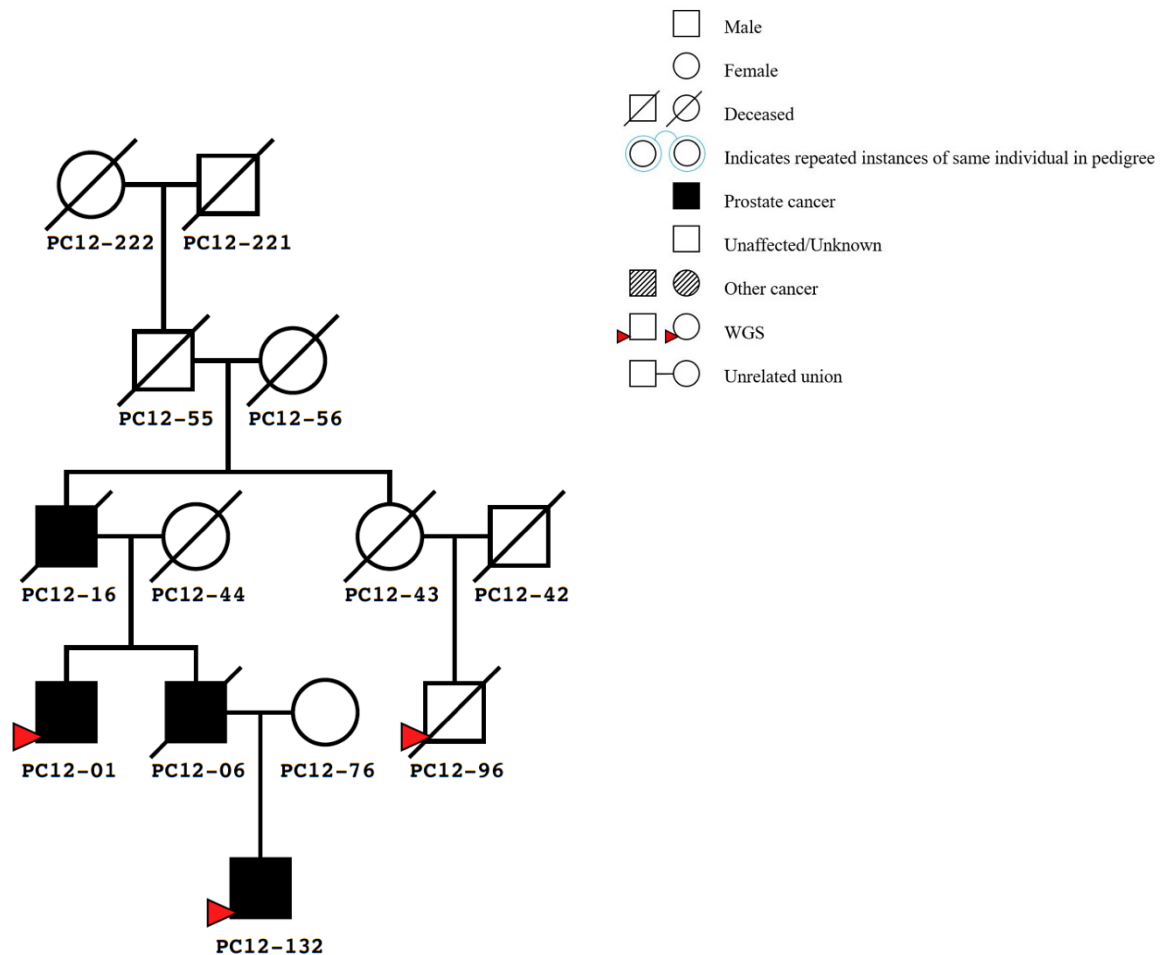
Figure 4.4 A condensed PcTas12 pedigree showing individuals chosen for whole-genome sequencing.

Individuals chosen for WGS are indicated by red arrows, in this case, two PCa cases and one unaffected male relative were chosen.

Four rare variants were prioritised as potential PCa candidates using filtering methods described in Chapter 3.2.2 (Table 4.2). All four variants were validated by Sanger sequencing of WGS individuals. However, following sequencing of four additional PCa cases and 11 unaffected relatives, only the variants in *ITGAD* and *EZH2* segregated with PCa in this family. The *ITGAD* variant was identified in an additional two PcTas12 PCa cases. The intronic *EZH2* variant was found in seven additional PcTas12 individuals; four PCa cases, a female relative and two older men who have been diagnosed with bowel cancer and lymphoma (self-reported; Figure 4.5). Given the *EZH2* variant appeared to segregate in a number of PcTas12 relatives, this variant was prioritised for additional study.

Table 4.2 Rare variants prioritised in the PcTas12 pedigree following whole-genome sequencing of two affected men and one older unaffected man.

Gene	rs number	Chromosome: base pair	ExAC[1] MAF (%)	Segregation in WGS individuals (affected carriers/unaffected carriers)	CADD[2] Score	Allele Change; Amino Acid Change	Number of Control Carriers	ClinVar Search[3]	Validation in WGS Individuals	Segregation in Entire Family
ITGAD	rs147321998	16:31,418,867	0.16	2 out of 2/ 0 out of 1	14.53	C > T; R246X	0 out of 8	Not reported	Yes	Yes
EZH2	rs78589034	7:148,508,818	0.19	2 out of 2/ 0 out of 1	11.7	G > A; Splice	0 out of 8	Weaver syndrome: Benign	Yes	Yes
EPS8	rs78763451	12:15,777,273	0.60	2 out of 2/ 0 out of 1	22.9	C > T; A705T	0 out of 8	Not specified: Benign	Yes	No
TIA1	rs115611153	2:70,441,562	0.63	2 out of 2/ 0 out of 1	22.1	T > C; Q318R	0 out of 8	Welander distal myopathy: Benign	Yes	No

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; MAF: Minor allele frequency; WGS: Whole-genome sequenced; [2]CADD: Combined Annotation Dependent Depletion [164]; Control: Control from the *Tasmanian Prostate Cancer Case-Control Study;* eight were WGS; [3]Associated condition: Interpretation of variant [163].
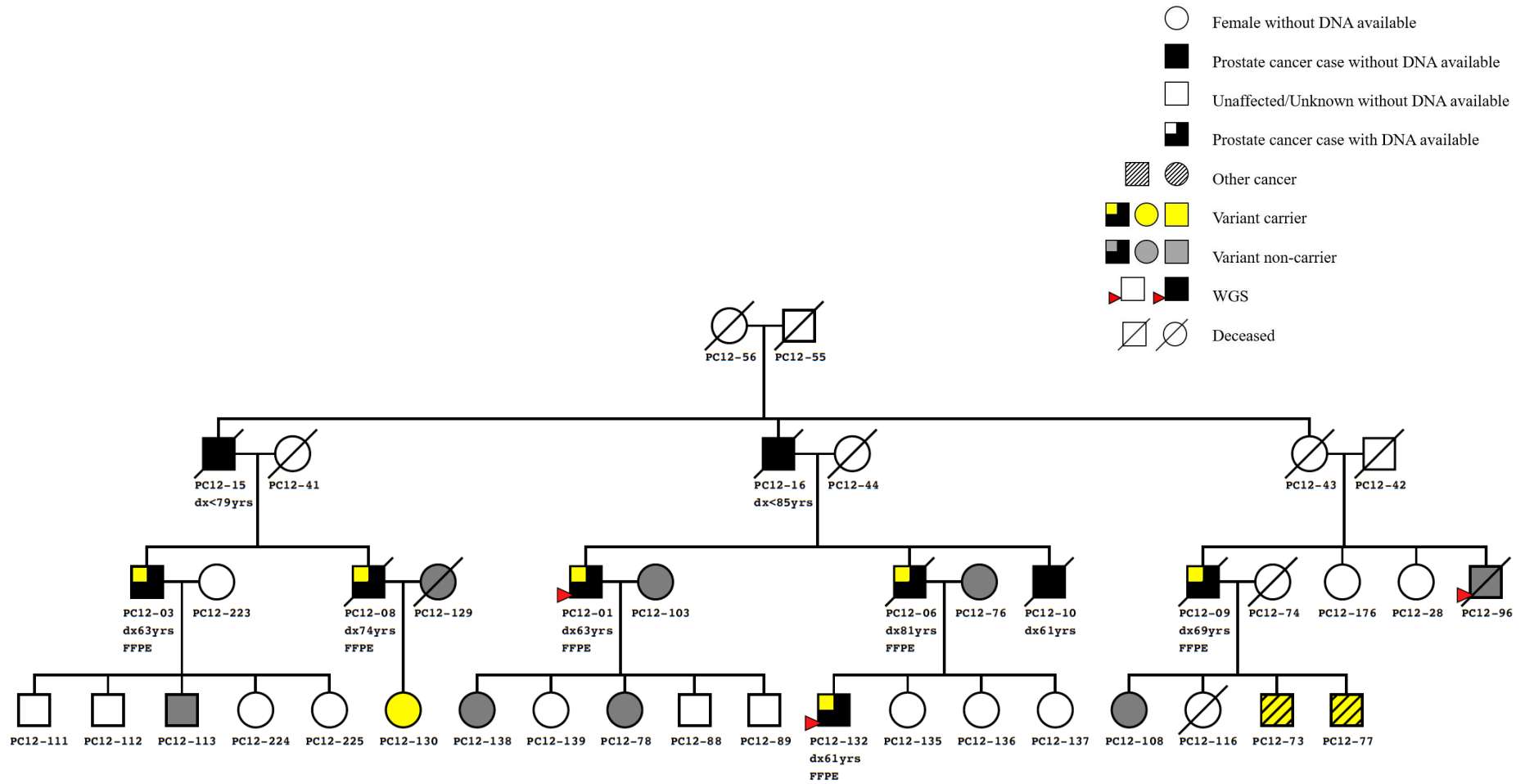
Figure 4.5 *EZH2* variant carriers in PcTas12.

This is a condensed pedigree of PcTas12 comprising all *EZH2* variant carriers (shown in yellow) and their relationship. Non-variant carrier family members are shown in grey and the individuals who were WGS are indicated by red arrows. Notably, the two unaffected male carriers to the right of the pedigree suffer from bowel cancer and lymphoma, respectively. Please note, the genotypes for PC12-03, 06, 08 and 09 were determined from sequencing prostate tumour DNA, which will be discussed below (Chapter 4.3.4).

4.3.2 Association of the *EZH2* variant with prostate cancer risk in Tasmania

Screening of 94 Tasmanian controls, as described in Chapter 2.1.3, revealed the absence of any *EZH2* rs78589034 carriers. Following TaqMan genotyping of the *Tasmanian Familial Prostate Cancer Study* cohorts, an additional PCa case (PcTas9 family) from the *Tasmanian Familial Prostate Cancer Cohort* (n=714), and 3 cases and 1 control from the *Tasmanian Prostate Cancer Case-Control Study* (n=853) were identified as variant carriers. M_{QLS} analysis [2] demonstrated a significant association of the variant with PCa risk in the Tasmanian population (OR=3.27, p=0.001). The number of familial case carriers was much higher compared to their unaffected family members (Table 4.3). The *EZH2* variant was also assessed for enrichment in groups within the *Tasmanian Familial Prostate Cancer Study,* as well as in comparison to ExAC and our Tasmanian controls (Table 4.4). However, there was found to be no enrichment of this variant within any of the groups assessed.

Table 4.3 The association of the *EZH2* variant with prostate cancer risk in the *Tasmanian Familial Prostate Cancer Study* cohorts.

Gene	Variant	Founder Family	Other Families	Familial Case Carriers (n=249)[1]	Familial Unaffected Carriers (n=439)[1]	Sporadic Case Carriers (n=494)[1]	Control Carriers (n=339)[1]	ExAC[2] MAF (%)	Odds Ratio	p-value
EZH2	rs78589034	PcTas12	PcTas9	4 (1.61%)	3 (0.68%)	3 (0.61%)	1 (0.29%)	0.19	3.27	0.001*

Familial case and familial unaffected comprise the *Tasmanian Familial Prostate Cancer Cohort*; Sporadic case and control comprise the *Tasmanian Prostate Cancer Case-Control Study;* [1](n=total sample size); [2]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; MAF: minor allele frequency; *Significant p-value.

Table 4.4 Comparison of *EZH2* variant carrier status in our *Tasmanian Familial Prostate Cancer Study* cohorts compared to ExAC or Tasmanian controls.

Gene	Variant		Entire Resource *versus* ExAC[1]	Familial & Sporadic Cases *versus* ExAC[1]	Familial Cases *versus* ExAC[1]	Sporadic Cases *versus* ExAC[1]	Controls *versus* ExAC[1]	Familial & Sporadic Cases *versus* Controls	Familial Cases *versus* Controls	Sporadic Cases *versus* Controls
EZH2	rs78589034 (ExAC[1] MAF 0.31%)	Chi square; 1df p-value	0.57 (-)[2] p=0.45	0.11 (-) p=0.74	0.22 (-) p=0.64	0.004 (-) p=0.95	1.23 (-) p=0.27	0.42 (+) p=0.52	0.16 (+) p=0.69	0.47 (+) p=0.50
		Number of carriers (n=total sample size)	11 (n=1,521) *versus* 84 (n=26,888)	7 (n=743) *versus* 84 (n=26,888)	4 (n=249) *versus* 84 (n=26,888)	3 (n=494) *versus* 84 (n=26,888)	1 (n=339) *versus* 84 (n=26,888)	7 (n=743) *versus* 1 (n=339)	4 (n=249) *versus* 1 (n=339)	3 (n=494) *versus* 1 (n=339)

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; Entire Resource includes the *Tasmanian Familial Prostate Cancer Cohort* and the *Tasmanian Prostate Cancer Case-Control Study*; Familial cases are a part of the *Tasmanian Familial Prostate Cancer Cohort*; Sporadic case and control comprise the *Tasmanian Prostate Cancer Case-Control Study*; [2]In the chi square test (+/-) indicates directionality, where (+) means the minor allele frequency is greater in the first named population *versus* the comparison dataset, whereas, (-) indicates it is more enriched in the second named population.

4.3.3 Association of the *EZH2* variant with clinical characteristics and tumour pathology

The *EZH2* variant was identified in three branches of the PcTas12 family and was shown to be segregating with PCa (Figure 4.5). No difference was found in the age of diagnosis between PcTas12 *EZH2* variant carrier cases (mean of 67.8 years, n=6) *versus* non-carrier cases (mean of 68.2 years, n=5, p=0.94; Table 4.5). Comparison of the Gleason score (GS) revealed no difference between *EZH2* carriers and non-carriers (p=0.54; Table 4.5); with the majority of men in each group having a GS of 6 (3+3).

Table 4.5 Clinicopathological characteristics of prostate cancer cases from the PcTas12 family, including *EZH2* carriers and non-carriers.

Sample Identification	Age at Diagnosis	Germline *EZH2* Genotype	Tumour *EZH2* Genotype	Tumour Grade[1]	Gleason Score[2]
PC12-02	80	GG	N/A	MD	6 (3+3)
PC12-04	63	GG	N/A	MD	6 (3+3)
PC12-05	64	GG	N/A	WD	-
PC12-07	59	N/A	GG	PD	9 (4+5)
PC12-254	75	GG	N/A	WD	6 (3+3)
PC12-01	63	GA	GA	MD	6 (3+3)
PC12-03	62	N/A	GA	WD	4 (2+2)
PC12-06	80	N/A	GA	PD	7 (3+4)
PC12-08	73	N/A	GA	-	6 (3+3)
PC12-09	68	N/A	GA	-	6 (3+3)
PC12-132	61	GA	GA	-	8 (4+4)

N/A: sample not available; [1]Tumour grade obtained from pathology report; [2]Gleason Score obtained from pathology report; WD: well differentiated; MD: moderately differentiated; PD: poorly differentiated; -: information not present in original pathology report.

4.3.4 Targeted collection of prostate tumour specimens from *EZH2* variant carriers

Targeted collection of FFPE prostate specimens from local pathology laboratories was undertaken for all PcTas12 tumour samples, and additional familial and sporadic *EZH2* variant carriers, as well as a random selection of non-carriers (Table 4.6). Tumour samples were obtained for 18 cases, and genotyping of malignant DNA confirmed three and identified four additional heterozygous *EZH2* carriers in PcTas12 (Table 4.6; Figure 4.5).

No difference was found in the age at diagnosis of *EZH2* variant carriers (n=7) *versus* non-carriers (n=11) used in the functional analyses of this chapter (p=0.41). Likewise, for those samples with a GS on their original pathology report, no difference was observed between carriers (n=7) and non-carriers (n=10, p=0.31; Table 4.6).

Subsequent genotyping of a number of prostate needle biopsies from the *Tasmanian Prostate Tissue Needle Biopsy Resource* (Chapter 2.1.6) identified an additional *EZH2* variant carrier (PT0018). Three needle biopsy samples deemed to be non-carriers were also included in this study (Table 4.7).

Table 4.6 Clinicopathological characteristics of FFPE prostate tumour samples obtained for *EZH2* carriers and non-carriers used in the functional analyses of this chapter.

Sample Identification	Age at Diagnosis	Germline *EZH2* Genotype	Tissue Source	Tumour *EZH2* Genotype	Tumour Grade[1]	Contemporary Gleason Score[2]
PC4-03	80	GG	TURP	GG	M/PD	7 (4+3)
PC11-11	85	N/A	TURP	GG	-	7 (3+4)
PC12-07	59	N/A	TURP	GG	PD	9 (4+5)
PC19-02	50	GG	RP	GG	-	6 (3+3)
PC60-01	58	GG	TURP	GG	WD	6 (3+3)
PC72-04	70	GG	TURP	GG	PD	9 (4+5)
PC72-06	62	GG	TURP	GG	W/MD	5 (3+2)
PC3250-01	51	GG	RP	GG	PD	9 (4+5)
DVA 216	64	GG	RP	GG	-	5 (3+2)
DVA 402	52	GG	RP	GG	MD	6 (3+3)
DVA 1002	61	GG	RP	GG	-	6 (3+3)
PC12-01	63	GA	RP	GA	MD	6 (3+3)
PC12-03	62	N/A	TURP	GA	WD	4 (2+2)
PC12-06	80	N/A	TURP	GA	PD	7 (3+4)
PC12-08	73	N/A	TURP	GA	-	6 (3+3)
PC12-09	68	N/A	TURP	GA	-	6 (3+3)
PC12-132	61	GA	RP	GA	-	8 (4+4)
DVA 416	62	GA	RP	GA	MD	6 (3+3)

N/A: sample not available; [1]Tumour grade obtained from pathology report; [2]Contemporary Gleason Score from FFPE tissue block chosen for macrodissection of nucleic acids and IHC; TURP: Transrectal resection of the prostate; RP: Radical prostatectomy; WD: well differentiated; MD: moderately differentiated; W/MD; well-moderately differentiated; PD: poorly differentiated; M/PD: moderately-poorly differentiated; -: information not present in original pathology report.

Table 4.7 Clinicopathological characteristics of the prostate needle biopsy samples obtained for an *EZH2* carrier and non-carriers used in the functional analyses of this chapter.

Sample Identification	Age at Diagnosis	Tissue Source	Tumour *EZH2* Genotype	Gleason Score[1]
PT0001	70	TRUS	GG	9 (4+5)
PT0002	73	TRUS	GG	6 (3+3)
PT0003	61	TRUS	GG	7 (4+3)
PT0018	59	TRUS	GA	6 (3+3)
TRUS; Transrectal ultrasound biopsy: [1]Gleason Score obtained from pathology report. Note: Germline samples are not available for any of these men.				

4.3.5 The effect of the *EZH2* variant on *EZH2* gene expression

To investigate *EZH2* expression, RNA was extracted from adjacent malignant and benign glands for all prostate specimens (n=18), except for three samples where only malignant glands were present. Amplification of the housekeeping genes, *GAPDH* and *β-Actin* showed moderate expression however, amplification of the gene of interest, *EZH2* was poor in both malignant and benign prostate glands. To determine whether these results were due to the poor quality FFPE samples, *EZH2* gene expression was then investigated in three PCa cell lines and the four needle biopsy cores (Appendix 8). *EZH2* expression was highest in the LNCaP cells, followed by 22Rv1 and PC3 cells, and in comparison, expression was relatively low in the needle biopsy samples, similar to that observed in the PC3 cells. Finally, *EZH2* expression in the two cores from the *EZH2* carrier (PT0018) appeared lower than the non-carriers, however, due to the small sample size, formal statistical analyses could not be undertaken (Figure 4.6).
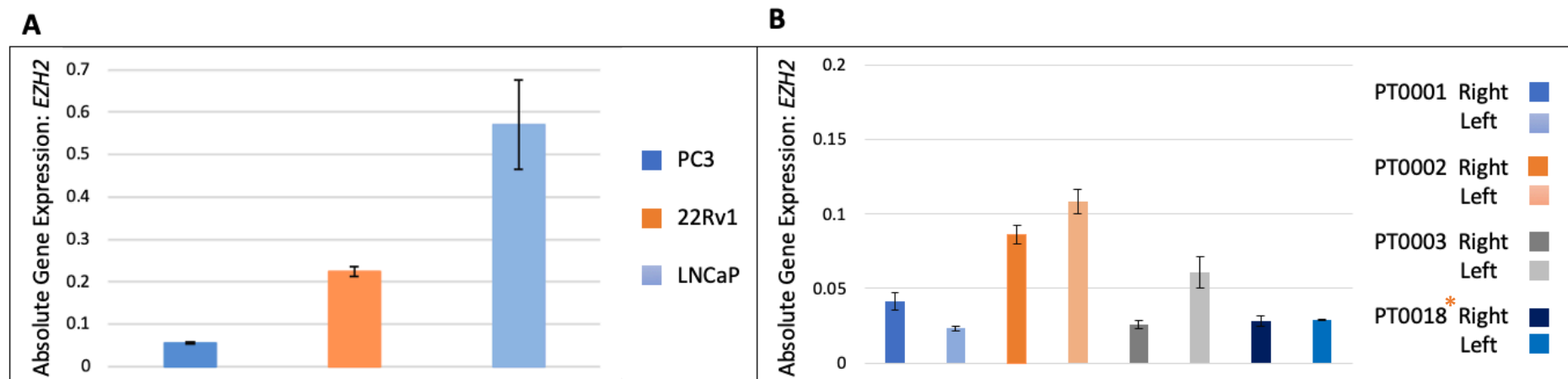
Figure 4.6 *EZH2* gene expression analysis in prostate cancer cell lines and prostate needle biopsy samples.

EZH2 expression was assessed in PCa cell lines and prostate needle biopsy samples. Absolute *EZH2* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. **A)** *EZH2* expression in individual PCa cell lines is shown here. **B)** *EZH2* expression in individual prostate biopsy cores is shown here. * *EZH2* variant carrier.

4.3.6   The effect of the *EZH2* variant on splicing

The *EZH2*, rs78589034 variant is located 6bp before the start of exon 16 and, therefore, could affect splicing. *EZH2* is a highly variable gene and multiple transcripts have been identified [137] (Figure 4.7). Notably, exon 16 is not included in three of the 12 most common transcripts expressed in the prostate. In this study, splicing was assessed by transient transfection of variant (A allele) and wild-type (G allele) constructs, including *EZH2* exons 16-19 and 200bp of intronic sequence either side, into PC3 and 22Rv1 cells, as shown in Figure 4.3. In both PC3 and 22Rv1 cells, cDNA sequencing of the transfected constructs showed presence of all exons downstream of the variant (within the construct), in both the variant and wild-type constructs (Figure 4.3E). These results suggest that the rs78589034 variant does not alter splicing in this model.
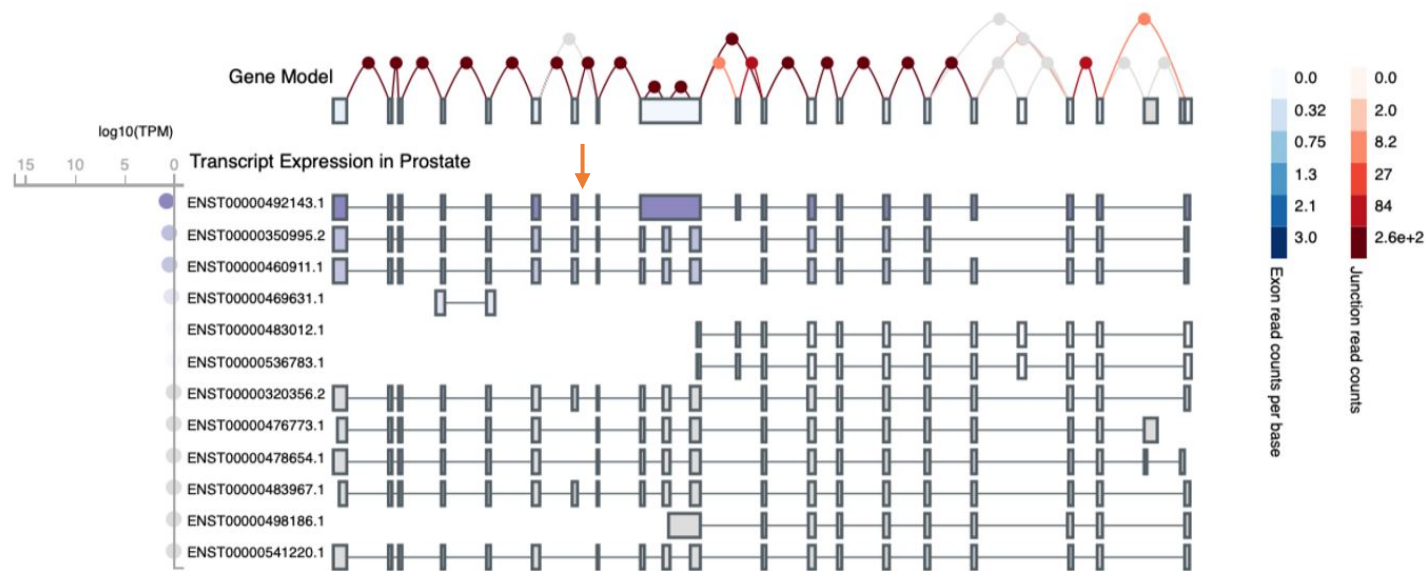
Figure 4.7 Screenshot of the GTEx Portal showing the most commonly expressed *EZH2* transcripts in the prostate.

Multiple transcripts of *EZH2* have been identified in the prostate. The schematic for each transcript consists of exons, which are shown as boxes and introns as lines. As per data from the GTEx Analysis Release V8 (dbGaP Accession phs000434.v8.p2; https://gtexportal.org/home/), exon expression is shown in a heatmap format, with greater median read count per base depicted in dark blue. Transcription is right to left. The most common transcript, ENST00000492143.1 was used for primer design for *EZH2* gene expression analyses and the rs78589034 is marked with an orange arrow [137].

Given the known differential splicing patterns of *EZH2*, and the predicted disruption by the presence of rs79589034; the presence/absence of selected exons was examined. The exon-level expression of *EZH2* in six regions across the gene was first assessed in the PCa cell lines and needle biopsy cores to determine whether particular exons were more consistently expressed in these samples (Appendix 8). Analysis of *EZH2* expression in the cell lines revealed that the regions of exon 8/9 and 20/21 were more highly expressed compared to the other regions (or more easily quantified). However, only exon 20/21 was significantly higher compared to the other regions in the needle biopsy samples (Figure 4.8). Interestingly, the *EZH2* carrier, PT0018 had the lowest *EZH2* exon 20/21 expression, with the expression level of both cores similar to that of the exon 8/9 region. The right biopsy from PT0001 and PT0002 had higher *EZH2* expression in the exon 20/21 region compared to the left lobe. Given our ability to quantify *EZH2* exon 20/21 expression in these samples, the 18 FFPE prostate tissue samples were examined. However, once again, amplification was poor and therefore, absolute expression of the regions of *EZH2* could not be determined in these samples.
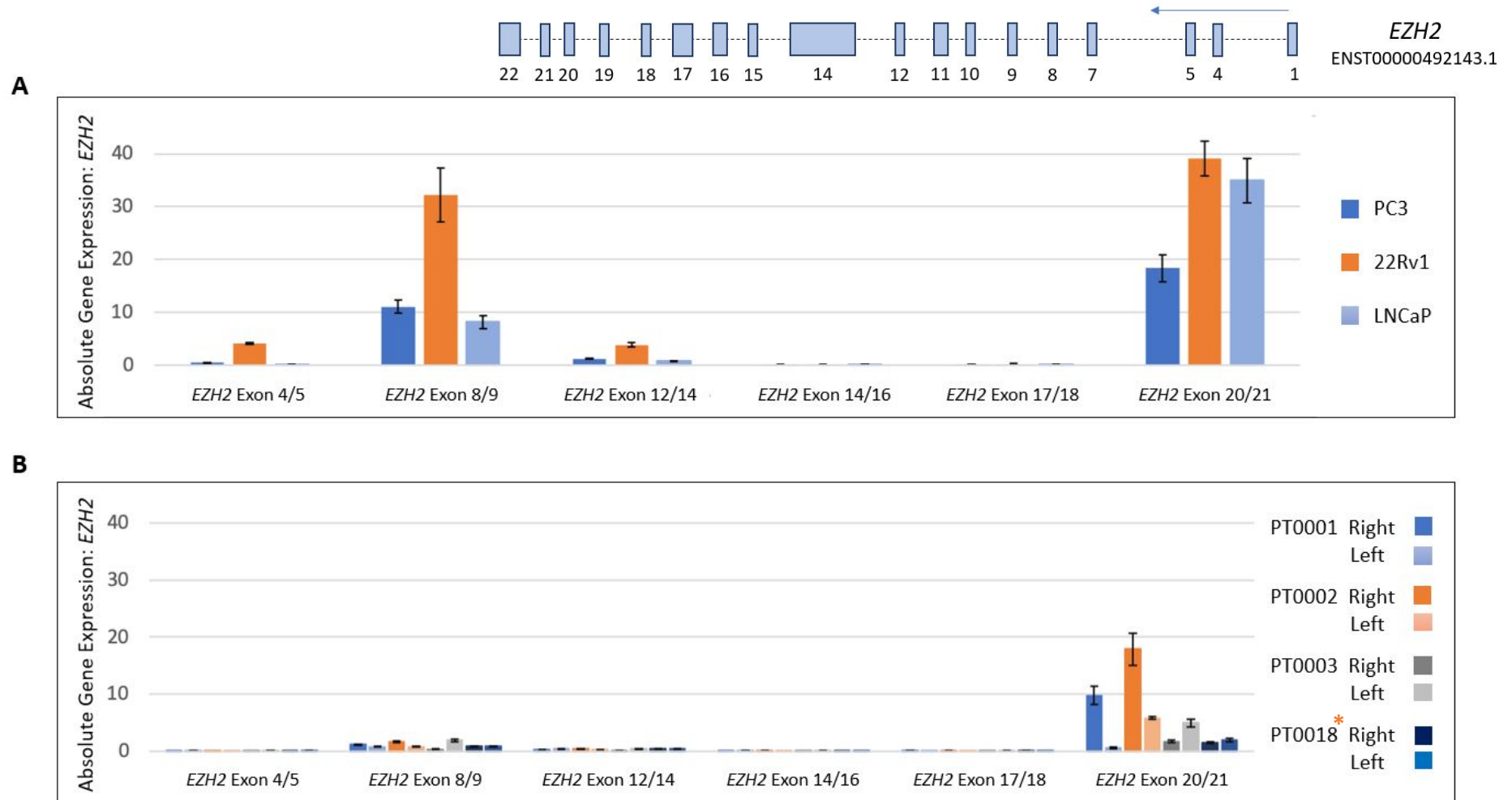
Figure 4.8 *EZH2* gene expression analysis in multiple regions of the gene in prostate cancer cell lines and prostate needle biopsy samples.

EZH2 expression in six different regions of the gene was assessed in PCa cell lines and prostate needle biopsy samples. A schematic of the most commonly transcribed isoform of *EZH2* in the prostate is shown at the top of the page. Absolute *EZH2* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. **A)** *EZH2* expression in individual PCa cell lines is shown here. **B)** *EZH2* expression in individual prostate biopsy cores is shown here. * *EZH2* variant carrier.

4.3.7 The effect of the *EZH2* variant on EZH2 protein expression

IHC was performed on all 18 FFPE prostate tumours and the four needle biopsy samples, with EZH2 protein expression assessed separately in malignant and benign glands. EZH2 staining was negative for all prostate samples analysed, in both malignant and benign glands (Figure 4.9C &D). Cytospins of HEK293 cells and a section of human colon tissue were used as positive controls. The HEK293 cells showed moderate to strong staining of EZH2 and the human colon tissue, weak to moderate (Figure 4.9A& B).
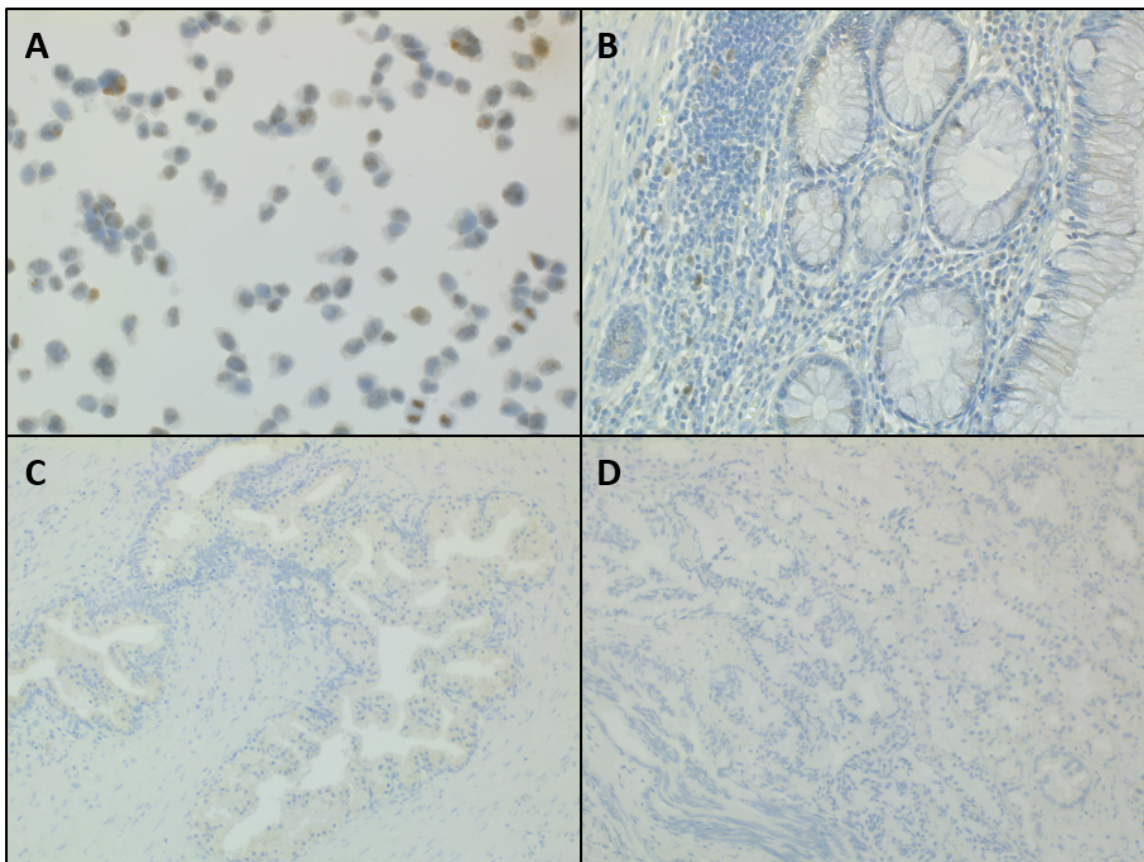
Figure 4.9 EZH2 protein expression in HEK293 cells, human colon and FFPE prostate tumour samples.

EZH2 protein expression was assessed in 18 prostate tumour specimens from the *Tasmanian Prostate Tissue Pathology Resource* to determine whether the intronic variant affected EZH2 protein levels. In short, IHC using an antibody targeting amino acid 696-745 of the EZH2 protein was utilised to assess protein expression. Staining intensity was scored as none, weak, moderate or strong. **A)** Moderate-strong staining of EZH2 in the nucleoplasm of HEK293 cells. **B)** Moderate-strong staining of EZH2 in the nucleoplasm of human colon glands. **C)** No staining of EZH2 in benign prostate glands. **D)** No staining of EZH2 in malignant prostate glands. Images were taken with a Leica 2500 microscope (x200) using the Leica Application Suite V3.

4.3.8 The effect of the *EZH2* variant on *EZH2* target gene expression

Due to low *EZH2* expression it was challenging to detect whether there were differences between *EZH2* carriers and non-carriers. Therefore, expression of *EZH2* target genes, *CDH1, HOXA9,* and *MSMB* were examined to determine whether the variant had an effect on their expression. It is possible that the variant may alter *EZH2* expression slightly and thus, could have a direct impact on the level of expression of its target genes. Studies have previously observed an inverse relationship of *CDH1* [188]*, HOXA9* [189] and *MSMB* [190] with *EZH2.*

Initially, expression levels were assessed in the PCa cell lines and the four needle biopsy samples (Appendix 9). *CDH1* and *HOXA9* expression was 5-fold higher in the androgen-refractory PC3 cells compared to the level of expression in the androgen-sensitive LNCaP and 22Rv1 cells (Figure 4.10). *CDH1* and *HOXA9* expression in the cell lines was inversely correlated with the overall average expression of *EZH2* (mean of the six regions), except *CDH1* expression was similar to *EZH2* in PC3 cells. An inverse relationship between *MSMB* and *EZH2* was detected, but only in the 22Rv1 and LNCaP cells (Figure 4.10).
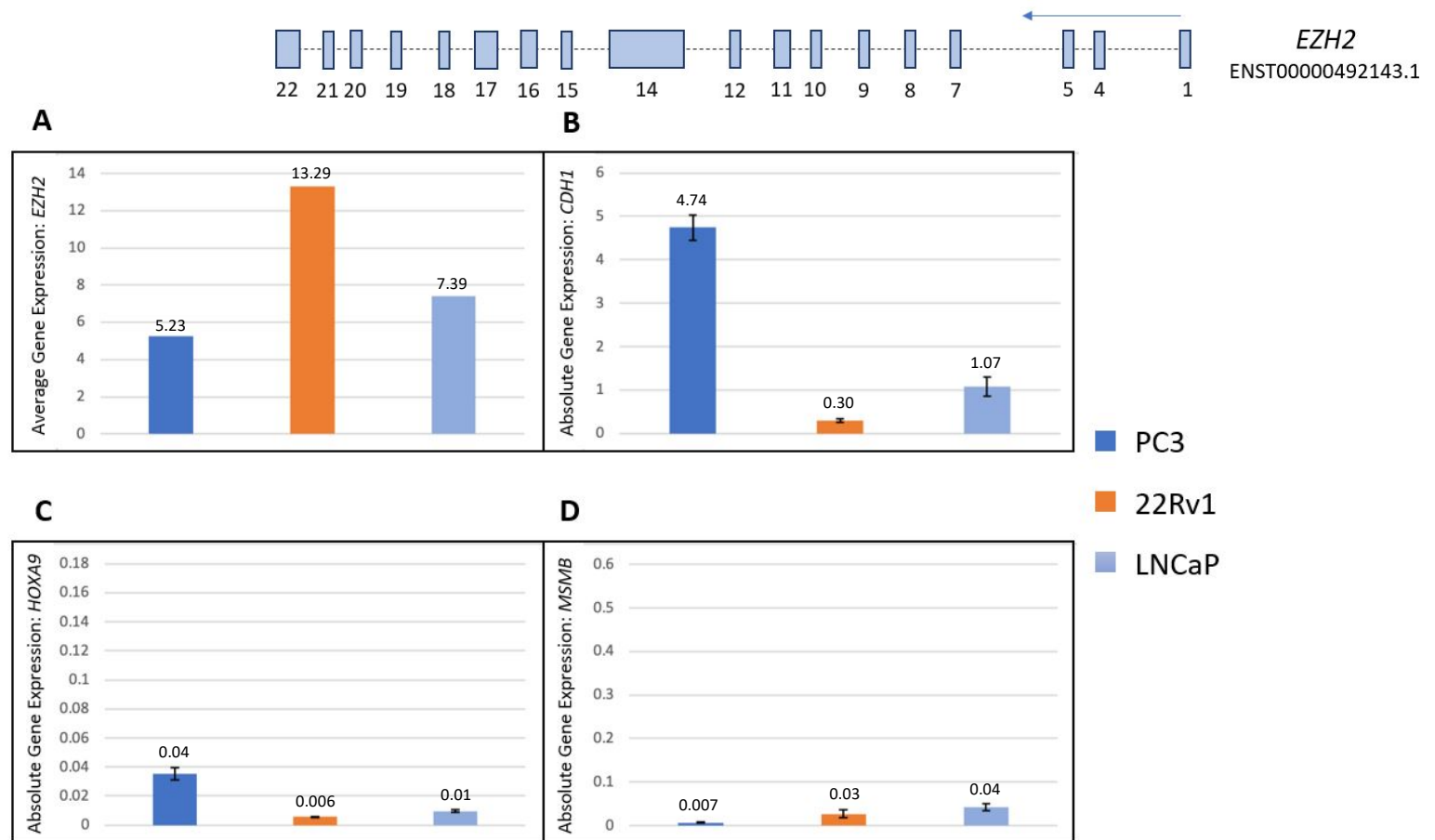
Figure 4.10 Average gene expression of *EZH2* and absolute gene expression of target genes, *CDH1, HOXA9* and *MSMB* in prostate cancer cell lines.
A) *EZH2* expression in six regions of the gene were assessed in three PCa cell lines. Average (mean) expression of all regions of *EZH2* assessed; exon 4/5, 8/9, 12/14, 14/16, 17/18, 20/21 is shown here. A schematic of the most commonly transcribed isoform of *EZH2* in the prostate is shown at the top of the page. *CDH1, HOXA9 and MSMB* expression was assessed in three PCa cell lines. Absolute *CDH1, HOXA9 and MSMB* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. **B)** Individual cell line *CDH1* expression.is shown here. **C)** Individual cell line *HOXA9* expression is shown here. **D)** Individual cell line *MSMB* expression is shown here.

Analysis of the four needle biopsy samples showed considerable variability across the gene expression profiles, similar to the cell lines. *CDH1* expression in the two PT0002 cores were inversely correlated with *EZH2* expression, however, the correlation was not as distinct compared to the cell lines (Figure 4.11). *HOXA9* expression in the needle biopsy samples was similar to PC3 cells, except the right core of PT0002 had higher expression compared to all other samples. Unlike the cell lines, no inverse trend in *HOXA9* expression with *EZH2* was found, though *MSMB* expression in samples from two of the needle biopsies (PT0002 and PT0003) was inversely correlated with *EZH2* expression. Expression of all target genes did not differ between the *EZH2* variant carrier (PT0018) and non-carriers (n=3), however given the small sample size statistical analyses were unable to be performed (Figure 4.11).
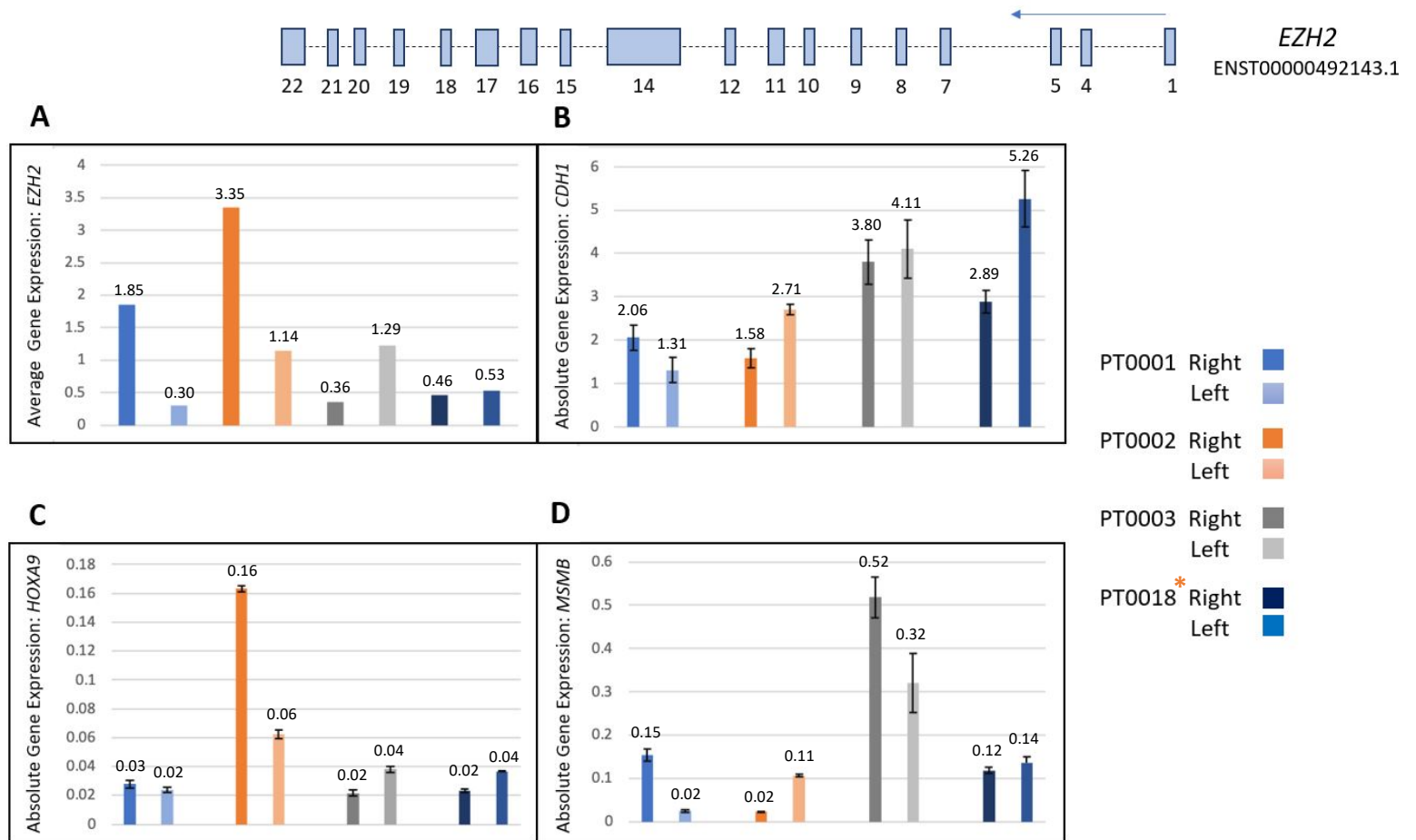
Figure 4.11 Average gene expression of *EZH2* and absolute gene expression of target genes, *CDH1, HOXA9* and *MSMB* in prostate needle biopsy samples.
A) *EZH2* expression in six regions of the gene were assessed in four prostate needle biopsy samples. Average (mean) expression of all regions of *EZH2* assessed; exon 4/5, 8/9, 12/14, 14/16, 17/18, 20/21 is shown here. A schematic of the most commonly transcribed isoform of *EZH2* in the prostate is shown at the top of the page. *CDH1, HOXA9* and *MSMB* expression was assessed in four prostate needle biopsy samples. Absolute *CDH1, HOXA9* and *MSMB* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. **B)** Individual biopsy core *CDH1* expression is shown here. **C)** Individual biopsy core *HOXA9* expression is shown here. **D)** Individual biopsy core *MSMB* expression is shown here. * *EZH2* variant carrier.

Given the known inverse relationship of *EZH2* and its target genes in the literature and our ability to quantitate *CDH1* and *MSMB* expression in the PCa cell lines and needle biopsy samples, expression was assessed in the 18 FFPE prostate tissue samples (Appendix 10). Firstly, differences in expression between malignant and adjacent benign glands was assessed in 14 tumour samples. There was found to be no significant difference in *CDH1* gene expression between paired malignant and benign prostate glands ($n=14_{pairs}$; $p=0.30$; Figure 4.12). *MSMB* expression was also unchanged between the two groups ($p=0.38$). Assessment of *CDH1* in tumours from *EZH2* carriers (n=7) and non-carriers (n=11) identified no difference in expression between malignant ($p=0.12$) and benign glands (n= 15, $p=0.44$), respectively. *MSMB* expression also appeared unaffected by the *EZH2* variant, in both malignant ($p=0.54$) and benign glands (n=15, $p=0.47$; Figure 4.12).
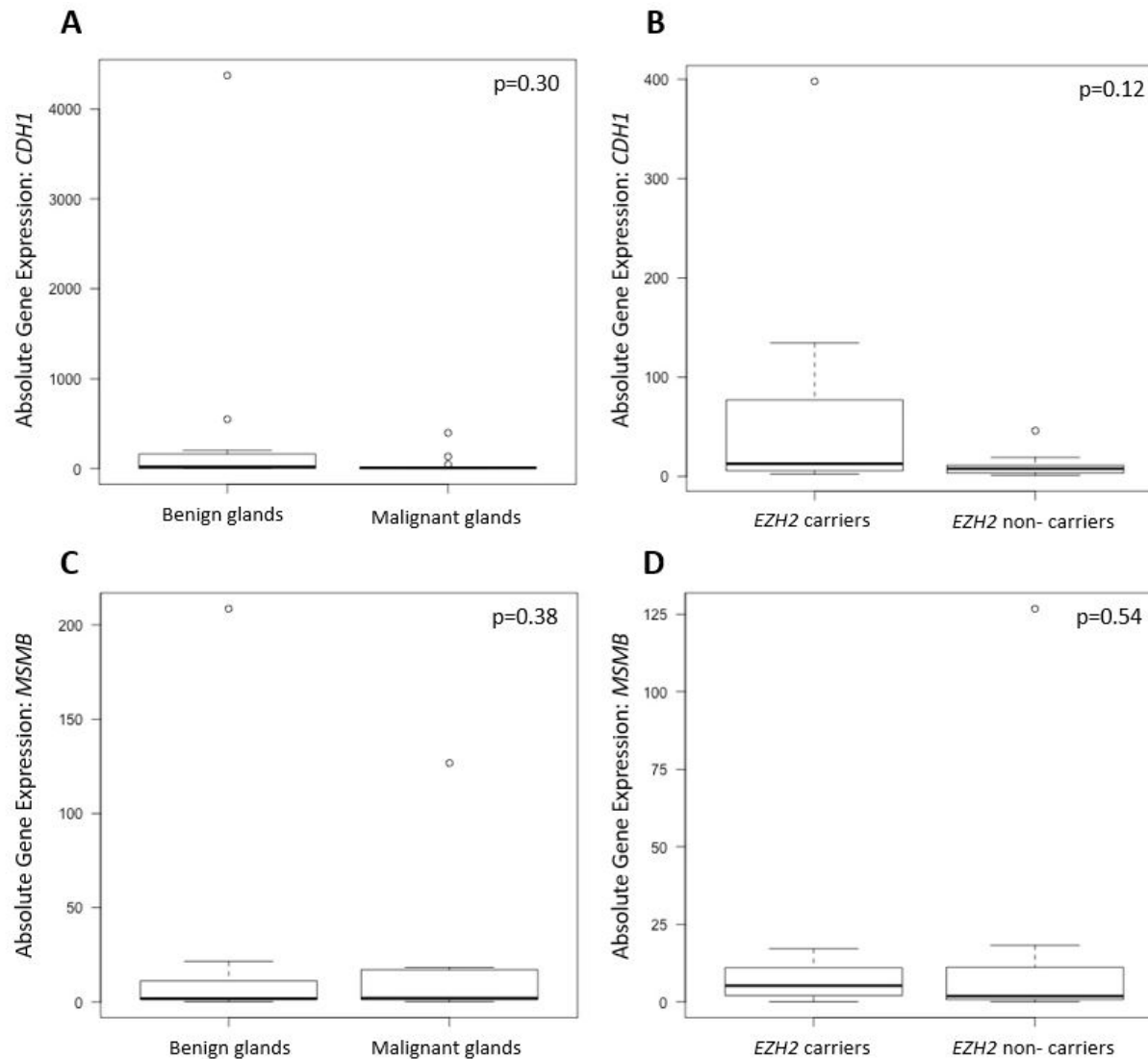
Figure 4.12 *CDH1* and *MSMB* gene expression analysis in malignant and benign prostate glands, and in malignant glands from *EZH2* variant carriers and non-carriers

The spread of the data is represented by a box and whisker plot. Median expression is shown by the thick black line, the interquartile range (middle 50% of data set) is represented by the box, and the minimum and maximum values by the whiskers (dotted lines). Individual outliers are shown with dots. **A/C)** *CDH1 and MSMB* expression was assessed in prostate tumours with matched malignant and benign glands (n_{pairs}=14). Absolute *CDH1* and *MSMB* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes and expression in malignant and benign glands was compared using a paired Student's t-test (*CDH1*: **A**; *MSMB*; **C**). **B/D)** *CDH1* and *MSMB* expression was assessed in malignant prostate glands from *EZH2* variant carriers (n=6) and non-carriers (n=11). Absolute *CDH1* and *MSMB* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes and expression in malignant glands from *EZH2* variant carriers and non-carriers was compared using an unpaired Student's t-test (*CDH1*: **C**; *MSMB*; **D**).

100

4.3.9 The effect of the *EZH2* variant on *EZH2* splicing factor expression

Due to low *EZH2* expression and the hypothesised functional effect of the variant on splicing, gene expression levels of *EZH2* splicing factors, *SF3B1, SF3B3* and *U2AF1* were determined. Initially, expression levels of these splicing factors were assessed in the PCa cell lines and the four needle biopsy samples (Appendix 9). Interestingly, *U2AF1* was highly expressed in all cell lines compared to *SF3B1* and *SF3B3*. Whilst *SF3B1* and *SF3B3* did not show an inverse trend with *EZH2*, *U2AF1* was inversely correlated with the overall average expression of *EZH2* (mean of the six regions) in all cell lines (Figure 4.13).
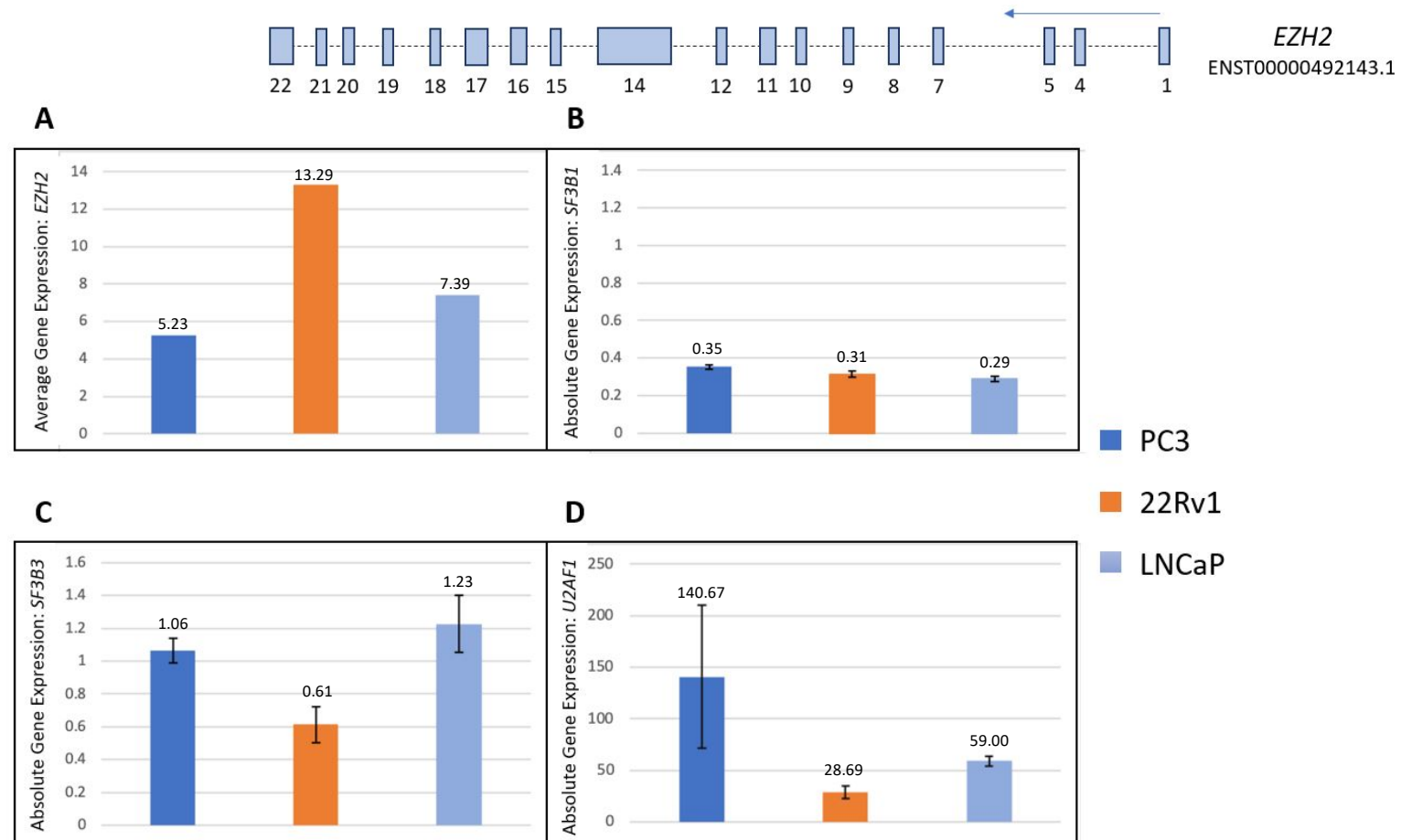
Figure 4.13 Average gene expression of *EZH2* and absolute gene expression of splicing factors, *SF3B1*, *SF3B3* and *U2AF1* in prostate cancer cell lines.
A) *EZH2* expression in six regions of the gene were assessed in three PCa cell lines. Average (mean) expression of all regions of *EZH2* assessed; exon 4/5, 8/9, 12/14, 14/16, 17/18, 20/21 is shown here. A schematic of the most commonly transcribed isoform of *EZH2* in the prostate is shown at the top of the page. *SF3B1, SF3B3* and *U2AF1* expression was assessed in three PCa cell lines. Absolute *SF3B1, SF3B3* and *U2AF1* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. **B)** Individual cell line *SF3B1* expression.is shown here. **C)** Individual cell line *SF3B3* expression is shown here. **D)** Individual cell line *U2AF1* expression is shown here.

The needle biopsy samples also had very high expression of the splicing factor, *U2AF1*. The expression pattern of *U2AF1* in these samples did not indicate that there was an inverse correlation with *EZH2* expression however, the two samples with the highest *U2AF1* expression (PT0003 right and PT0018 right) did have the lowest average expression of *EZH2*. But, unlike the cell lines, *SF3B1* and *SF3B3* expression appeared to follow a similar trend in expression to *EZH2* (Figure 4.14). While the expression of the splicing factors did not appear to differ between the variant carrier (PT0018) and non-carriers, this was not able to be confirmed statistically due to a limited sample size.
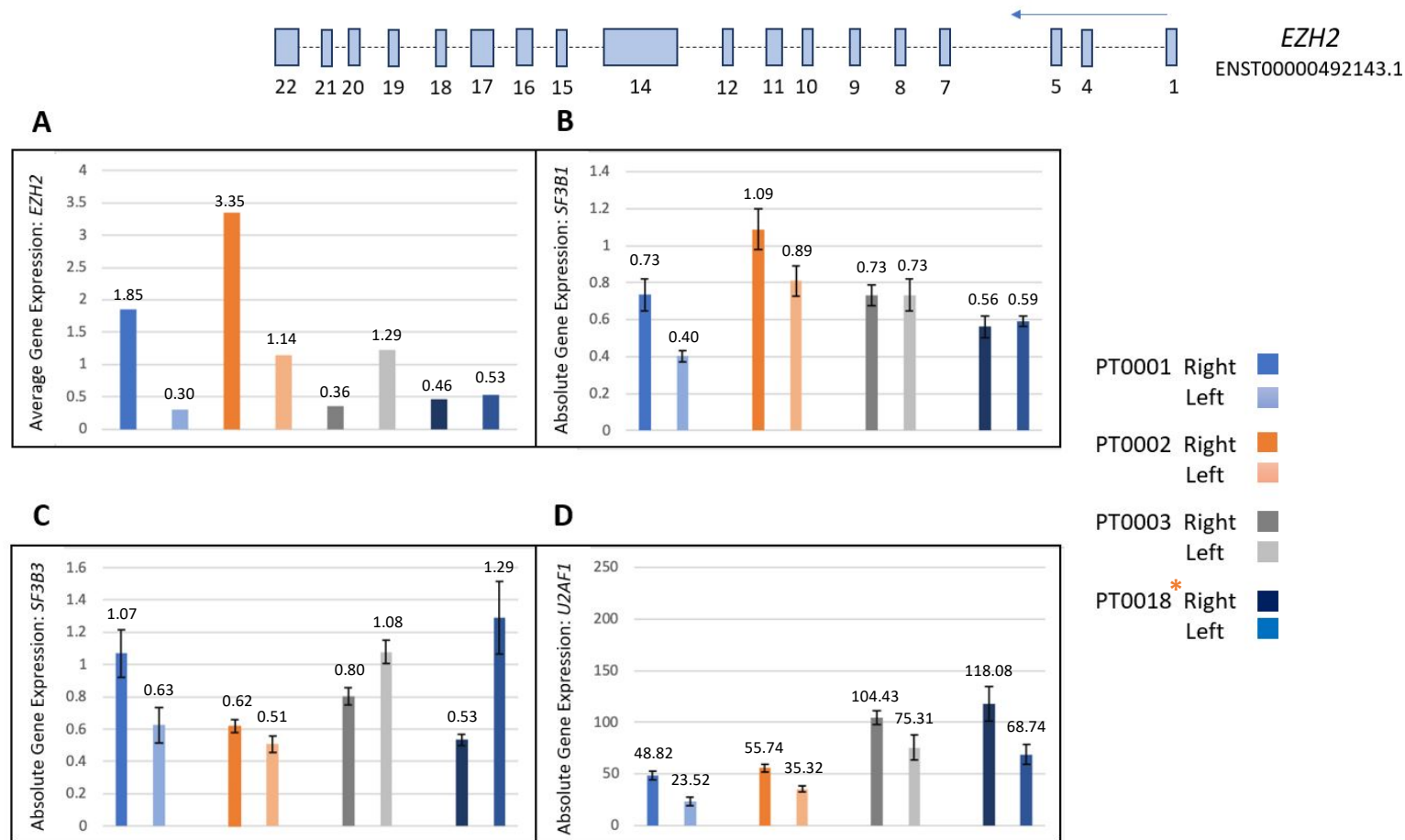
Figure 4.14 Average gene expression of *EZH2* and absolute gene expression of splicing factors, *SF3B1, SF3B3* and *U2AF1* in prostate needle biopsy samples.
A) *EZH2* expression in six regions of the gene were assessed in four prostate needle biopsy samples. Average (mean) expression of all regions of *EZH2* assessed; exon 4/5, 8/9, 12/14, 14/16, 17/18, 20/21 is shown here. A schematic of the most commonly transcribed isoform of *EZH2* in the prostate is shown at the top of the page. *SF3B1, SF3B3* and *U2AF1* expression was assessed in four prostate needle biopsy samples. Absolute *SF3B1, SF3B3* and *U2AF1* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. **B)** Individual biopsy core *SF3B1* expression is shown here. **C)** Individual biopsy core *SF3B3* expression is shown here. **D)** Individual biopsy core *U2AF1* expression is shown here. * *EZH2* variant carrier.

As *U2AF1* expression was very high in all cell line and needle biopsy samples, *U2AF1* expression was assayed in the 18 FFPE prostate tissue samples (Appendix 10). Not surprisingly, given the quality of RNA, *U2AF1* expression levels were much lower in the FFPE tumours compared to the cell lines and needle biopsy samples. Whilst malignant gland expression was generally higher than benign, there was no significant difference in *U2AF1* gene expression between paired malignant and benign prostate glands (n_{pairs}=11, p=0.11; Figure 4.15). In malignant glands, the majority of *EZH2* variant carriers (n=6) had lower *U2AF1* expression than non-carriers (n=11), however this was not statistically significant (p=0.12). A similar expression pattern was observed in benign prostate glands, however the difference between *EZH2* carriers (n=5) and non-carriers (n=8) was statistically significant (p=0.03; Figure 4.15).
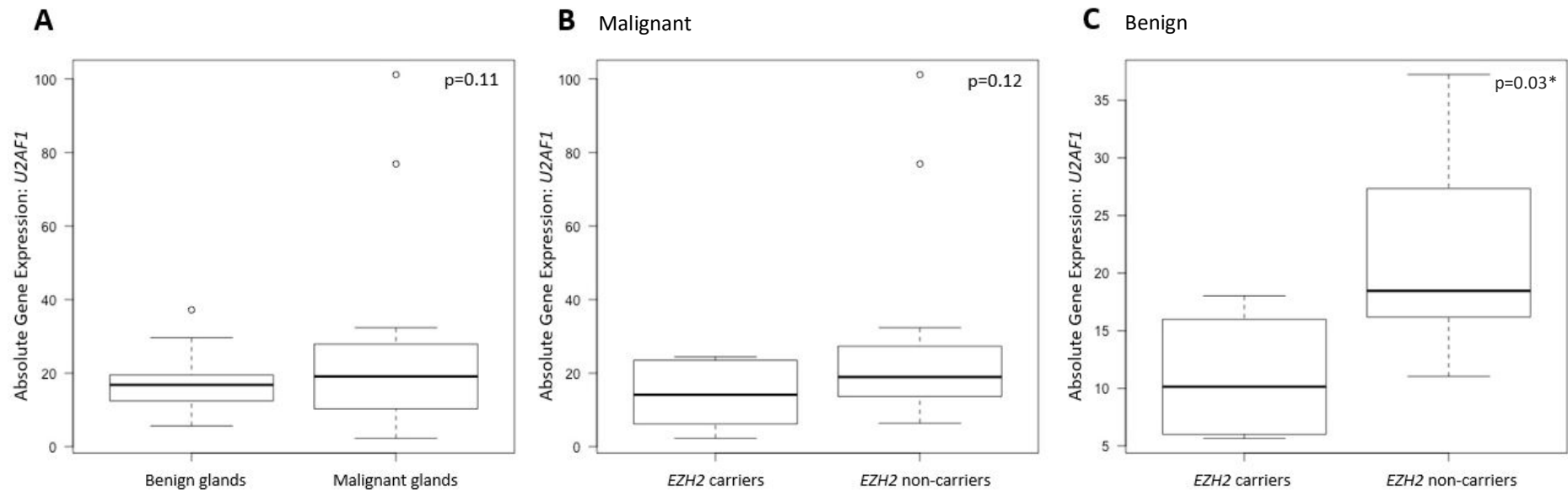
Figure 4.15 *U2AF1* **gene expression analysis in malignant and benign prostate glands; in malignant glands from** *EZH2* **variant carriers and non-carriers, and in benign glands from** *EZH2* **variant carriers and non-carriers**

The spread of the data is represented by a box and whisker plot. Median expression is shown by the thick black line, the interquartile range (middle 50% of data set) is represented by the box, and the minimum and maximum values by the whiskers (dotted lines). Individual outliers are shown with dots. **A)** *U2AF1* expression was assessed in prostate tumours with matched malignant and benign glands (n_{pairs}=11). Absolute *U2AF1* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. *U2AF1* expression in malignant and benign glands was compared using a paired Student's t-test. **B)** *U2AF1* expression was assessed in malignant prostate glands from *EZH2* variant carriers (n=6) and non-carriers (n=11). Absolute *U2AF1* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. *U2AF1* expression in malignant glands from *EZH2* variant carriers and non-carriers was compared using an unpaired Student's t-test and the spread of the data is shown here. **C)** *U2AF1* expression was assessed in benign prostate glands from *EZH2* variant carriers (n=5) and non-carriers (n=8). Absolute *U2AF1* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. *U2AF1* expression in benign glands from *EZH2* variant carriers and non-carriers was compared using an unpaired Student's t-test.

106

4.4   DISCUSSION

4.4.1   *EZH2* as a potential prostate cancer risk variant

A rare variant in *EZH2* (rs78589034) was initially identified in two PCa cases in PcTas12. It is an intronic variant that occurs 6bp from the beginning of exon 16, which according to the Human Splicing Finder is an acceptor splice site [191], therefore it may affect expression of *EZH2* transcripts. It has a CADD score of 11.7; predicting it to be in the top 1% of all damaging variants in the genome [161] and is highly conserved across species.

The polycomb group (PcG) protein enhancer of zeste homolog 2, *EZH2*, is the catalytic subunit of the Polycomb Repressive Complex 2 (PRC2). Overall, *EZH2* acts as a histone methyltransferase (HMTase), trimethylating lysine 27 on the histone H3 protein subunit (H3K27me3) and as part of the PRC2 complex is ultimately responsible for long term transcriptional repression of its target genes [192-194]. Whilst it is known that *EZH2* expression is highly correlated with the progression of PCa, and is associated with disease aggressiveness and a poor prognosis [195,196], the mechanism by which the expression of *EZH2* increases during PCa is currently unknown [196-199]. We hypothesised that the rs78589034 variant may contribute to dysregulated *EZH2* expression. Whilst this intronic variant has been identified in a case of parathyroid neoplasm [200], this is the first study to find an association with PCa risk. In other cancers, germline and acquired variations in *EZH2* have been found to have both activating and inactivating effects in cancer, including in B-cell lymphomas, follicular lymphoma, and myelodysplastic and myeloproliferative disorders [201-203]. Interestingly, a male carrier of this variant in our family suffers from lymphoma (Figure 4.5; PC12-73).

Overall, the *EZH2* variant was determined to be significantly associated with PCa risk in the Tasmanian population (OR=3.27, p=0.001). Notably, the variant was not found to be enriched in any groups within our resource or compared to ExAC. This is because the carrier frequency of each group was not significantly different, however M_{QLS} analysis did find an association with disease. This is due to the fact that M_{QLS} [2] takes in to account the relatedness of individuals, can distinguish between unaffected controls and controls of unknown phenotype, incorporates phenotype data about relatives who have missing genotype data and obtains more power by giving increased weighting to those individuals with closely related disease-carrying relatives [2]. Though, the *EZH2* variant was more common in men with a family history of disease (1.61%) compared to those with no family history (0.61%), suggesting an inherited predisposition.

4.4.2   Examining the effect of the *EZH2* variant on *EZH2* gene and protein expression in prostate tumours

In the normal prostate, *EZH2* expression is relatively low, however it is reported to be overexpressed in diverse cancer types, including PCa [196]. Clinically localised PCa with high *EZH2* expression has a poorer prognosis compared to tumours with low expression [196]. Plus, metastatic PCa has been associated with higher levels of *EZH2* at both the transcriptional and translational level compared to clinically localised PCa [196]. Therefore, it has been suggested that *EZH2* expression could potentially predict disease progression and treatment outcomes [204]. A study by Saramaki and colleagues (2006) found that *EZH2* was upregulated in more than half of hormone refractory PCa tumours, compared to only 27% of early untreated PCa tumours [199]. *EZH2* is also recurrently mutated in several forms of cancer, particularly Non-Hodgkin's lymphoma, with heterozygous mutations at Y641 and A678V known as gain-of-function mutations which lead to hypertrimethylation of H3K27 [201,202]. Here, we hypothesised that the intronic *EZH2* variant may alter the regulation and expression of *EZH2*, given that *EZH2* is tightly regulated at the transcriptional, post-transcriptional and post-translational level [205,206]. EZH2 gene and protein expression was assessed in PCa cell lines and prostate tumours of variant carriers and non-carriers. Previous studies have identified higher *EZH2* expression in PC3 cells compared to LNCaP cells [190], however here, we found expression to be lowest in PC3 cells, which may be due to different growth conditions leading to altered expression. *EZH2* expression in all exonic regions was consistently lower in the two needle biopsy cores of the *EZH2* variant compared to the three non-carrier samples. Unfortunately, no conclusions could be made in regards to the difference in expression between variant carriers and non-carriers, due to very low levels of *EZH2* expression in the FFPE samples.

Protein expression data from the FFPE samples supported the gene expression findings and showed no EZH2 staining in any of the samples. The Human Protein Atlas (https://www.proteinatlas.org) reports that there is reasonably low expression of EZH2 in the prostate, with the majority of tumours showing low or undetectable expression [207]. In contrast, cDNA microarray profiling enabled Varambally *et al.* (2002) to conclude that *EZH2* was found to be overexpressed in invasive and hormone-refractory metastatic PCa [196]. This overexpression is thought to be due to amplification of the gene itself, or transcriptional upregulation by *MYC* and *ETS* gene family members [208,209], however we were unable to quantitate EZH2 gene or protein expression. The samples analysed in this study were primary

tumours of low to moderate GS (unlikely to become metastatic), therefore, low expression of *EZH2* at both the transcriptional and translational level is plausible.

4.4.3 Examining the effect of the *EZH2* variant on *EZH2* target gene expression in prostate tumours

Recent reports suggest that *EZH2* may promote PCa progression by repressing tumour suppressor gene targets, such as *CDH1, HOXA9* and *MSMB* [188,190,210]. These genes are silenced via two mechanisms; *EZH2* can directly bind to their promoter, or cause histone methylation, specifically H3K27me3. Both mechanisms lead to reduced expression of the target genes and increased cancer cell migration and invasion [210-212]. Reduced *CDH1* expression has been linked to metastasis in breast cancer, following studies of epithelial cell lines [213-215]. In our study, we observed the highest expression in bone metastasis cells, PC3, and the lowest expression in primary PCa cells, 22Rv1. We also observed an inverse relationship between *EZH2* and *CDH1* expression in 22Rv1 and LNCaP cells, but not in PC3 cell lines. It has been proposed that transcriptional repression of *CDH1* during *EZH2* overexpression is the result of PRC2 recruitment to the *CDH1* promoter by SNAIL [188,216]. Previous studies show that transient down-regulation of *CDH1* occurs in localised PCa [217], however no difference in expression was found between malignant and benign prostate glands in our study, following analysis of tumours from the *Tasmanian Prostate Tissue Pathology Resource*.

Changes in *HOXA9* expression have been associated with *EZH2* mutations in acute myeloid leukemia cases (p=0.048) [189]. Here, *HOXA9* expression was very low in all cell lines and needle biopsy samples, and expression appeared to be the same in the *EZH2* carrier compared to the non-carrier needle biopsy samples. It is known that *MSMB* expression is silenced by *EZH2* in advanced PCa cells, as the *MSMB* promoter binds to PRC2 and H3K27me3 when *EZH2* is overexpressed [190]. The H3K27 methylation-associated silencing of *MSMB* in such cells is believed to contribute to their increased growth, proliferation and invasive potential [218]. In fact, several studies have shown higher *MSMB* expression in benign *versus* malignant prostate tissue after a radical prostatectomy (RP) [219]. Our study included eight RP samples, however *MSMB* expression was no different between malignant and benign glands in these samples. Overall, expression of *MSMB* was very low in our FFPE samples and it was not possible to draw any conclusions from the small dataset.

4.4.4 Examining the effect of the intronic *EZH2* variant on splicing mechanisms

Splicing dysregulation is one of the molecular hallmarks of cancer [220] and the literature suggests that carcinogenesis often involves alternative splicing, which can result in protein diversity [220,221]. Chen *et al.* (2017) observed that alternative splicing involving the inclusion of exon 14 of *EZH2* plays a major role in the tumourigenesis of renal cancer, in their study of 24 clear cell renal cell carcinomas with matched malignant and benign cells [222]. The most common *EZH2* transcript in the prostate involves the inclusion of the full-sized exon 14, however the second and third most common transcript involve alternative splicing of exon 14 into three smaller exons (Figure 4.7). The *EZH2* gene can give rise to over 30 different mRNA transcripts [223] and multiple transcripts can exist in tissues; the functional implications of which are not yet known. Here, it was hypothesised that the rare intronic variant identified in PcTas12 may affect splicing. The *EZH2* variant lies 6bp away from the beginning of exon 16, which according to the Human Splicing Finder is an acceptor splice site [191]. The GTEx portal (https://gtexportal.org/home/) predicts that the variant causes protein truncation, following the identification and assessment of the variant in a sample of a tibial artery [137]. Using an *in vitro* splicing assay, our study detected no effect on *EZH2* splicing in the presence of the variant *versus* the wild-type allele. However, cell models may not accurately mimic the *in vivo* environment. In addition, only the exons downstream of exon 16 were assessed and the potential disruption to upstream splicing was not.

The literature suggests that disrupted expression of *EZH2* splicing factors, *SF3B1, SF3B3* and *U2AF1* can cause aberrant splicing and defective *EZH2* mRNA production [224]. Whilst this was not the focus of this study, these studies highlight the relationship of *EZH2* and its splicing factors, and overall, the potential effect of variants in a splice site recognition sequence. Sequence changes in recognition sites have been shown to affect splicing, and all splicing factors have preferred recognition sites. To determine whether there are particular splicing factors expressed in PCa, expression of *SF3B1, SF3B3* and *U2AF1* was initially assessed in PCa cells lines and our needle biopsy samples. Overall, *U2AF1* was highly expressed in all assayed samples compared to *SF3B1* and *SF3B3,* indicating that *U2AF1* is more prominent in the prostate. A study by Daures *et al.* (2018) of prostate biopsies divided into three clinical grades; normal (n=23), GS ≤7 tumour (n=20) and GS >7 tumour (n=19), identified that upregulation of six genes correlated with tumour severity, two of which were *EZH2* and *U2AF1*

[225]. Here, *U2AF1* expression was significantly higher in the metastatic cell lines, PC3 and LNCaP (GS >7 inferred) compared to the localised PCa cells, 22Rv1's (GS of 6 or 7 inferred).

U2AF1 is ultimately responsible for pre-mRNA splicing and mRNA 3'-end processing by recognising the AG di-nucleotide marking the end of the intron, and interestingly, binds directly to *EZH2* [226]. This 3' splice site recognition takes place in conjunction with a larger subunit, *U2AF2*. *U2AF1* recognises a polypyrimidine tract preceding the 3' splice site and directly positions *U2AF1* to recognise the downstream AG sequence. Thus, the variant may disrupt the target sequence for *U2AF2* to recognise the 3' splice site, and ultimately, may affect the affinity of *U2AF1* to bind to *EZH2*, thus resulting in ineffective *EZH2* mRNA transcript synthesis [227]. A polypyrimidine tract is considered strong if it contains four consecutive T bases within 30 nucleotides upstream of the 3' splice site [228]. Interestingly, the rs78589034 variant causes four consecutive T bases to occur (in comparison to the wild-type sequence; TT**C**T) just three nucleotides upstream of the AG start site. Thus, the *EZH2* variant forms a strong pyrimidine tract, which is recognised by *U2AF2,* causing *U2AF1* to be directly positioned over the AG start site, potentially resulting in an alternate transcript of *EZH2*.

There is also evidence to suggest that differential splicing is not an all or nothing phenomenon, rather that splicing is influenced by the bases surrounding the AG site in the target sequence [229]. Therefore, we aimed to investigate whether the formation of a strong polypyrimidine tract by the *EZH2* rs78589034 variant results in large or subtle effects of splicing factor expression. Following *U2AF1* expression analysis in the FFPE samples from the *Tasmanian Prostate Tissue Pathology Resource*, statistical analysis revealed that *EZH2* non-carriers trended towards having higher *U2AF1* expression compared to variant carriers, but this finding was only statistically significant in benign glands. It is unclear as to why this finding is only observable in benign prostate glands, but given that the variant form of *EZH2* is preferentially targeted by *U2AF2*, followed by *U2AF1*, it is likely that the difference we see is not solely due to the variant allele. Overall there are hundreds of other regions of the genome that *U2AF2* and *U2AF1* can bind to therefore, it is likely that a number of unknown factors are also contributing to reduced expression. Though, it is worth remembering that the rs78589034 variant is a risk allele, therefore it is possible that it has no effect on the splicing of the gene, or expression of its target genes and associated splicing factors, and if it does, the effects could be subtle.

Notably, there are two isoforms of *U2AF1* (*U2AF1α* and *U2AF1β*) and they differ by seven amino acids in the second RNA recognition motif within the U2AF homology motifs domain [230]. Tissue expression analysis demonstrated that *U2AF1α* is more highly expressed compared to *U2AF1β*, which may be due to the fact that these isoforms have different target sequence preferences [231,232]. Kralovicova and colleagues (2015) examined the effect of knocking down *U2AF1* (and its isoforms) in HEK293 cells and found that a small number of transcripts exhibited distinct responses to one isoform over the other, supporting the existence of isoform-specific interactions [231]. Analysis of altered targeted sequences (50 nucleotides) observed that 6bp from the beginning of an exon was important in the splicing process, which is the intronic location of the *EZH2* rs78589034 variant. It is plausible that the *EZH2* risk variant preferences a particular *U2AF1* isoform, however due to time constraints and resources we were unable to determine this here. It appears that the *EZH2* variant produces a 3' splice site which is preferred by *U2AF1β* [233], however this would need to be confirmed by targeted RNA sequencing of *U2AF1*. This would enable us to determine which transcript is preferentially expressed in our FFPE prostate samples and specifically, the *EZH2* carriers. It is hypothesised that preference for the *U2AF1β* isoform to bind to the variant form of *EZH2* could promote the inclusion of exon 16 in more *EZH2* transcripts. However, given that *U2AF1β* is expressed at a lower level to *U2AF1α* it is likely that that the variant may have only subtle effects on the splicing mechanisms of *EZH2*, slightly altering the expression of *EZH2* transcripts, yet this was undetectable here.

4.4.5   Limitations of this study

This study has identified a Tasmanian PCa risk variant in *EZH2* and has assessed its effect on the splicing mechanisms of *EZH2*, however, several limitations of the study must be acknowledged. To date, this study is the first to find an association of the rs78589034 variant with PCa risk, therefore it has not been replicated in independent populations of PCa cases and controls. Thus, we are unaware as to whether it contributes to PCa risk in other populations. Overall, availability of FFPE samples and the rarity of the variant restricted our opportunity to identify *EZH2* variant carriers, and thus limited the availability of informative tumour tissue specimens from carriers. A small sample size results in reduced power and therefore lowers the likelihood of detecting real associations. Thus, the concepts explored in this study should be accessed in a larger tissue cohort of *EZH2* carriers. The quality of DNA and RNA extracted from FFPE tissue is also fairly poor, therefore it is important that our findings are validated in

larger FFPE cohorts. Data from the GTEx Portal (https://gtexportal.org/home/) [137] suggests that each region of *EZH2* assessed have similar expression levels, therefore the discrepancies in levels of *EZH2* expression between different regions of the gene is likely due to poor quality RNA samples and different primer efficiencies. Thus, we were only able to compare expression within regions and not between. The analysis discussed earlier (4.3.6) was used to determine the region of *EZH2* that we were able to efficiently quantify. In terms of the needle biopsy samples we do not definitively know which lobe contains benign or malignant tissue, therefore we cannot draw any real conclusions here. On another note, the splicing assay was performed *in vitro* in PCa cells, PC3 and 22Rv1's. This approach directly detects the effect of the intronic variant on splicing in these cells, but does not replicate the *in vivo* environment in the prostate. For example, this assay does not take into account the effects of other events associated with splicing, such as transcription, capping and polyadenylation, or other proteins and complexes involved in splicing [234]. Lastly, the only positive control we had for this experiment was the splicing together of the rat insulin exons.

4.5   FUTURE DIRECTIONS

Overall, this study identified that the *EZH2* rs78589034 variant is significantly associated with PCa risk in the Tasmanian population, although enrichment in familial or sporadic cases was not demonstrated. Given that it is yet to be replicated in other populations, we aim to further explore the contribution of this variant to independent PCa populations through collaboration with members of the International Consortium of Prostate Cancer Genetics (ICPCG) and the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium. *EZH2* expression was unable to be detected in the FFPE prostate tissue samples, however it may be because *EZH2* expression in these FFPE prostate samples is too low to be detected by RT-qPCR. It would be advantageous to use a more precise platform for gene expression quantification, such as droplet digital PCR (ddPCR™, Bio-Rad) or the QuantStudio 3D Digital PCR Chip platform (ThermoFisher Scientific). Both of these systems can detect very low gene copy numbers from minimal sample input; therefore, they could be utilised to determine if the rs78589034 variant alters *EZH2* expression in our *Tasmanian Prostate Tissue Pathology Resource*. *U2AF1* expression could be quantitated in these samples and it was found to be downregulated in *EZH2* carriers *versus* non-carriers, however this was only statistically significant in benign cells. Follow-up *in vitro* studies analysing the interaction of *EZH2* and *U2AF1* in cell lines with and without the intronic variant

would be valuable. This could be achieved by ChIP-sequencing, which analyses protein and DNA interactions. Overall, the *Tasmanian Prostate Tissue Pathology Resource* consists of only seven *EZH2* variant carriers, therefore it would be worthwhile to undertake targeted collection of FFPE prostate samples from newly diagnosed or 'pathology only' cases from PcTas12. Collection of additional tumours from PCa cases in other *Tasmanian Familial Prostate Cancer Cohort* families will increase our sample size and statistical power. This study has also suggested that the presence of the *EZH2* variant produced no detectable difference in *EZH2* splicing. As mentioned, the splicing assay assessed exons downstream of the variant, therefore it would be beneficial to assess splicing of exons 15 and 16 in *EZH2* variant carriers in comparison to non-carriers to determine if they are affected. Alternatively, RNA sequencing data for the single needle biopsy *EZH2* carrier (PT0018) is now available, therefore, we aim to assess expression of *EZH2* transcripts in this variant carrier.

4.6   CONCLUSION

This study aimed to prioritise rare variants segregating with PCa, following WGS of individuals from family, PcTas12. Subsequent genotyping of the larger *Tasmanian Familial Prostate Cancer Study* cohorts found an intronic variant in *EZH2* (rs78589034) to be significantly associated with PCa risk (OR=3.27, p=0.001). Given that this association has not been previously described, validation in larger cohorts of PCa cases and controls is warranted. Presented here are preliminary findings assessing the functional effect of the intronic variant on *EZH2* gene and protein expression, the splicing capabilities of *EZH2* and *EZH2* splicing factor and target gene expression. A larger sample size of fresh, frozen prostate tissue will prove fruitful for this study.

CHAPTER 5 : IDENTIFICATION AND FUNCTIONAL ASSESSMENT OF A RARE PROSTATE CANCER RISK VARIANT IN *HOXB13*

Publications arising from this chapter:

FitzGerald LM*, **Raspin K***, Marthick JR, *et al*. **Impact of the G84E variant on HOXB13 gene and protein expression in formalin-fixed, paraffin embedded prostate tumours**. *Sci Rep* 2017; 7:17778. *Joint first authors.

5.1   INTRODUCTION

In recent years, a number of rare prostate cancer (PCa) susceptibility genes have been identified, however the *HOXB13* gene is the only one that has been consistently replicated [11,151-156]. It has also been shown that many breast and ovarian cancer predisposition genes, including *BRCA1, BRCA2, CHEK2* and *ATM* increase the risk of PCa [235-238], suggesting that shared genetic factors predispose to multiple cancer types. To explore this theory, Leongamornlert *et al.* (2019) recently used a targeted sequencing approach to screen 1281 young-onset PCa cases and 1160 controls for protein truncating variants in known prostate, breast and ovarian cancer predisposition genes [239]. The study identified 233 unique variants in 97 genes, each of which had minor allele frequency's (MAF) of less than 0.50% in their control population. Gene-set analysis found a subset of 20 genes associated with increased PCa risk (OR=3.2, p=4.1x10^{-3}) [239]. The gene list covered 167 DNA repair genes and eight PCa candidate genes, with many of these from the breast and ovarian cancer associated (BROCA) cancer risk panel designed by Walsh and colleagues (2010) [172-176]. DNA repair genes are crucial regulators of DNA damage and repair, and therefore, their dysregulation can lead to genomic instability and ultimately, cancer [240]. Previous studies have observed variants in DNA repair genes in only 2% of early low-to-intermediate risk PCa, whereas this frequency increases to 6% in high-risk localised disease and up to 12% in metastatic disease [241,242]. In fact, it is now recommended that germline testing for variants in *BRCA2* and *ATM* is undertaken in all men with high-risk localised PCa, or more advanced, metastatic disease [243].

Previously we have used an agonistic approach to gene discovery, however a more targeted approach was used here. A targeted approach enables the concurrent identification of novel PCa predisposition variants and validation of previous associations. In this chapter we took an alternative approach by selecting a panel of 36 genes for examination. Whole-genome sequencing (WGS) data from five *Tasmanian Familial Prostate Cancer Cohort* families were screened for potential disease-associated rare coding variants in 36 genes (Appendix 11). The gene list comprised known PCa predisposition genes, including *HOXB13* [11], *MSR1* [239], *TANGO2* and *CHAD* [244], genes associated with breast cancer, such as *BRCA1* and *BRCA2* [235], and other DNA repair genes from the BROCA gene set, including *ATM* [172-176]. The results presented in this chapter are published in Scientific Reports [1].

5.2   METHODS

5.2.1   Whole-genome sequencing analysis

In total, 33 individuals from five Tasmanian PCa families were selected for WGS as described in Chapter 3.2.1. The data were analysed, annotated and variants called as previously mentioned (Chapter 3.2.2), with each family analysed separately. In this study, only variants in the 36 candidate genes were prioritised further (Appendix 11). Prioritisation was guided by the frequency of the variant in the Exome Aggregation Consortium (ExAC; non-Finnish European, non-TCGA (The Cancer Genome Atlas) population), a publicly available database consisting of sequencing data from 60,706 unrelated individuals [158]. Variants with a MAF of <2% in ExAC were prioritised for segregation analyses. Rare, segregating variants with evidence of functional consequences using *in silico* functional prediction tools, such as SIFT [159], PolyPhen2 [160] and CADD (Combined Annotation Dependent Depletion; model v1.3) [161] were prioritised. The carrier frequency of the prioritised variants were determined in the eight Tasmanian controls with WGS data, plus a literature search using online search engines, ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) [163] and PubMed to determine if the variant has previously been associated with cancer.

5.2.2   Validation, segregation and association analysis of prioritised rare variants

Variants prioritised from the WGS data were validated by Sanger sequencing and sequencing of additional family members was used to track segregation with disease, as described previously (Chapter 3.2.3). For those cases without gDNA, formalin-fixed paraffin embedded (FFPE) DNA was sequenced for the prioritised rare variants (Appendix 2). If found to

segregate, rare variants were screened in the entire *Tasmanian Familial Prostate Cancer Cohort* and the *Tasmanian Prostate Cancer Case-Control Study*, as discussed in Chapter 3.2.4 (Appendix 6). M_{QLS} analysis [2] was used to determine if there was an association between the prioritised rare variants and PCa risk in the Tasmanian population as discussed in Chapter 3.2.5.

5.2.3   Quantification of *HOXB13* gene expression

Absolute *HOXB13* gene expression (ENST00000290295.7) was analysed as discussed in Chapter 2.3 (Appendix 3). The absolute copy number of *HOXB13* was normalised to the copy number of the housekeeping genes, *β-Actin* and *GAPDH*. RT-qPCR primers were designed to the most commonly transcribed isoform in the prostate (as per GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2; https://gtexportal.org/home/)) [137] and are displayed in Appendix 3.

5.2.4   Allele-specific next-generation sequencing

A region surrounding the *HOXB13* variant was analysed using the Illumina MiSeq next-generation sequencing approach (Appendix 12). cDNA samples were PCR amplified and visually assessed by agarose gel electrophoresis. Amplicons were then quantitated with the Qubit 2.0 Fluorometer, using the dsDNA broad range sensitivity kit (Life Technologies), according to the manufacturer's instructions. PCR products were diluted to 0.5ng/μL and then barcoded with a forward and reverse tag each, according to the conditions in Appendix 1. Each sample was barcoded with its own unique combination of forward and reverse tags, which were 10bp in length, with i5 and i7 adaptors 20bp in length. These barcodes were designed by our collaborator, Andrea Polanowski (Australian Antarctic Division; Appendix 13). Barcoded DNA fragments were pooled, purified and quantitated, as previously described, and the 2nM library was sequenced on the Illumina MiSeq platform using the MiSeq® V2 300 Cycle Reagent Kit (Illumina).

FastQ files were aligned to the reference genome (hg19) using the web interface wrapper, Galaxy version 16.04 [245,246]. The FastQ files were converted to Sanger and Illumina 1.8+ format using the FASTQ Groomer tool, followed by realignment using BWA-MEM. The allele frequency at the variant position (rs138213197, G84E) was visualised using IGV 2.3.68 [247].

FastQC of BAM files was used to assess the quality of the raw sequence data (an example is shown in Appendix 14).

5.2.5 Quantification of HOXB13 protein expression

Immunohistochemistry (IHC) was undertaken as per Chapter 2.4 (Appendix 5). Normal prostate glands (Abcam) ascertained as wild-type for the *HOXB13* variant by Sanger sequencing, were used as a positive control. Negative controls included primary antibody only, secondary antibody only, and a mouse IgG$_1$ isotype control (Dako).

5.2.6 Allele-specific methylation analysis

FFPE DNA (~200ng) was bisulphite converted using the EZ DNA Methylation-Gold^{TM} Kit (Zymo Research Corp), as per the manufacturer's instructions. Two primer sets were designed to amplify fragments covering the *HOXB13* promoter/exon 1 CpG island and a CpG island ~4.5kb upstream of the transcription start site (Appendix 12), using MethPrimer [248]. Amplification was performed according to the conditions in Appendix 1. Fragments were purified using the QIAquick Gel Extraction Kit (Qiagen), as per the manufacturer's instructions, and were cloned into the p-GEM®-T Easy Vector Kit (Promega Corporation), using a 3:1 ratio of insert to vector. Top10 competent cells (Invitrogen) were transformed with 2μL of ligations. Ten white clones per sample were selected for amplification and DNA extraction, using the QIAprep Spin Miniprep Kit (Qiagen), as per the manufacturer's instructions. Inserts in the clones were sequenced using the reverse *Sp6* primer (ThermoFisher Scientific). Each CpG site, for each clone was scored as either 1, methylated or 0, unmethylated, and bubble maps were generated using the CpG Bubble Chart Generator, Version 20061209 Alpha, created by Mark A Miranda.

A 175bp region of *HOXB13,* including the G84E mutation and nine surrounding CpG sites, was PCR amplified using bisulphite-converted FFPE DNA, as described previously in Chapter 2.3 (Appendix 12). Products were barcoded with unique forward and reverse tags (Appendix 13) and sequenced on the Illumina MiSeq platform, as described above (Chapter 5.2.4). FASTQ files were quality score checked and separated into reads containing the G84E variant allele and the wild-type allele. A beta value (β), the ratio of methylated *versus* unmethylated reads, was determined for all nine CpG sites. An unpaired Student's t-tests was used to compare

methylation in reads containing the G84E variant allele *versus* reads with the wild-type (comparison of β values). P values <0.05 were considered to be statistically significant.

5.3   RESULTS

5.3.1   Rare variant prioritisation

Thirty-three individuals from five PCa families were successfully WGS, including five affected men from PcTas3 (Table 3.2 and Figure 3.3), five individuals from PcTas4 (Table 3.4 and Figure 3.7), eleven individuals from PcTas22 (Table 3.6/8 and Figure 3.10), three individuals from PcTas12 (Table 4.1 and Figure 4.4) and nine individuals from PcTas72 (Figure 5.1, Table 5.1 and Figure 5.2).
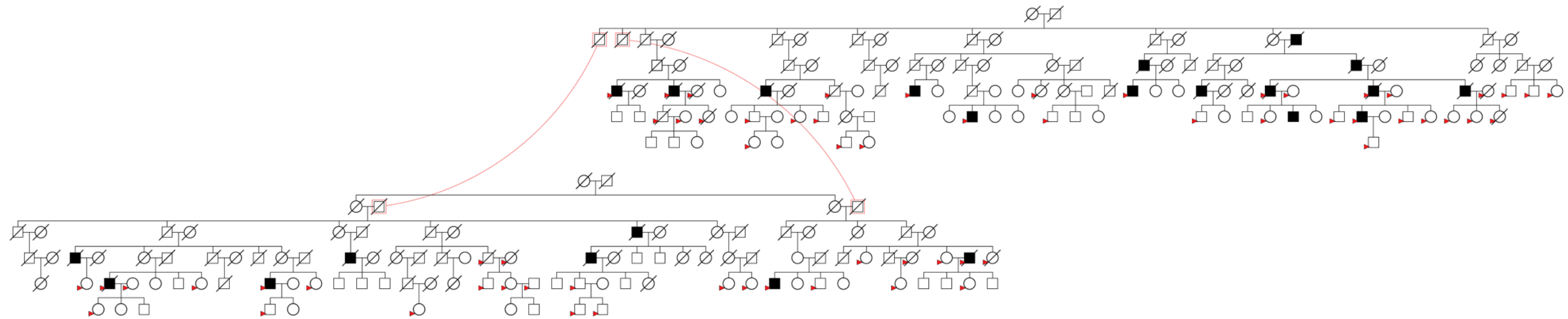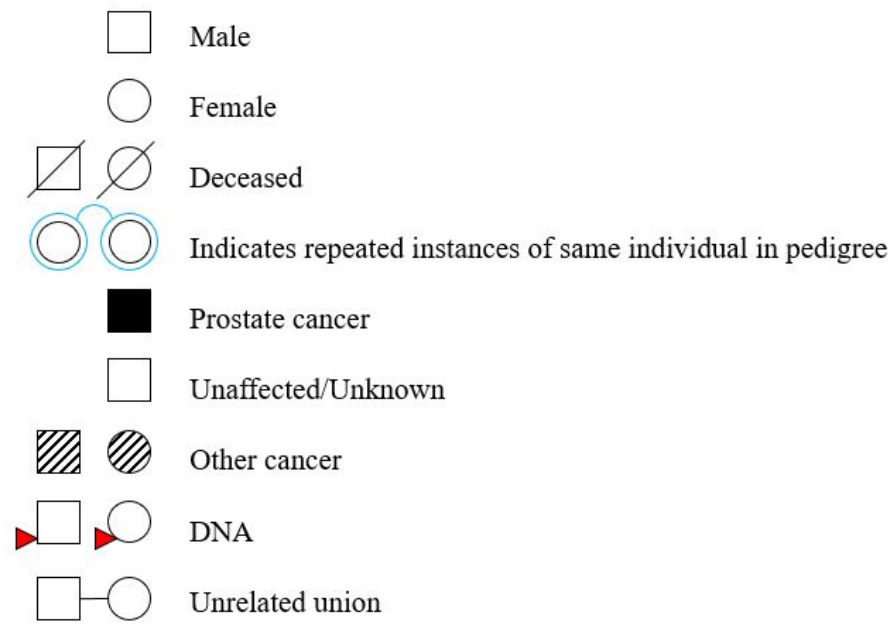
Figure 5.1 PcTas72 Pedigree.

This a pedigree of family, PcTas72, depicting the number and relationships of PCa cases (shown in shaded squares), as well the availability of DNA from cases and their unaffected relatives, which is represented by red arrows. Please note, in each instance, the disease status for earlier generations is generally unknown, unless this information was obtained from clinical records. And if so, these individuals have been marked as affected in the pedigrees. This pedigree is included to illustrate the size of the pedigree only, please refer to Figure 5.2 and 5.3 for individual annotations.

Table 5.1 Clinicopathological characteristics of individuals from PcTas72 chosen for whole-genome sequencing.

Sample Identification	Sex	Prostate Cancer Affection Status	Age at diagnosis	Tumour Grade[1]	Contemporary Gleason score[2]
PC72-02	Male	Affected	76	MD	6 (3+3)
PC72-03	Male	Affected	67	WD	4 (2+2)
PC72-04	Male	Affected	70	PD	9 (4+5)
PC72-75 [WES]	Female	N/A	N/A	N/A	N/A
PC72-94	Male	Unaffected	68^	N/A	N/A
PC72-97	Female	N/A	N/A	N/A	N/A
PC72-106	Male	Unaffected	58*	N/A	N/A
PC72-126	Male	Affected	51	-	6 (3+3)
PC72-188	Male	Unaffected	33*	N/A	N/A

WES: Whole-exome sequenced; ^: Age at death; *Unaffected, age at WGS; [1]Tumour grade obtained from pathology report; [2]Contemporary Gleason Score from FFPE tissue block chosen for macrodissection of nucleic acids and IHC; WD: well differentiated; MD: moderately differentiated;

PD: poorly differentiated; -: information not present in original pathology report; N/A: not applicable.
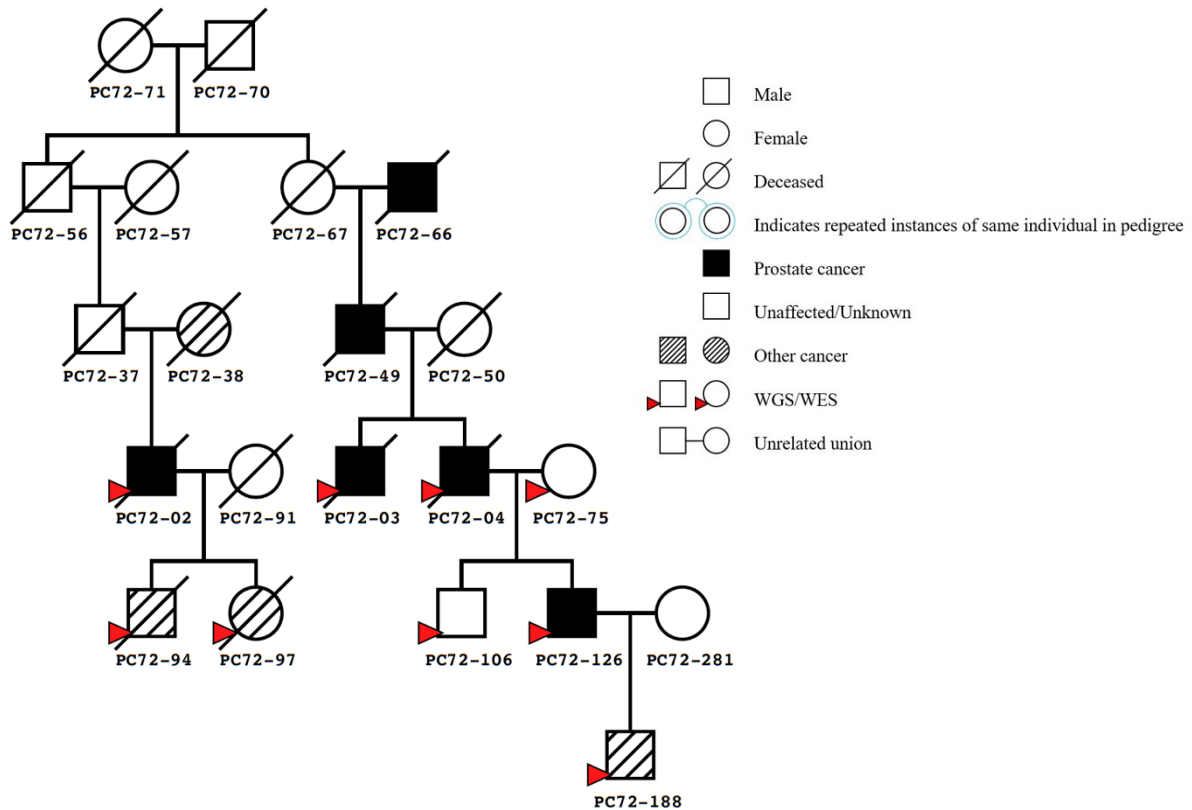
Figure 5.2 A condensed PcTas72 pedigree showing individuals chosen for whole-genome or whole-exome sequencing.

Individuals chosen for WGS are indicated by red arrows, in this case, four PCa cases, three PCa unaffected male relatives and 2 female relatives were chosen. Note: PC72-75 has WES data.

The WGS data from the five families was screened for rare variants in 36 cancer-associated genes, including DNA repair genes and previously identified PCa predisposition genes (Appendix 11). In total, 17 rare variants and seven novel variants were identified (in *AR, ATM, BRCA1, BRCA2, HOXB13, NBN, NKX3-1, OR5H14, PALB2, RAD51C, RNASEL, SLX4* and *TANGO2;* Table 5.2 and Appendix 15), however only five had a CADD score >15 and were found in two or more affected individuals (Table 5.2). The known PCa risk variant, *HOXB13* G84E was also identified in a single PCa case in PcTas72. Given that the *HOXB13* G84E variant is known to be associated with PCa [11], the focus of the remainder of this chapter is characterising the contribution of this variant to the Tasmanian population and, secondly, understanding its functional impact as this has not yet been established.

Subsequent genotyping of PcTas72 identified an additional five carriers, including three PCa cases and a female carrier, PC72-97, with Mantle Cell Lymphoma (Figure 5.3). Segregation of the *HOXB13* variant was observed in two branches of PcTas72.

Table 5.2 Prioritised rare variants in known cancer-associated genes following whole-genome sequencing of five Tasmanian prostate cancer families.

Gene	rs number	Chromosome: base pair	ExAC[1] MAF (%)	Family Identification	Segregation in WGS individuals (affected carriers/unaffected carriers)	CADD[2] Score	Allele Change; Amino Acid Change	Number of Control Carriers	ClinVar Search[3]
ATM	rs56128736	11:108,119,723	0.20	PcTas22 Main	2 out of 5/ 0 out of 1	23.4	T > C; V410A	0 out of 8	Not reported
ATM	rs1800058	11:108,160,100	1.27	PcTas22 Sub	2 out of 4/ 1out of 1	16.54	C > T; L1420F	1 out of 8	Hereditary cancer: Benign
ATM	rs4986761	11:108,124,511	0.70	PcTas72	2 out of 4/ 1 out of 4	19.39	T > C; S707P	0 out of 8	Hereditary cancer: Benign

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; MAF: Minor allele frequency; N/A: Not found in ExAC or ClinVar; WGS: Whole-genome sequenced; [2]CADD: Combined Annotation Dependent Depletion [164]; Control: Control from the *Tasmanian Prostate Cancer Case-Control Study;* eight were WGS; [3]Associated condition: Interpretation of variant [163]. An additional 18 variants were also identified in the 36 cancer-associated genes that were selected (Appendix 11) and they are shown in Appendix 15.

Gene	rs number	Chromosome: base pair	ExAC[1] MAF (%)	Family Identification	Segregation in WGS individuals (affected carriers/unaffected carriers)	CADD[2] Score	Allele Change; Amino Acid Change	Number of Control Carriers	ClinVar Search[3]
HOXB13	rs138213197	17:46,805,455	0.22	PcTas72	1 out of 4/ 1 out of 4	22.7	C > T; G84E	0 out of 8	Hereditary prostate cancer: Pathogenic
RAD51C	rs61758784	17:56,772,272	0.35	PcTas72	3 out of 4/ 0 out of 4	21.7	G > A; A126T	0 out of 8	Hereditary cancer: Benign
RNASEL	Novel	1:182,555,547	N/A	PcTas3	2 out of 5	16.21	G > A; A132V	0 out of 8	N/A

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; MAF: Minor allele frequency; N/A: Not found in ExAC or ClinVar; WGS: Whole-genome sequenced; [2]CADD: Combined Annotation Dependent Depletion [164]; Control: Control from the *Tasmanian Prostate Cancer Case-Control Study;* eight were WGS; [3]Associated condition: Interpretation of variant [163]. An additional 18 variants were also identified in the 36 cancer-associated genes that were selected (Appendix 11) and they are shown in Appendix 15.
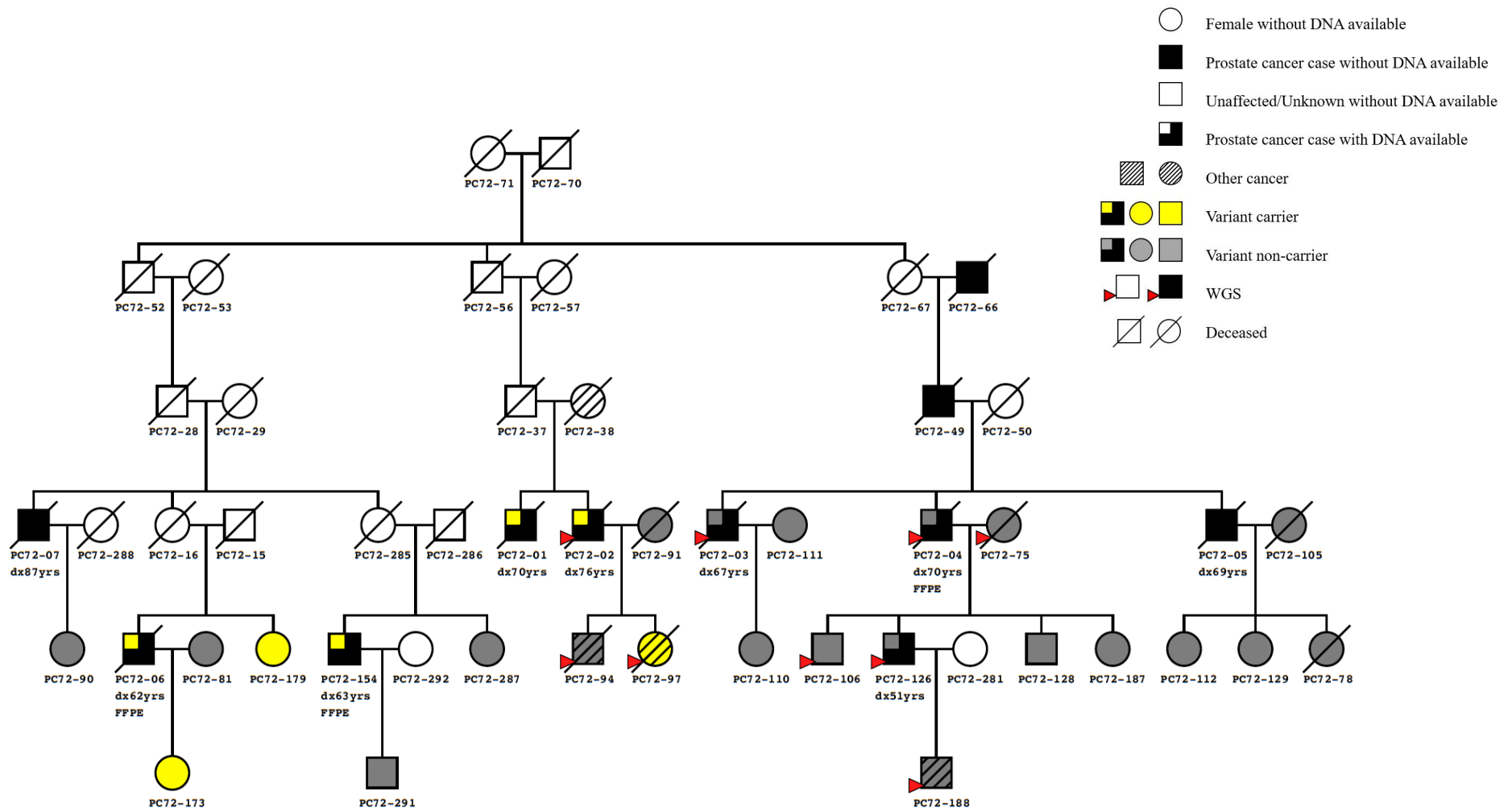
Figure 5.3 *HOXB13* G84E variant carriers in PcTas72.

This is a condensed pedigree of PcTas72 comprising all *HOXB13* variant carriers (shown in yellow) and their relationship. Non-variant carrier family members are shown in grey and the individuals who were WGS are indicated by red arrows.

5.3.2 Association of the *HOXB13* variant with prostate cancer risk in Tasmania

Screening of 94 Tasmanian controls, as described in Chapter 2.1.3, revealed the absence of any *HOXB13* G84E carriers. Following TaqMan genotyping of our entire *Tasmanian Familial Prostate Cancer Cohort* (n=714) and *Tasmanian Prostate Cancer Case-Control Study* (n=853), a further 8 familial cases from an additional five PcTas families, plus 3 unaffected relatives, 3 sporadic cases and 1 control were identified. M_{QLS} analysis [2] demonstrated a significant association between the *HOXB13* variant and PCa risk in the Tasmanian population (OR=6.59, p=$4.2x10^{-5}$). The number of familial case carriers was much higher compared to their unaffected family members, and the sporadic cases carried the variant at a lower percentage than the affected and unaffected family members (Table 5.3). The variant was also assessed for enrichment in groups within the *Tasmanian Familial Prostate Cancer Study* cohorts, as well as in comparison to ExAC (Table 5.4). The *HOXB13* G84E variant was enriched in the Tasmanian familial PCa cases *versus* the controls (p=0.03), however it was not enriched in any Tasmanian patient group compared to ExAC. Familial case carriers in our cohort did have a higher carrier frequency compared to ExAC, yet this was not significant (p=0.08).

Table 5.3 The association of the *HOXB13* variant with prostate cancer risk in the *Tasmanian Familial Prostate Cancer Study* cohorts.

Gene	Variant	Founder Family	Other PcTas Families	Familial Case Carriers (n=249)[1]	Familial Unaffected Carriers (n=448)[1]	Sporadic Case Carriers (n=495)[1]	Control Carriers (n=341)[1]	ExAC[2] MAF (%)	Odds Ratio	p-value
HOXB13	rs138213197	PcTas72	12, 22, 63, 213, 3250	8 (3.21%)	6 (1.34%)	3 (0.61%)	1 (0.29%)	0.22	6.59	4.22×10^{-5}*

Familial case and familial unaffected comprise the *Tasmanian Familial Prostate Cancer Cohort*; Sporadic case and control comprise the *Tasmanian Prostate Cancer Case-Control Study;* [1](n=total sample size); [2]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; MAF: minor allele frequency; *Significant p-value

Table 5.4 Comparison of *HOXB13* variant carrier status in our *Tasmanian Familial Prostate Cancer Study* cohorts compared to ExAC or Tasmanian controls.

Gene	Variant		Entire Resource *versus* ExAC[1]	Familial & Sporadic Cases *versus* ExAC[1]	Familial Cases *versus* ExAC[1]	Sporadic Cases *versus* ExAC[1]	Controls *versus* ExAC[1]	Familial & Sporadic Cases *versus* Controls	Familial Cases *versus* Controls	Sporadic Cases *versus* Controls
HOXB13	rs138213197 (ExAC[1] MAF 0.33%)	Chi square; 1df p-value	0.78 (+)[2] p=0.38	1.80 (+) p=0.18	3.02 (+) p=0.08	0.03 (-) p=0.86	1.56 (-) p=0.21	3.44 (+) p=0.06	4.46 (+) p=0.03*	0.46 (+) p=0.50
		Number of carriers (n=total sample size)	18 (n=1,533) *versus* 89 (n=26,596)	11 (n=744) *versus* 89 (n=26,596)	8 (n=249) *versus* 89 (n=26,596)	3 (n=495) *versus* 89 (n=26,596)	1 (n=341) *versus* 89 (n=26,596)	11 (n=744) *versus* 1 (n=341)	8 (n=249) *versus* 1 (n=341)	3 (n=495) *versus* 1 (n=341)

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; Entire Resource includes the *Tasmanian Familial Prostate Cancer Cohort* and the *Tasmanian Prostate Cancer Case-Control Study*; Familial cases are a part of the *Tasmanian Familial Prostate Cancer Cohort*; Sporadic case and control comprise the *Tasmanian Prostate Cancer Case-Control Study;* [2]In the chi square test (+/-) indicates directionality, where (+) means the minor allele frequency is greater in the first named population *versus* the comparison dataset, whereas, (-) indicates it is more enriched in the second named population; *Significant p-value.

5.3.3 Association of the *HOXB13* variant with clinical characteristics and tumour pathology

Following genotyping of the *Tasmanian Familial Prostate Cancer Study* cohorts, the *HOXB13* G84E variant was identified in an additional five families. The age at diagnosis of carriers *versus* non-carriers was similar in all six families (Table 5.5). Likewise, for tumour pathology, the Gleason score (GS) of carriers and non-carriers was similar. On average the majority of tumours had a GS of 6 (3+3) or 7 (3+4) (Table 5.5).

Table 5.5 Clinicopathological characteristics of prostate cancer cases from the six *HOXB13* variant carrier families, including G84E carriers and non-carriers.

Sample Identification	*HOXB13* G84E Genotype	Age at Diagnosis	Tumour Grade[1]	Gleason Score[2]
PC72-01	CT	70	WD	-
PC72-02	CT	76	MD	6 (3+3)
PC72-06	CT	62	-	8 (4+4)
PC72-154	CT	63	WD	4 (2+2)
PC72-03	CC	67	WD	4 (2+2)
PC72-04	CC	70	PD	9 (4+5)
PC72-05	CC	69	WD	6 (3+3)
PC72-07	CC	87	-	-
PC72-08	CC	75	-	-
PC72-09	CC	69	WD	-
PC72-77	CC	66	PD	10 (5+5)
PC72-114	CC	57	-	-
PC72-126	CC	51	-	6 (3+3)
PC72-134	CC	76	-	-
PC72-150	CC	96	PD	-
PC72-293	CC	57	M/PD	7 (3+4)
PC72-303	CC	59	-	-
PC72-306	CC	50	M/PD	7 (4+3)
PC72-307	CC	73	-	8 (4+4)

[1]Tumour grade obtained from pathology report; [2]Gleason Score obtained from pathology report; WD: well differentiated; MD: moderately differentiated; PD: poorly differentiated; M/PD: moderately-poorly differentiated; -: information not present in original pathology report.

Sample Identification	*HOXB13* G84E Genotype	Age at Diagnosis	Tumour Grade[1]	Gleason Score[2]
PC12-03	CT	62	WD	4 (2+2)
PC12-07	CT	59	PD	9 (4+5)
PC12-08	CT	73	-	6 (3+3)
PC12-01	CC	63	MD	6 (3+3)
PC12-02	CC	80	MD	6 (3+3)
PC12-04	CC	63	MD	6 (3+3)
PC12-05	CC	65	WD	-
PC12-06	CC	80	PD	7 (3+4)
PC12-09	CC	68	-	6 (3+3)
PC12-132	CC	61	-	8 (4+4)
PC12-187	CC	71	-	6 (3+3)
PC12-254	CC	76	WD	6 (3+3)
PC22-203	CT	79	PD	8 (4+4)
PC22-576	CT	69	M/PD	7 (3+4)
PC22-637	CT	70	-	8 (4+4)
PC22-01	CC	72	WD	5 (2+3)
PC22-02	CC	64	MD	6 (3+3)
PC22-03	CC	62	WD	-
PC22-04	CC	57	MD	6 (3+3)
PC22-05	CC	85	M/PD	8 (4+4)
PC22-06	CC	63	WD	-
PC22-07	CC	61	WD	-
PC22-16	CC	74	WD	-
PC22-17	CC	56	MD	6 (3+3)
PC22-21	CC	69	-	6 (3+3)
PC22-167	CC	69	WD	-
PC22-169	CC	60	M/PD	7 (3+4)
PC22-183	CC	69	MD	6 (3+3)
PC22-239	CC	64	MD	7 (3+4)
PC22-246	CC	66	MD	6 (3+3)
PC22-249	CC	59	-	-
PC22-387	CC	83	-	8 (4+4)
PC22-416	CC	58	MD	8 (3+5)

[1]Tumour grade obtained from pathology report; [2]Gleason Score obtained from pathology report; WD: well differentiated; MD: moderately differentiated; PD: poorly differentiated; M/PD: moderately-poorly differentiated; -: information not present in original pathology report.

Sample Identification	*HOXB13* G84E Genotype	Age at Diagnosis	Tumour Grade[1]	Gleason Score[2]
PC22-387	CC	83	-	8 (4+4)
PC22-416	CC	58	MD	8 (3+5)
PC22-584	CC	63	MD	7 (3+4)
PC22-589	CC	72	-	7 (4+3)
PC22-657	CC	60	-	6 (3+3)
PC22-660	CC	69	-	-
PC22-698	CC	55	-	-
PC63-01	CC	67	WD	-
PC63-02	CC	63	WD	-
PC63-03	CC	74	MD	-
PC63-06	CC	72	-	7 (3+4)
PC63-12	CC	78	WD	2 (1+1)
PC63-18	CC	62	PD	9 (5+4)
PC63-24	CC	67	MD	6 (3+3)
PC63-74	CC	65	W/MD	7 (3+4)
PC63-133	CC	60	-	-
PC63-286	CC	62	MD	6 (3+3)
PC63-293	CC	63	-	6 (3+3)
PC213-13	CT	59	MD	6 (3+3)
PC213-01	CC	68	WD	6 (3+3)
PC213-17	CC	65	-	6 (3+3)
PC213-106	CC	75	-	-
PC213-516	CC	68	-	8 (5+3)
PC213-712	CC	61	-	6 (3+3)
PC213-718	CC	86	WD	4 (2+2)
PC213-731	CC	75	-	6 (3+3)
PC213-756	CC	71	-	7 (3+4)
PC213-772	CC	73	-	9 (4+5)
PC213-833	CC	61	-	7 (3+4)
PC213-845	CC	63	-	6 (3+3)
PC213-861	CC	58	-	7 (4+3)
PC213-874	CC	61	-	6 (3+3)
PC213-881	CC	64	-	6 (3+3)

[1]Tumour grade obtained from pathology report; [2]Gleason Score obtained from pathology report; WD: well differentiated; MD: moderately differentiated; PD: poorly differentiated; W/MD: well-moderately differentiated; -: information not present in original pathology report.

Sample Identification	*HOXB13* G84E Genotype	Age at Diagnosis	Tumour Grade[1]	Gleason Score[2]
PC213-883	CC	72	-	9 (4+5)
PC213-935	CC	62	-	7 (4+3)
PC213-938	CC	75	-	8 (4+4)
PC213-946	CC	58	-	7 (3+4)
PC213-971	CC	55	-	7 (4+3)
PC213-991	CC	68	PD	9 (4+5)
PC3250-01	CT	51	PD	9 (4+5)

[1]Tumour grade obtained from pathology report; [2]Gleason Score obtained from pathology report; PD: poorly differentiated; -: information not present in original pathology report.

5.3.4   Targeted collection of prostate tumour specimens from *HOXB13* variant carriers

Targeted collection of FFPE specimens from local pathology laboratories was undertaken for known *HOXB13* variant carriers (n=4), as well as a random selection of G84E non-carriers (n=7). Where possible, we also collected tumour specimens from affected relatives of known carriers for whom we didn't have a germline sample available (n=11). Genotyping of prostate tissue DNA from the 22 blocks confirmed four and identified five additional heterozygous G84E carriers, including a case whose germline DNA was genotyped as wild-type (PC22–203; Table 5.6). Repeat genotyping of PC22–203 germline and re-extracted tumour DNA samples confirmed the discordant result. First-degree relatives of this individual were genotyped as G84E wild-type. No additional samples were available for this individual (deceased) therefore, this anomaly could not be resolved to determine whether a pathology sample mix-up had occurred, mosaicism was present in the individual or the variant arose somatically.

Clinical analyses of the FFPE specimens revealed no detectable difference in the age at diagnosis of the G84E variant carriers (n=9) *versus* non-carriers (n=13; p=0.22, Table 5.6). For those samples with malignant glands present, there was no detectable difference observed in GS between carriers (n=9) and non-carriers (n=10; p= 0.86, Table 5.6).

Table 5.6 Clinicopathological characteristics of FFPE prostate tumour samples obtained for *HOXB13* **carriers and non-carriers used in the functional analyses of this chapter.**

Sample Identification	Age at Diagnosis	Germline G84E Genotype	Tissue Source	Tumour G84E Genotype	Tumour Grade[1]	Contemporary Gleason Score[2]
PC4-03	80	CC	TURP	CC	M/PD	7 (4+3)
PC11-11	85	N/A	TURP	CC	-	7 (3+4)
PC11-12	58	N/A	TURP	CC	-	9 (4+5)
PC11-13	72	N/A	TURP	CC	-	Benign
PC11-16	78	N/A	TURP	CC	-	5 (2+3)
PC12-01	63	CC	RP	CC	MD	6 (3+3)
PC12-06	80	N/A	TURP	CC	PD	7 (3+4)
PC12-09	68	N/A	TURP	CC	-	6 (3+3)
PC22-06	63	CC	TURP	CC	WD	Benign
PC47-02	68	CC	TURP	CC	WD	Benign
PC60-01	58	CC	TURP	CC	WD	6 (3+3)
PC63-24	67	N/A	TRUS	CC	MD	6 (3+3)
PC72-04	70	CC	TURP	CC	PD	9 (4+5)
PC12-03	62	N/A	TURP	CT	WD	4 (2+2)
PC12-07	59	N/A	TURP	CT	PD	9 (4+5)
PC12-08	73	N/A	TURP	CT	-	6 (3+3)
PC22-203	79	CC	TRUS	CT	PD	8 (4+4)
PC22-576	69	N/A	RP	CT	M/PD	7 (3+4)
PC22-637	70	CT	TRUS	CT	PD	8 (4+4)
PC72-06	62	CT	TURP	CT	W/MD	5 (3+2)

N/A: sample not available; TRUS: Transrectal ultrasound-guided biopsy: TURP: Transrectal resection of the prostate; RP: Radical prostatectomy; [1]Tumour grade obtained from pathology report; [2]Contemporary Gleason Score from FFPE tissue block chosen for macrodissection of nucleic acids and IHC; WD: well differentiated; MD: moderately differentiated; PD: poorly differentiated; W/MD: well-moderately differentiated; M/PD: moderately-poorly differentiated; -: information not present in original pathology report.

Sample Identification	Age at Diagnosis	Germline G84E Genotype	Tissue Source	Tumour G84E Genotype	Tumour Grade[1]	Contemporary Gleason Score[2]
PC72-154	63	CT	TRUS	CT	WD	4 (2+2)
PC3250-01	51	CT	RP	CT	PD	9 (4+5)

TRUS: Transrectal ultrasound-guided biopsy: RP: Radical prostatectomy; [1]Tumour grade obtained from pathology report; [2]Contemporary Gleason Score from FFPE tissue block chosen for macrodissection of nucleic acids and IHC; WD: well differentiated; PD: poorly differentiated; -: information not present in original pathology report.

5.3.5   The effect of the G84E variant on *HOXB13* gene expression

To investigate *HOXB13* gene expression, RT-qPCR was undertaken. RNA was extracted separately from adjacent malignant and benign glands for 10 cases, and due to limited tissue availability, from benign glands only and malignant glands only for four cases each and a mixed cell population for one case (Appendix 16). *HOXB13* expression was initially assessed in the 10 paired malignant-benign samples for both carriers and non-carriers. Significantly higher expression was observed in malignant compared to benign cells (1.5-fold increase; p=0.01; Figure 5.4A). However, when *HOXB13* expression was statistically compared between the malignant glands of G84E variant carriers (n=6) and non-carriers (n=8), there was no significant difference detected (p=0.21; Figure 5.4B). There was also no detectable difference in *HOXB13* gene expression between the benign glands of variant carriers (n=4) and non-carriers (n=10; p=0.29; Appendix 16).
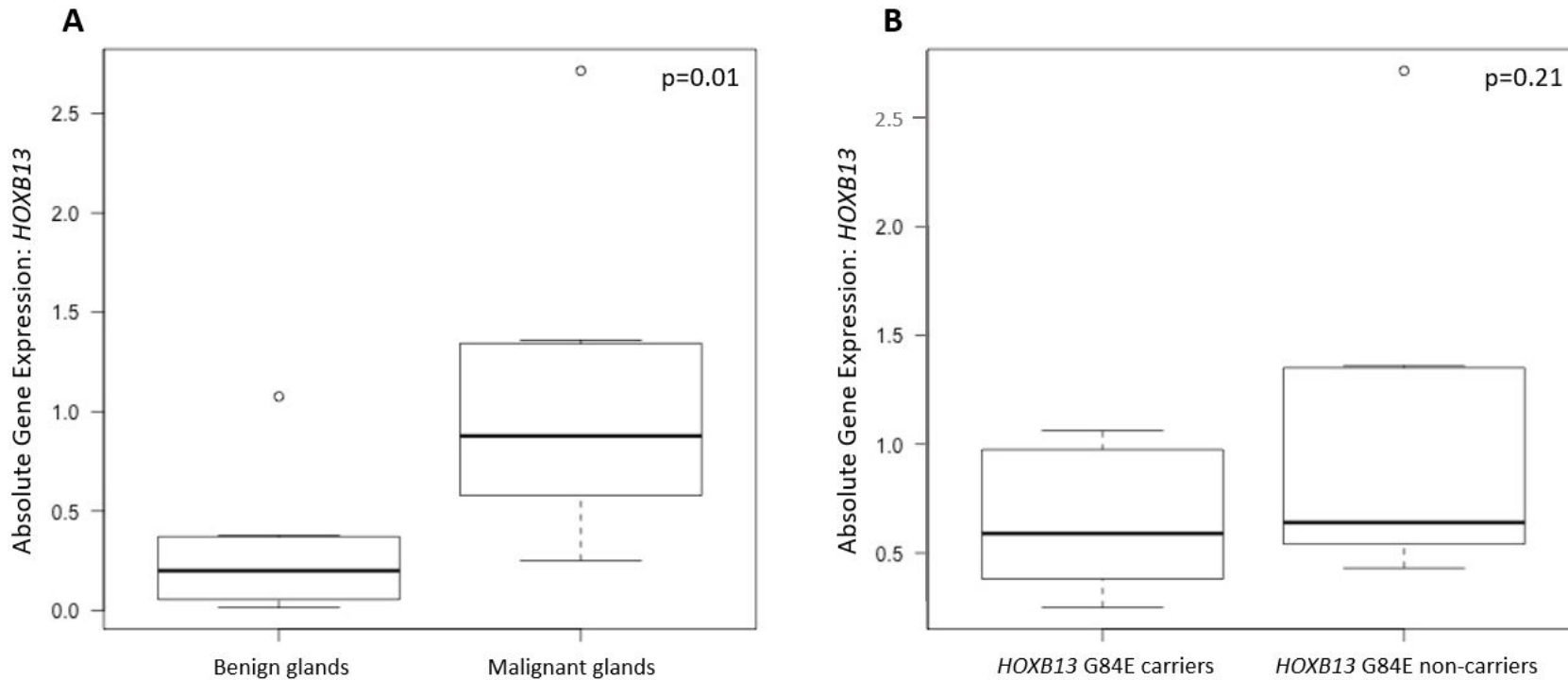
Figure 5.4 *HOXB13* gene expression analysis in malignant and benign prostate glands, and in malignant glands from G84E carriers and non-carriers.

The spread of the data is represented by a box and whisker plot. Median expression is shown by the thick black line, the interquartile range (middle 50% of data set) is represented by the box, and the minimum and maximum values by the whiskers (dotted lines). Individual outliers are shown with dots. **A)** *HOXB13* expression was assessed in prostate tumours with matched malignant and benign glands (n_{pairs}=10). Absolute *DAPK3* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. *HOXB13* expression in malignant and benign glands was compared using a paired Student's t-test. **B)** *HOXB13* expression was assessed in malignant prostate glands from G84E carriers (n=6) and non-carriers (n=8). Absolute *HOXB13* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. *HOXB13* expression in malignant glands from G84E carriers and non-carriers was compared using an unpaired Student's t-test.

We next examined whether the variant allele was detectable in the tumour tissue of seven G84E variant carriers. Next-generation sequencing applied to cDNA from freshly cut FFPE sections indicated that only two of seven variant carriers had evidence of variant allele transcription (Table 5.7). The variant allele was detectable in both malignant and benign glands in one individual (PC12-03) and in benign cells only in the second individual (PC72-06). In all cases, the variant allele was transcribed less than the wild-type thus, suggesting imbalanced allele transcription.

Table 5.7 Transcription of the G84E variant allele by *HOXB13* variant carriers.

Sample Identification	Tissue Cell Type	G84E Variant Allele Transcription[1]
PC12-03	Malignant	+
	Benign	+
PC12-07	Malignant	-
PC12-08	Malignant	-
PC22-203	Malignant	-
PC22-576	Benign/Malignant	-
PC72-06	Malignant	-
	Benign	+
PC3250-01	Malignant	-
	Benign	-
[1]Transcribes (+) or does not transcribe (-) the G84E variant allele.		

To determine whether imbalanced allele transcription was related to *HOXB13* G84E carrier status, allele-specific transcription was determined for another variant in relatively close proximity to the G84E variant. The *HOXB13* variant, rs9900627 (MAF 11.2%), is 262bp centromeric to G84E and is also located in exon 1 (Appendix 12). Genotyping of our tumour tissue samples identified one carrier of rs9900627 (PC11-11; G84E negative). Unlike carriers of G84E, the variant and wild-type alleles of rs9900627 were detectable in equal proportions in this tumour.

5.3.6 The effect of the G84E variant on HOXB13 protein expression

IHC was performed on all 22 FFPE pathology samples to determine whether protein expression differed between benign and malignant prostate tissue, and between *HOXB13* variant carriers and non-carriers. HOXB13 staining intensity ranged from weak (1) to strong (3) across the dataset, and the percentage of HOXB13 positive nuclei ranged from approximately 50-100% (Appendix 16). Analyses of the quasi-continuous nuclear scores (staining intensity x % of HOXB13 positive nuclei) of 16 samples with paired malignant and benign glands did not reveal any significant difference in HOXB13 protein expression between malignant and benign glands (p=0.45; Figure 5.5 and Figure 5.6). Analysis of malignant glands from G84E variant carriers (n=9) *versus* non-carriers (n=9) also indicated no significant difference between the two groups (p=0.68; Figure 5.6). A similar result was observed for carriers (n=8) and non-carriers (n=12) in benign glands (p=0.84).
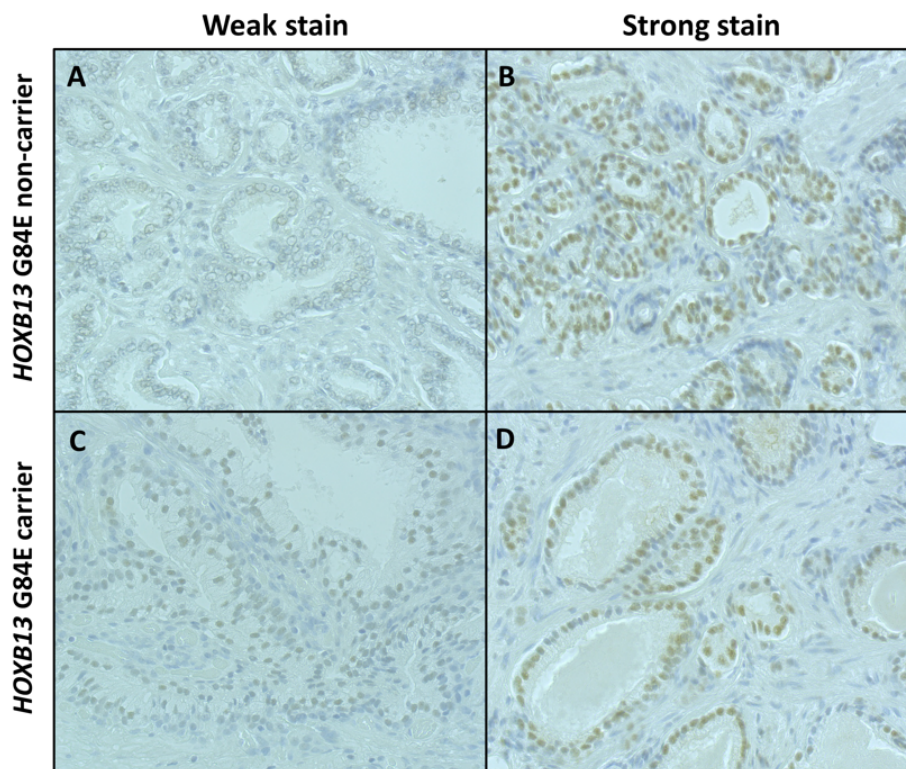
Figure 5.5 HOXB13 protein expression in FFPE prostate tumour samples.

HOXB13 protein expression was assessed in 22 prostate tumour specimens from the *Tasmanian Prostate Tissue Pathology Resource* to determine whether the G84E variant affected HOXB13 protein levels. In short, IHC using an antibody targeting amino acid 1-284 of the HOXB13 protein was utilised to assess protein expression. Staining intensity was scored as weak, moderate or strong. **A/C)** Weak staining of HOXB13 in the nucleoplasm of malignant prostate glands in a G84E non-carrier **(A)** and carrier **(C)**. **B/D)** Strong staining of HOXB13 in the nucleoplasm of malignant prostate glands in a G84E non-carrier **(B)** and carrier **(D)**. Images were taken with a Leica 2500 microscope (x200) using the Leica Application Suite V3.
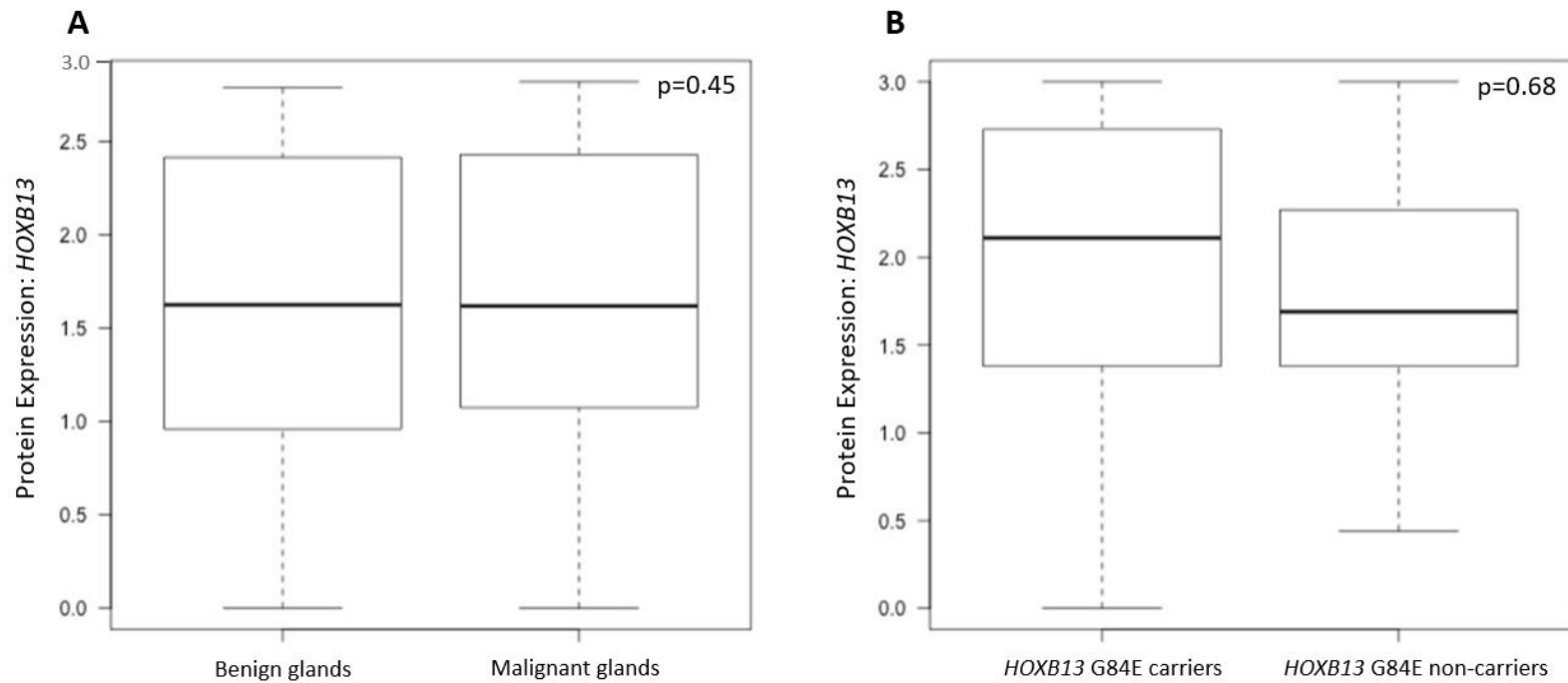
Figure 5.6 HOXB13 protein expression analysis in malignant and benign prostate glands, and in malignant glands from G84E carriers and non-carriers.

HOXB13 protein expression was calculated as a quasi-continuous score (staining intensity x % of HOXB13 positive nuclei) for both malignant and benign glands in all samples. The spread of the data is represented by a box and whisker plot. Median expression is shown by the thick black line, the interquartile range (middle 50% of data set) is represented by the box, and the minimum and maximum values by the whiskers. Individual outliers are shown with dots. **A)** HOXB13 expression was assessed in prostate tumours with matched malignant and benign glands (n_{pairs}=16). HOXB13 expression in malignant and benign glands was compared using a paired Student's t-test **B)** HOXB13 expression was assessed in malignant prostate glands from HOXB13 G84E carriers (n=9) and non-carriers (n=9). HOXB13 expression in malignant glands from G84E carriers and non-carriers was compared using an unpaired Student's t-test.

5.3.7   The effect of the G84E variant on *HOXB13* CpG island methylation

DNA methylation was investigated at two *HOXB13* CpG islands, one spanning the promoter region and exon 1 of the gene (19 CpG sites) and the other located ~4.5 kb upstream of the *HOXB13* transcription start site (22 CpG sites; Figure 5.7). Very low levels of DNA methylation was observed across both CpG islands in variant carriers (n=3) and non-carriers (n=3; Figure 5.8). Patterns of methylation differed between individuals, however there was no correlation between DNA methylation and G84E carrier status (Appendix 16).
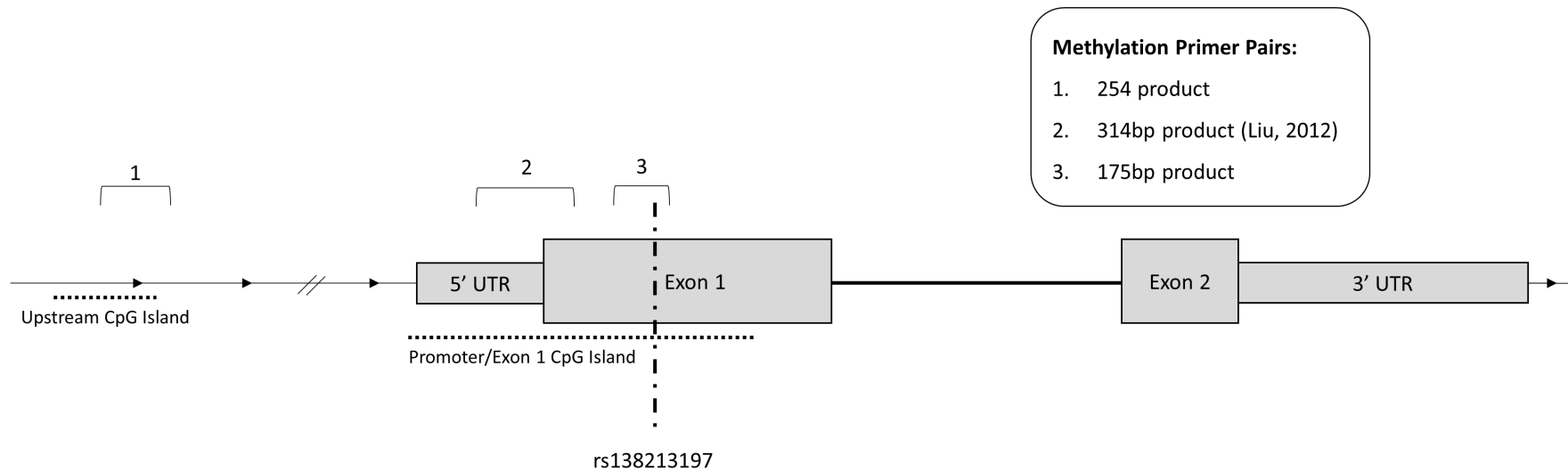
Figure 5.7 Schematic of the *HOXB13* gene indicating specific primer pairs used to analyse CpG island DNA methylation.

This figure depicts the structure of the *HOXB13* gene; the two exons are shown in large boxes, with the untranslated regions on either side. The location of the *HOXB13* G84E variant is marked with a dashed line in exon 1 (rs138213197). Two CpG islands are marked; one spans the promoter/exon 1 region and the other is located ~4.5kb upstream of the *HOXB13* transcription start site. DNA methylation was investigated in these two regions using primer pairs 1 and 2. Allele-specific methylation was examined across nine CpG sites in the promoter/exon 1 CpG island, surrounding the G84E variant (Primer pair 3). Primer sequences were designed using MethPrimer [248] and are shown in Appendix 12. Please note, this diagram is not to scale.
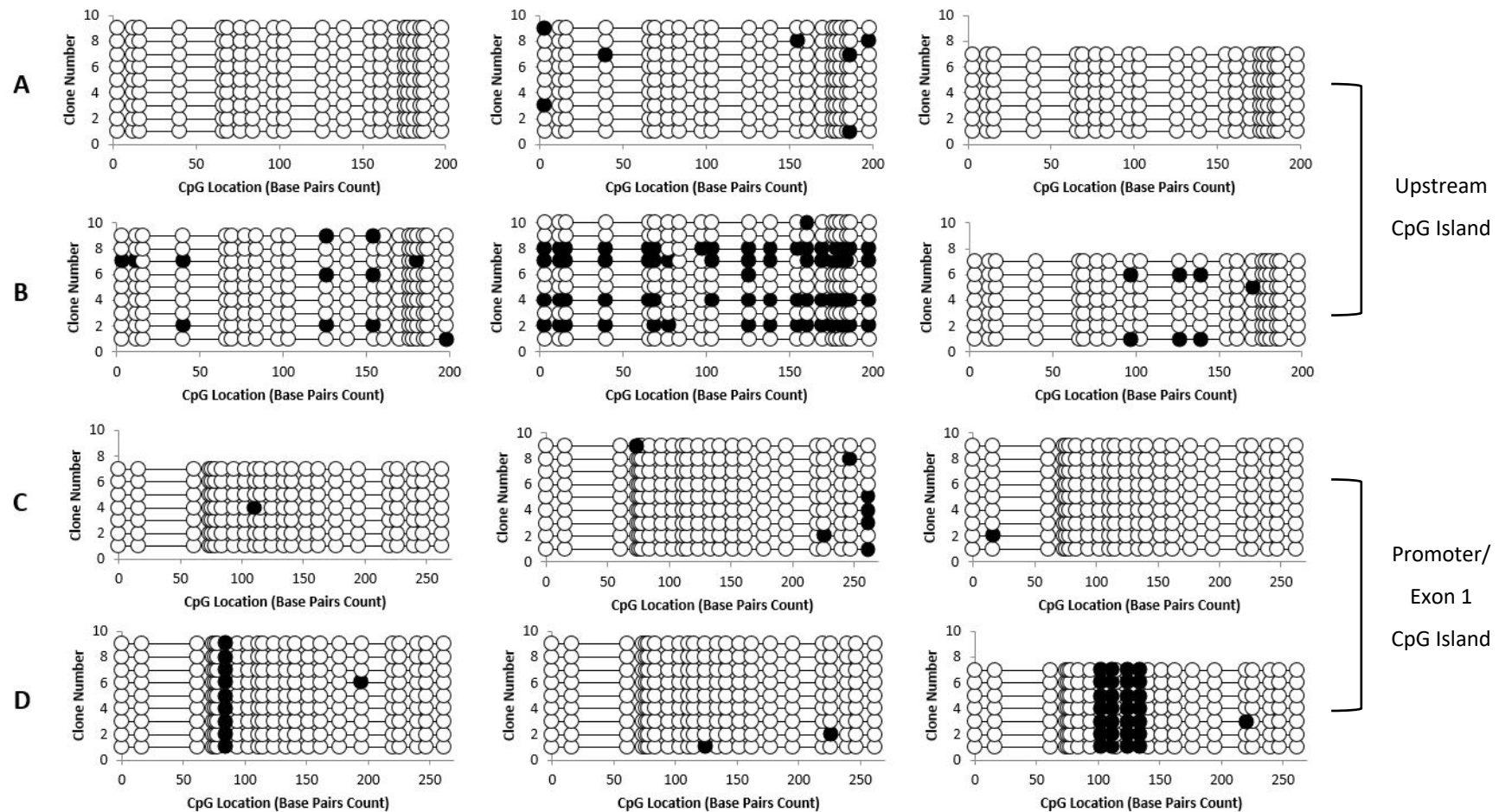
Figure 5.8 Bubble maps showing methylation patterns across the two *HOXB13* CpG islands in G84E carriers and non-carriers.

DNA methylation was investigated at two *HOXB13* CpG islands; the first ~4.5kb upstream of the *HOXB13* transcription start site **(A and B**; primer pair 1 on Figure 5.7) and the promoter/exon 1 region **(C and D**; primer pair 2 on Figure 5.7) in G84E carriers **(A and C)** and non-carriers **(B and D)**. Bubble maps were produced using CpG Bubble Chart Generator, Version 20061209 Alpha. The location of the CpG site is shown from left to right, with every sequenced clone depicted one above the other. Open circles indicate non-methylated CpG sites while coloured circles indicate methylated sites. Overall, very low levels of FFPE DNA methylation was observed, however patterns differed greatly between individuals, regardless of G84E carrier status.

Allele-specific methylation was also examined across nine CpG sites within the promoter/exon 1 CpG island (surrounding the G84E variant) to determine if differential methylation explained the observed unbalanced allele transcription. Allele-specific methylation was also consistently low across all nine CpG sites in both variant carriers (n=10) and non-carriers (n=7). However, methylation of the variant allele was lower than that of the wild-type allele in all instances (Figure 5.9). Significant differences in CpG site-specific methylation between the variant and wild-type alleles of both carriers and non-carriers was observed at three CpG sites (p<0.05), while no difference was observed between the wild-type alleles of carriers and non-carriers (Figure 5.9). No statistical correlation between methylation and transcription of the G84E variant allele, or absolute *HOXB13* gene expression was observed (Appendix 16).
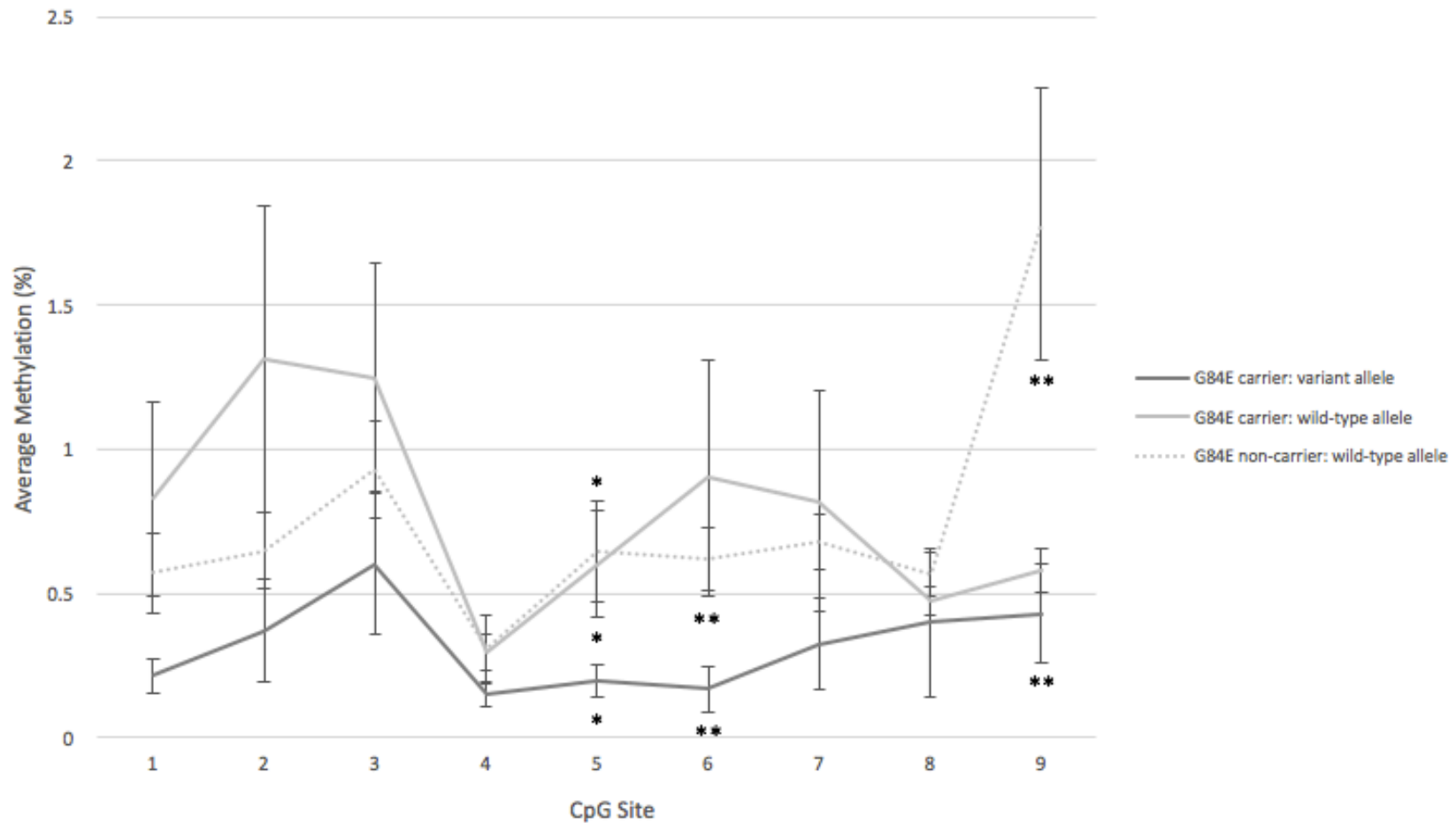
Figure 5.9 Average methylation across nine CpG sites located within the CpG island surrouding the G84E variant.

Allele-specific methylation was examined across nine CpG sites within the *HOXB13* promoter/exon 1 CpG island. CpG sites are labelled 1-9, which are left to right in Figure 5.7. Both G84E carriers (n=10) and non-carriers (n=7) were examined. Wild-type and variant allele methylation were assessed separately for G84E carriers, and in all instances the percentage of methylated reads *versus* total reads was calculated, which is shown here. Average methylation was compared between the three groups at each CpG site and those statistically significant are marked (*p=<0.05, **p=<0.01).

5.4   DISCUSSION

5.4.1   The *HOXB13* G84E variant and prostate cancer risk

The *HOXB13* G84E variant, rs138213197, was initially identified as a PCa risk variant by Ewing and colleagues (2012), and has since been replicated in several population and family-based case-control studies [11]. G84E is a missense variant in exon 1 of the *HOXB13* gene which results in a glycine to glutamic acid substitution at amino acid position 84. The variant amino acid residue is larger, less hydrophobic and negatively charged compared to the wild-type, suggesting that the variant allele may affect protein structure and/or function [168]. It has a CADD score of 22.7; predicting it to be in the top 0.1% of all damaging variants in the genome [161], is highly conserved across species and is predicted to be deleterious and probably damaging by computational algorithms, SIFT and PolyPhen2, respectively [159,160].

In the normal prostate, the highly expressed HOXB13 transcription factor plays a key role in prostate development [249]. Notably, HOXB13 has been shown to interact with the androgen receptor (AR), a protein essential for prostate development and required for all stages of PCa growth [250]. Norris and colleagues (2009) demonstrated that HOXB13 acts as both a repressor and coactivator of AR target genes [250]; in target genes with an androgen-response element the HOXB13:AR complex inhibits transcription, but in genes with a HOX element, the complex enhances transcription [250]. HOXB13 has been reported to function as a growth promoter and growth suppressor in PCa models, depending on factors such as tumour androgen sensitivity status and cellular localisation of the protein (reviewed in [251]). Therefore, the role of HOXB13 in prostate tumour development appears complex.

The *HOXB13* G84E variant was initially identified in probands from four PCa families, following targeted whole-exome sequencing (WES) of a known and replicated linkage peak [11], as described in Chapters 1.7 and 3.1. Subsequent population and family-based studies have confirmed the association of the variant with early-onset, familial disease, including an Australian study that established a relative risk of 16.4 (95% CI: 2.5-107.2) [252]. Other studies have simply shown that the G84E variant is more frequently observed among men with PCa compared to men without cancer [151,253]. This association with overall PCa risk was replicated here in our study by M_{QLS} analysis [2] (OR=6.59, p=4.22×10^{-5}). Enrichment analysis found the G84E variant to be enriched in the familial PCa case cohort compared to our population controls. When comparing familial and sporadic cases to controls, this significant finding was

diminished, which may indicate that the G84E variant is more relevant in men with a family history of disease. However, the carrier frequency of our familial cohort was not significantly different to ExAC. Overall, enrichment analysis doesn't take in to account the relatedness of individuals, whereas M_{QLS} can test for association with risk by taking into account all available relationship data.

Notably, the mechanism by which the *HOXB13* gene and, specifically, the G84E variant promotes prostate carcinogenesis, is largely unknown. Further analyses are required to determine whether the G84E variant causes a gain or loss of gene function, or increases PCa risk through other mechanisms.

5.4.2   Examining the effect of the G84E variant on *HOXB13* gene and protein expression and methylation patterns in prostate tumours

The G84E amino acid change could cause the torsion angles in the wild-type backbone to be forced in to an incorrect conformation, which could lead to disturbance of the protein structure [168]. However, a computational modeling study by Chandrasekaran and colleagues (2017) has suggested that the G84E variant increases HOXB13 protein stability [254], which may in turn cause increased transcription of downstream target genes promoting cell proliferation and invasion [255]. In an *in vitro* cell model study using site directed mutagenesis, Cardoso and colleagues (2016) found that the G84E variant had no phenotypic impact thus, proliferation and apoptotic potential was comparable to the wild-type cell model [256]. In our study of FFPE prostate tumour tissue, no difference in HOXB13 protein expression was found between G84E carriers and non-carriers; a finding supported by a larger IHC study of radical prostatectomy samples from 101 G84E carriers and 99 non-carriers [257].

Furthermore, we demonstrated that gene expression was comparable between G84E variant carriers and non-carriers. Although tumour tissue samples from carriers were demonstrated to be heterozygous for the G84E variant, the variant allele was rarely detectable in G84E carrier prostate tissue (benign or malignant glands). In fact, the variant allele was only detectable in two of seven carriers and at lower levels than the wild-type allele. To further examine *HOXB13* allelic expression, transcription of another *HOXB13* variant (rs9900627) in close proximity to G84E was examined. Comparable transcription of both the rs9900627 wild-type and variant alleles was observed in the malignant glands of a non-G84E carrier. We therefore hypothesise that the unbalanced allele transcription may be related to the presence of the G84E variant.

Unbalanced allele transcription has previously been reported in a study of breast cancer patients [258]. Benz and colleagues (2006) investigated the common *ERBB2* variant, G1170C, in *ERBB2*-positive and *ERBB2*-negative breast cancer patients and found that although tumour genotyping supported the heterozygous state, similar to our study, 70% of tumours showed preferential transcription of one allele, or unbalanced allele transcription [258]. The authors suggested that the unbalanced allele transcription in *ERBB2*-negative tumours may be due to epigenetic mechanisms, whereby methylation silences a particular allele [258].

Two CpG islands are located within or near the *HOXB13* gene; the first spans the promoter and exon 1 region of the gene and the second is ~4.5 kb upstream of the *HOXB13* transcription start site [259]. In a study of colorectal cancer, Ghoshal and colleagues (2010) found very little methylation in the promoter/exon 1 CpG island in both tumour and normal cell lines, whilst the upstream CpG island was significantly more methylated in tumour compared to normal cell lines [259]. They found that hypermethylation of the upstream CpG island partially suppressed *HOXB13* expression and speculate that this region may function as an enhancer [259]. In our study, we observed very low levels of DNA methylation at both CpG islands in all prostate tumour samples tested, thus no association with *HOXB13* expression was able to be examined. When we looked further at allele-specific methylation of nine CpG sites surrounding the G84E variant in exon 1, overall level of methylation across this region was again very low, however methylation was lower at three CpG sites on the variant allele compared to the wild-type. Overall, our sample size reduced our statistical power of finding an association between patterns of methylation and G84E carrier status. There are significant cis-expression quantitative trait loci (cis-eQTL) encompassing *HOXB13*, which explain how differentially methylated CpG sites may act as mediators between genetic variation and gene expression [260]. Even though the GTEx Portal (https://gtexportal.org/home/) has found no significant cis-eQTLs in prostate tissue [137], it is possible that other methylation differences explain the unbalanced allele transcription we observe in the G84E carriers. Alternatively, copy number variation at the *HOXB13* site or rapid targeted degradation of the variant mRNA transcript may underpin the observed allelic imbalance and warrants further investigation.

5.4.3 Association of G84E carrier status with clinical characteristics and tumour pathology

Several previous studies have investigated possible associations between the G84E variant and clinicopathological factors, and the majority have found no association between carrier status and GS [11,154,261,262]. However, two studies have presented contrary results. A Danish study of 995 cases (25 G84E carriers) found G84E carrier status was significantly associated with GS ≥7 *versus* GS <7 (p=0.032) [263]; that is the variant is associated with more aggressive disease. Another study of 1,457 cases (18 G84E carriers) observed that the G84E variant was more strongly associated with GS ≥7 (4+3) disease (OR=4.13), but this was not significantly different to the association with GS ≤7 (3+4) disease (OR=2.71) [155]. Following analysis of all Tasmanian PCa cases in the six carrier families, our study found that the G84E variant was not associated with GS, with the majority of men, irrespective of carrier status, having a GS of 6 (3+3) or 7 (3+4). Interestingly, while the numbers were too small for formal analyses, it appeared that clinical characteristics differed between G84E variant carriers who did or did not transcribe the variant allele. The tumours from PC12–03 and PC72–06, where the variant allele was transcribed, were well- to moderately-differentiated with a GS <6, whereas tumours where the variant allele was not transcribed, were predominantly poorly differentiated with a GS ≥7, with the exception of one sample (PC12–08; GS 6 (3+3)). Due to insufficient tumour material (TRUS biopsy), allele-specific transcription was not able to be determined for two variant carriers, PC22–637 and PC72–154.

In previous studies, G84E carrier status has been identified to be associated with an earlier age of disease onset. The initial study by Ewing and colleagues (2012) found that G84E carriers were more likely to be diagnosed at ≤55 years compared to non-carriers [11]. Here, the age of diagnosis between G84E variant carriers and non-carriers in the six carrier families was similar (mean of 66 *versus* 67 years, respectively). While it has to be noted that our observations are based on limited numbers, the conflicting results of the studies described above may be due to the underlying variability in G84E variant allele transcription that we have observed, and this should be explored in a larger prostate tumour dataset, consisting of more G84E variant carriers.

5.4.4   Other prioritised rare variants in cancer associated genes

This study also identified a number of other rare potential PCa risk variants (Table 5.2). The *RAD51C* A126T variant, like the *HOXB13* G84E variant was identified in PcTas72, however instead of only being present in one WGS affected case, it segregated with disease in three out of four of the WGS individuals. It has a CADD score of 21.7 and has been reported by ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) [163] to be associated with familial breast and ovarian cancer, yet it is predicted to be benign. The variant causes a substitution of a small amino acid, alanine to a nucleophilic amino acid, threonine, which may cause the protein structure to be disturbed [168]. In fact, the variant is located on the surface of the protein, which may disturb interactions with other molecules or parts of the protein [168]. The variants in *ATM* and *RNASEL* are interesting candidates too. As mentioned in Chapter 3, *ATM* is a DNA repair gene which is responsible for recognising damaged or broken DNA strands, but it also controls the rate at which cells grow and divide [116]. The highest prioritised *ATM* variant, rs5612873, was identified in two out of five PCa cases in the PcTas22 main pedigree and is predicted to be damaging to protein function, with a CADD score of 23.4. The variant results in a smaller amino acid, which may lead to loss in interactions and an inability to repair defective DNA [168]. *RNASEL* is a known PCa susceptibility gene, which has been found to be associated with disease in families with five or more affected relatives, father to son transmission, a younger age of diagnosis and a higher GS [94]. Here, a novel variant was identified in two affected men from PcTas3. It has a CADD score of 16.21 and causes the small glycine amino acid to be substituted with a hydrophobic, valine [168]. The variant is located within a stretch of residues that is repeated in the protein, which is known as an ankyrin repeat domain. Thus, the variant may disturb this repeat and consequently, its function, which is to bind to other molecules [168].

5.4.5   Limitations of this study

This study has provided important insights into the effect the *HOXB13* variant has on gene transcription in prostate tumour tissue, but there are some limitations. Due to the rarity of the variant and the limited availability of informative tumour tissue specimens, the number of samples available for G84E variant carriers was restricted. A small sample size results in reduced power and, therefore lowers the likelihood of detecting statistically significant differences between groups. For example, the sample size of our methylation assays significantly hampered our power to link carrier status with DNA methylation patterns. Thus, the concepts explored in this study should be followed-up in a larger tissue cohort of G84E carriers. The quality of DNA and RNA extracted from FFPE tissue is also fairly poor, therefore

it is important that our findings are validated in larger FFPE cohorts or, if available, fresh frozen samples. Lastly, in our IHC experiment, the antibody used was not specific to the variant form of the HOXB13 protein and it would be valuable to verify our gene expression results with a variant-specific protein antibody.

5.4.6   Possible interactions between two prostate cancer risk genes identified in our Tasmanian cohort

Previous literature suggests that the PCa risk genes identified in this study, *EZH2* and *HOXB13*, may interact. In a study of 148 non-small cell lung cancer, HOXB13 was found to upregulate EZH2 expression, via binding directly to the *EZH2* promoter [264]. Liu and colleagues (2012) also observed that *EZH2* represses *HOXB13* expression through recruitment of DNMT3b to the *HOXB13* promoter [265]. In fact, Xiong *et al.* (2018) showed that the overexpression of a long noncoding RNA, HOXB13-AS1, increased DNMT3b-mediated methylation of the *HOXB13* gene promoter by binding to *EZH2*, epigenetically suppressing *HOXB13* expression [266]. Therefore, given that we had DNA methylation data from the *HOXB13* promoter region, as well as ~4.5 kb upstream of the transcription start site, we assessed whether differential methylation of *HOXB13* was present between *EZH2* carriers and non-carriers. Two *EZH2* variant carriers (PC12-03 and PC12-09) and three non-carriers (PC4-03, PC11-11 and PC12-07) were examined and although our analyses lacked statistical power, Figure 5.6 shows low methylation across the *HOXB13* region in all tumour DNA samples (some were completely unmethylated). Overall, there appeared to be no difference in *HOXB13* methylation between *EZH2* carriers and non-carriers.

5.5   FUTURE DIRECTIONS

This study has provided insight into the effect of the *HOXB13* G84E variant on HOXB13 expression at the transcriptional and translational level, however it is still unclear how the mutation functionally leads to increased cancer risk. It is possible that the G84E variant affects the developing prostate during embryonic development when *HOXB13* expression levels are very high. Future studies should investigate the variants effect on the developing prostate, as well as the pathways that may be affected by this variant. Chandrasekaran and colleagues (2017) suggested that the G84E variant may cause increased transcription of downstream target genes, such as *MEIS, AR* and *FOXA1* and FOXA2, therefore it is possible that these interactions are affected by the presence of the variant. Given that *MEIS* expression has been implicated in

collaboration with *HOX* genes in the development of leukaemia, these interactions are an ideal area for future research, as variants in *HOXB13* could affect the function of *MEIS* itself, or its target genes [267]. *HOXB13* interactions with *FOXA1* and *FOXA2* are also interesting. *FOXA1* is enriched at tumour-specific AR binding regions, just like *HOXB13,* and *FOXA2* may have the ability to bind to the AR enhancer and regulate *HOXB13* expression [268,269]. To determine whether the G84E variant has an effect on these interactions, gene expression assays and pathway analysis could be performed in an *in vitro* setting, using cell lines with and without the variant. Overall, through collaboration with members of the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium, we aim to further explore the function of this variant.

This study also identified a number of other rare PCa risk variants, however further follow-up studies are required. Table 5.2 details an additional five variants in *ATM, RAD51C* and *RNASEL* that were prioritised, however given the replicated association of the *HOXB13* G84E variant this was the only one followed-up. Appendix 15 details an additional eight novel variants and 12 rare variants in other DNA repair genes and PCa associated genes that have not been further assessed at this stage. Rare variants in *ATM, BRAC1, BRCA2* and *RNASEL* that have been previously associated with breast and PCa (*RNASEL* A132V) were identified here (Appendix 15), therefore follow-up analysis is required to characterise their contribution to the Tasmanian population. Overall, validation, segregation and association analysis of the *RAD51C* variant is currently underway, however it is also possible to follow up the other identified rare/novel variants using the study design described in this thesis.

5.6   CONCLUSION

In conclusion, this study has found that the well-known *HOXB13* G84E variant also contributes to PCa risk in the Tasmanian population (OR=6.59, p=4.22×10^{-5}). Functional assessment of the effect this variant has on gene and protein expression provided some insight into the expression levels of *HOXB13* in malignant and benign prostate glands of G84E carriers and non-carriers, however questions still remain regarding how this variant promotes cancer development. This is because the sample size presented here is too small to make definitive conclusions about the functional consequence of this variant. Therefore, it is important that our findings are validated in larger FFPE cohorts or, if available, fresh frozen samples. Overall,

this chapter has proven that a targeted approach to rare variant prioritisation can aid in the timely identification of pathogenic variants in previously identified candidate genes.

CHAPTER 6 : CHROMOSOMAL ABERRATIONS IN TASMANIAN PROSTATE TUMOURS

6.1   INTRODUCTION

Prostate tumours are extremely heterogenous at the molecular, genetic and phenotypic level [270]. Tumour heterogeneity is the phenomenon of individual tumour foci, and even individual cells, presenting distinct characteristics [271]. Despite tremendous progress over the last decade, we still lack understanding of the extent and effect of intra-tumour heterogeneity, particularly in prostate tumours. This makes the diagnosis and treatment of prostate cancer (PCa) difficult and can result in poor outcomes for the patient.

Mapping chromosomal aberrations has provided insight into the genetic makeup of a range of tumours and, in some instances, formed the basis of cancer classification systems used to stratify patients and determine their treatment option, for example in haematological malignancies [272,273]. Chromosomal aberrations include deletions, amplifications, inversions and translocations. Chromosomal deletions, inversions and translocations can result in the fusion of two separate genes and this phenomenon will be discussed in Chapter 7. In this chapter, deletions and amplifications resulting in DNA copy number variations (CNVs) will be examined. CNVs are frequent in PCa tumours of high grade and advanced stage [274,275] and have previously been identified in the clinic by traditional chromosomal karyotyping in blood samples. However, in research, comparative genomic hybridisation (CGH) is commonly used, which is far more advantageous, as DNA can be obtained from cell lines, and fresh, frozen or formalin-fixed paraffin embedded (FFPE) tumour tissue.

CGH is a molecular cytogenetic method developed by Kallioniemi and colleagues (1992), which examines a tumour genome for DNA sequence CNVs [276]. In the late 1990's, array-based CGH (aCGH) superseded CGH, due to its increased resolution (5-10Mb to 1.4Mb) [277,278]. It has provided the flexibility to gain a genome-wide view of abnormalities, but also provides the opportunity to target specific regions of the genome to gain an in-depth picture of CNVs. aCGH also significantly improves the detection of genomic aberrations in cancer cells compared to previously established whole-genome methodologies [279].

To date, most CGH studies have been performed using sporadic tumours, and have confirmed loci previously identified by traditional methods. Consistent regions of gain may potentially harbour causative proto-oncogenes, whereas regions of loss could identify tumour suppressor genes [276,280]. Thus far, almost all of the chromosomes have been found to be gained or lost in sporadic prostate tumours. Overall, the most frequently altered chromosomes include 6, 7, 8, 10, 13, 16, 17 and X [281]. The 8p chromosomal region is the most commonly deleted region in the prostate tumour genome, affecting about a third of all tumours and half of advanced tumours [282]. This alteration was first described by Matsuyama *et al.* (1994) in a study of primary and metastatic deposits of PCa [283]. Since then, many groups have used a variety of methods to fine-map this loss to 8p22 (but not exclusively) [284,285]. The long arm of chromosome 8 is also frequently gained in PCa. In fact, it is the most commonly gained region, affecting about a quarter of all tumours and half of advanced tumours [282]. The 8q region of gain was shown to harbour the *c-MYC* gene at 8q24 by Jenkins and colleagues (1997), and was one of the first chromosomal regions to be linked to a causal gene [286]. However, as the amplification event on 8q is quite large, this suggests that many genes may be affected. Another locus identified by CGH that has led to the identification of a candidate gene is the deletion of 10q23, which harbours the candidate tumour suppressor gene, *PTEN* [287,288]. The *PTEN* deletion is now considered a likely useful biomarker for the diagnosis of lethal PCa [289,290]. The 16q region is also frequently deleted in PCa and following fine-mapping by Sun and colleagues (2005) resulted in the identified of *ATBF1*. It is thought that the loss of *ATBF1* is one mechanism that defines the absence of growth control in PCa [291]. Despite these discoveries, there are very few additional examples of observed chromosomal aberrations in sporadic prostate tumours where the underlying casual gene has been identified.

To date, there are only two published studies that have assessed chromosomal aberrations in familial prostate tumours. Verhagen *et al.* (2000) undertook the first CGH study of high risk PCa families, which included six familial cases with sufficient prostate tumour tissue [292]. This study also included seven sporadic tumours (defined as no linkage to 1q24-25 or Xq27-28). Loss of 7q and 10q, and gain of 8q were consistently identified aberrations in both familial and sporadic tumours. Distinctive abnormalities observed in familial tumours only, included loss of 3p12-3p22 in five tumours (83%) and gain of 6q11-6q21 in four tumours (67%) [292]. A later CGH analysis of 21 prostate tumours from 19 Finnish PCa families identified common losses at 13q14-13q22 (29%), 8p12-pter (24%) and 6q13-6q16 (14%), and gains at 19p (25%), 19q (14%) and 7q (14%) [293]. Overall, there are many consistently altered regions that have been

identified in both sporadic and familial prostate tumours. Notably, two chromosomal regions, 16q and 18q are consistently lost in sporadic prostate tumours, which was supported by Verhagen and colleagues (2000) [292]. These two regions were not identified in any of the familial tumours included in the familial CGH study by Rokman *et al.* (2001) [293], suggesting that familial PCa tumours harbour some unique genetic changes compared with sporadic prostate tumours. This may indicate that underlying genetic predisposition may cause familial tumours to acquire different CNVs compared to sporadic tumours. Overall, the completion of further familial PCa CGH studies will aid in the identification of somatic tumour alterations and the possible link between these and PCa predisposition variants.

Previously our group has used the Spectral Chip™ 2600 BAC array to highlight regions of loss and gain in one of our familial PCa families, PcTas9 (Table 6.1) [294]. Since this work, next-generation platforms have emerged as a useful tool for the identification of chromosomal abnormalities. These CGH platforms provide far greater resolution and can be used with far more confidence on FFPE tumour DNA. This chapter aims to expand on our preliminary findings through the analysis of Agilent Oligonucleotide aCGH data from a larger collection of PcTas9 tumour samples. We aim to identify consistent regions of loss and gain in these tumours that may be caused by underlying inherited germline variants.

Table 6.1 Regions of chromosomal loss and gain previously identified in PcTas9 prostate tumour samples.

Loss	Gain
1p22-1p31.1	1p36.21-1p36.22
1q23.3-1q25.2	6p22.1-6p22.3
6p25.1-6p25.3	6p24-6p25
6q22-6q22.1	6q25.3-6q27
7p21-7p21.3	17p13-17p13.3
10q26.2	20p12-20p12.2
17p13-17p13.3	
19p13.3	

6.2   METHODS

6.2.1   Array-Based Comparative Genomic Hybridisation

Twelve PcTas9 tumours were assayed on a customised SurePrint G3 Human 8 x 60K Microarray (Agilent Technologies), designed by Dr Liesel FitzGerald (Menzies Institute for Medical Research (AUS). Regions of loss and gain previously identified in tumours from PcTas9 were targeted for fine-mapping (Table 6.1; Appendix 17). In addition to this, the array also assays the entire genome, providing genome-wide data for each tumour. The aCGH procedure and analysis was carried out by the Molecular Anatomical Pathology laboratory at PathWest, according to the manufacturer's instructions for FFPE tissue samples (report in Appendix 18). The reference sample used for this analysis was a female, therefore to pass quality control (QC) tumour samples had to show loss of chromosome X and gain of chromosome Y. Data were visualised in Cytogenomics 5.0.2.5 (Agilent) and analysed for CNVs using the Default Analysis Method CGHv2. Regions with a log ratio of > 0.3 (gain) or < -0.3 (loss), regardless of the number of probes, were considered chromosomal aberrations.

6.2.2   Quantification of *EEF2* and *DAPK3* gene expression

EEF2 (ENST00000309011.6) and *DAPK3* (ENST00000301264.3) gene expression in FFPE prostate tumour samples was assessed by RT-qPCR analysis (Appendix 3). Expression was normalised to the expression of two housekeeping genes, as discussed in Chapter 2.3. RT-qPCR primers were designed to the most commonly transcribed isoform in the prostate (as per GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2; https://gtexportal.org/home/)) [137] and are displayed in Appendix 3. Absolute gene expression was compared between tumours from PcTas9 and non-PcTas9 cases. The non-PcTas9 patient group comprised DVA sporadic tumours from the *Tasmanian Prostate Cancer Case-Control Study* and tumours from other *Tasmanian Familial Prostate Cancer Cohort* families.

6.2.3   Statistical analysis of absolute *EEF2* gene expression

Random intercepts models were used to estimate and compare mean levels of absolute *EEF2* gene expression in the PCa families. A modified Bonferroni procedure was used to prevent the family-wise error rate rising above the pre-specified alpha of 0.05. With families ranked in terms of descending mean levels of absolute *EEF2* gene expression, a binary (0/1) covariate for the family with the highest mean was included in the model. If the Wald test of its estimated coefficient yielded a p-value less than 0.05, a binary (0/1) covariate for the family with the

second-highest mean was included in the model. If the Wald test of its estimated coefficient yielded a p-value less than 0.05/2, a binary (0/1) covariate for the family with the third-highest mean was included in the model and tested at the 0.05/3 significance level. This sequential process was terminated when the null hypothesis was accepted at any step. This analysis was performed under the guidance of biostatistician, Professor Leigh Blizzard, Menzies Institute for Medical Research (AUS).

6.2.4   Quantification of EEF2 protein expression

EEF2 protein expression in FFPE prostate tumours was assessed by immunohistochemistry (IHC), as discussed in Chapter 2.4 (Appendix 5). Cytospins of HEK293 cells, and sections of colon and skin were used as positive controls. Negative controls included primary antibody only, secondary antibody only, and a mouse IgG$_1$ isotype control (Dako). EEF2 protein expression was compared between tumours from PcTas9 and non-PcTas9 cases.

6.3   RESULTS

6.3.1   Targeted collection of prostate tumour samples from PcTas9 men for array comparative genomic hybridisation analysis

Targeted collection of prostate tissue specimens from local pathology laboratories was undertaken for affected men in the Tasmanian family, PcTas9 (Figure 6.1). In total, 26 FFPE samples from PcTas9 PCa cases were obtained (Figure 6.2). In addition, tissue specimens from 27 familial cases from 14 additional Tasmanian PCa families, and 15 sporadic cases were available for this study and together, these 42 FFPE specimens comprised the non-PcTas9 patient group.

To investigate the prevalence of chromosomal aberrations in PcTas9, 12 samples from across the pedigree, with sufficient high quality tumour DNA were assayed by aCGH (Figure 6.2; Table 6.2).
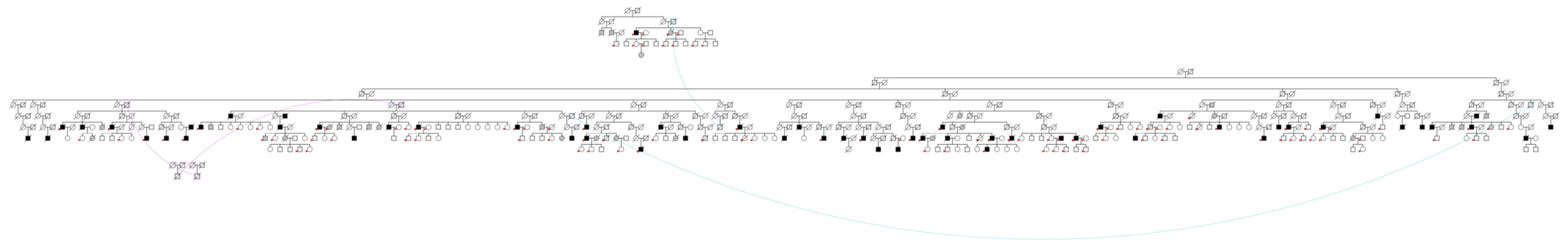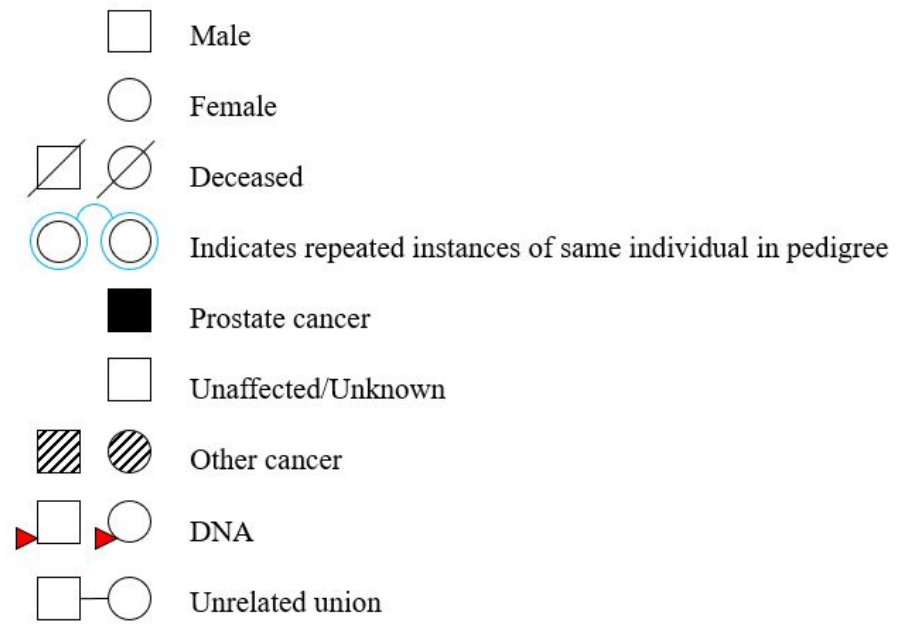
Figure 6.1 PcTas9 Pedigree.

PcTas9 pedigree, depicting the number and relationships of PCa cases (shown in shaded squares), as well the availability of DNA from cases and their unaffected relatives, which is represented by red arrows. The disease status for earlier generations is generally unknown, unless this information was obtained from clinical records. And if so, these individuals have been marked as affected in the pedigrees. This pedigree is included to illustrate the size of the pedigree only, please refer to Figure 6.2 for individual annotations.
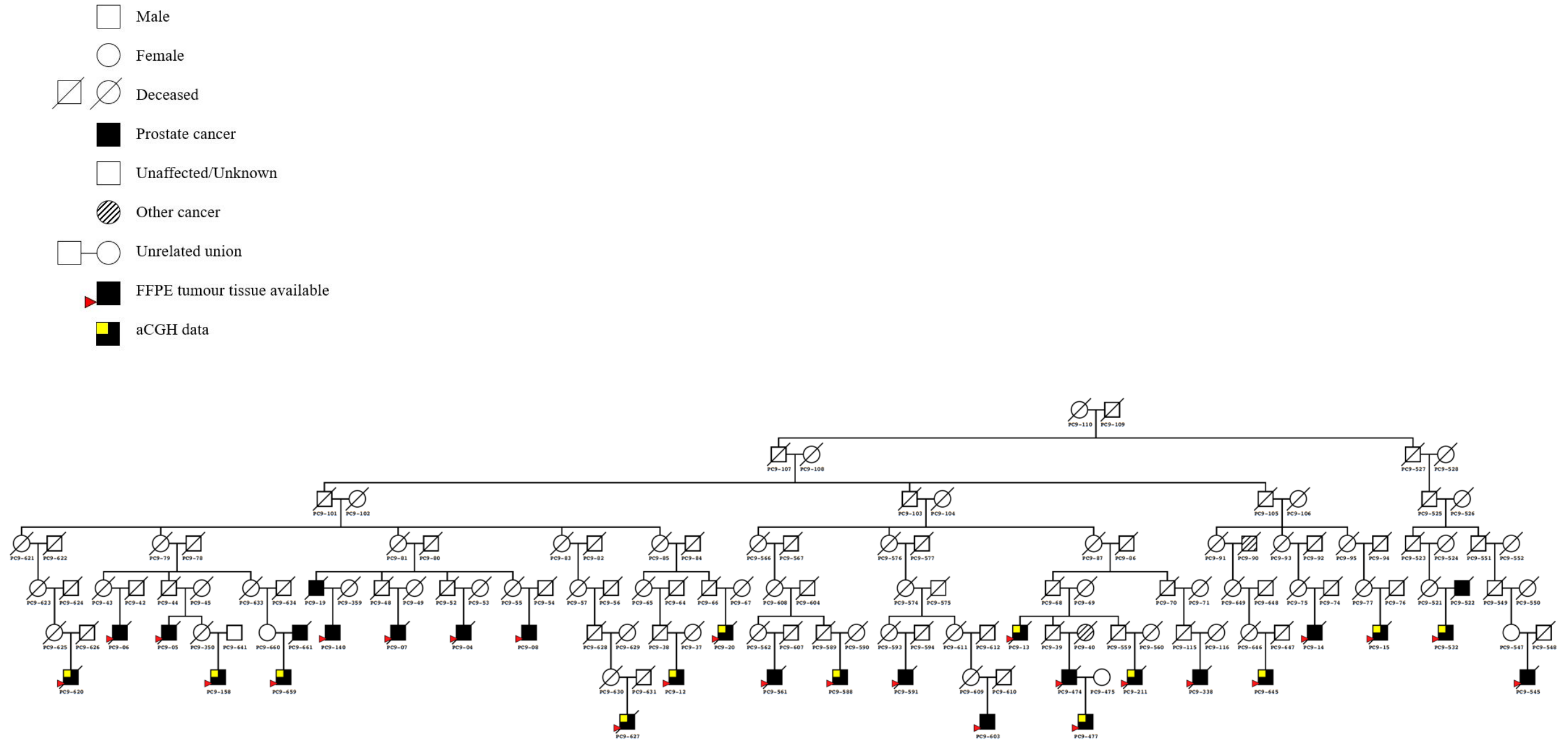
Figure 6.2 A condensed PcTas9 pedigree showing tumour samples chosen for array Comparative Genomic Hybridisation analysis.

This condensed version of PcTas9 indicates those PCa cases with available prostate tumour specimens (shown by red arrows) and their relationship. Tumours chosen for aCGH analysis are shown in yellow.

Table 6.2 PcTas9 tumour samples chosen for array Comparative Genomic Hybridisation, including clinicopathological characteristics.

Sample Identification	Age at Diagnosis	Tissue Source	Tumour Grade[1]	Contemporary Gleason Score[2]
PC9-12	66	RP	MD	6 (3+3)
PC9-13	83	TURP	-	9 (4+5)
PC9-20	76	TURP	PD	9 (4+5)
PC9-158	63	RP	-	6 (3+3)
PC9-211	68	TURP	PD	9 (4+5)
PC9-477	55	RP	-	6 (3+3)
PC9-532	70	RP	-	6 (3+3)
PC9-588	63	RP	MD	6 (3+3)
PC9-620	71	RP	PD	9 (4+5)
PC9-627	65	RP	-	7 (3+4)
PC9-645	60	RP	-	7 (3+4)
PC9-659	65	RP	PD	9 (4+5)

RP: Radical prostatectomy; TURP: Transurethral resection of the prostate; [1]Tumour grade obtained from pathology report; [2]Contemporary Gleason Score from FFPE tissue block chosen for macrodissection of nucleic acids and IHC; MD: moderately differentiated; PD: poorly differentiated; -: information not present in original pathology report.

6.3.2 Quality assessment of array data

The 12 FFPE DNA samples were assayed across two separate arrays, with four replicates for QC (Table 6.3). The Derivative Log Ratio (DLR) spread was considered the most important QC metric, which calculates the probe to probe log ratio noise of an array. A DLR spread of >0.3 is defined by Agilent as poor, however as FFPE samples normally lie within 0.3 and 0.6, a DLR spread threshold of ≤0.6 was considered acceptable, but only if the sex chromosome patterns were as expected (loss of chromosome X and gain of Y due to a female reference sample; Table 6.3; report in Appendix 18). In total, three tumours, including PC9-13, PC9-211 and PC9-659 failed QC as they had gain of chromosome Y, but no loss of chromosome X (Appendix 19). PC9-158 also failed QC as the DLR spread of both replicates was >0.6. Overall, eight of the 12 tumours passed QC thresholds (Appendix 20 and 21).

Table 6.3 Derivative Log Ratio Spread and quality assessment of the assayed PcTas9 tumour samples.

Sample Identification	Array	DLR Spread[1]	Loss of chromosome X^2	Quality Control
PC9-12	2	0.52	Yes	Pass
PC9-13	1 & 2	0.64 & 0.52	No & No	Fail & Fail
PC9-20	2	0.48	Yes	Pass
PC9-158	1 & 2	0.65 & 0.62	Yes & Yes	Fail & Fail
PC9-211	1 & 2	0.66 & 0.58	No & No	Fail & Fail
PC9-477	1	0.42	Yes	Pass
PC9-532	1	0.47	Yes	Pass
PC9-588	1 & 2	0.57 & 0.51	Yes & Yes	Pass & Pass
PC9-620	1	0.40	Yes	Pass
PC9-627	2	0.44	Yes	Pass
PC9-645	1	0.52	Yes	Pass
PC9-659	2	0.60	No	Fail
DLR: Derivative Log Ratio; [1]Tumour DNA passed quality control if the DLR spread was ≤0.6 and there was loss of chromosome X^2.				

6.3.3 The identification of chromosomal aberrations

Seven tumours from those that passed QC (n=8) showed regions of chromosomal loss and all eight tumours had regions of gain (Table 6.4; Appendix 22). Overall, four tumours were shown to harbour four or more chromosomal losses, with only PC9-620 having more than 10. Seven tumour samples showed gain at four or more chromosomal regions with one tumour, PC9-645 shown to harbour more than 10 amplifications. Chromosomal aberrations were considered to be consistent across PcTas9 tumours if they were identified in three or more tumours. The most consistent losses observed in PcTas9 tumours were found at chromosomal regions 1p36.21 and 19p13.3 (Table 6.5). The most consistent regions of gain were at 6p23-p22.3, 6p24.2, 17p13.3 and 19p13.3 (Table 6.5). The 19p13.3 region was amplified across three separate genes, including *PTPRS* (38%), *ZBTB7A* (50%) and, notably, *EEF2* (*Eukaryotic Translation Elongation Factor 2*) was gained in all eight tumours (Figure 6.3).

Table 6.4 Chromosomal aberrations identified by array Comparative Genomic Hybridisation analysis of prostate tumour samples from PcTas9 cases.

Sample Identification	Chromosomal Losses	Chromosomal Gains
PC9-12	17p13.2	3q11.1-q26.32, 19p13.3
PC9-20	6p23-p22.3, 8p23.3-p11.21, 16q22.1-q24.3, 17q25.1-q25.3, 18q21.32-q23, 21q22.11-q22.3	1p34.3-p13.2, 3q12.3-q29, 7p21.2, 17p13.3, 19p13.3, 20p12.3-p11.21, 20p12.2
PC9-477	1p36.21, 6p24.3-p24.2, 6p23-p22.3, 17p13.3, 19p13.3, 19p13.3-p11	6p24.2, 7p21.1, 17p13.3, 19p13.3
PC9-532	1p36.21, 6p24.3-p24.2, 6p23-p22.3, 19p13.3	3q13.11-q25.32, 5q11.2-q12.1, 6p24.2, 6p23-p22.3, 7p21.3, 17p13.3, 19p13.3
PC9-588[#]	None	6q22.31-q26*, 7p22.1-p15.3*, 8q12.1-q24.3*, 10q25.2-q26.2, 11p15.1-p13*, 20p12.3-p11.1
PC9-620	1p36.21, 2p13.1-p11.1, 6p24.3-p24.2, 6p23-p22.3, 6q12-q21, 10p15.1-p11.21, 13q14.12-q34, 16q22.2-q24.1, 17p13.3, 17p13.2, 19p13.3	1p36.22, 6p25.3, 6p24.2, 7p22.3-p11.2, 7q21.11-q22.1,17p13.3, 19p13.3
PC9-627	6p23-p22.3, 19p13.2-p12	6p24.2, 10q26.2, 17p13.3, 19p13.3
PC9-645	17p13.3	3q13.31-q26.2, 4q12-q35.2, 6q12-q26, 7p22.3-p11.2, 7q11.21-q36.3, 10q25.1-q26.2, 11q12.1-q24.1, 17p13.3-p13.2, 19p13.3, 20p12.3-11.1
[#]Duplicated samples on both arrays; *Chromosomal aberration not identified on both arrays. Please Note: PC9-13, PC9-158, PC9-211 and PC9-659 did not pass QC.		

Table 6.5 Recurrent chromosomal aberrations identified in PcTas9 prostate tumours.

Loss or Gain	Chromosomal Region	Frequency of CNV in PcTas9 tumours	Tumours with CNV	Known association with cancer	Interesting genes underlying the region of alteration*
Loss	1p36.21	38% (3/8)	PC9-477, 532, 620	Ovarian cancer [295]	*PRAME* and *HNRNPCL* gene families
Loss	19p13.3	38% (3/8)	PC9-477, 532, 620	Prostate cancer [296]	*TINCR* (lncRNA00036)
Gain	6p23-p22.3	63% (5/8)	PC9-20, 477, 532, 620, 627	Bladder cancer [297] and retinoblastoma tumours [298]	*JARID2*
Gain	6p24.2	50% (4/8)	PC9-477, 532, 620, 627	No known association	*NEDD9*
Gain	17p13.3	63% (5/8)	PC9-20, 477, 532, 620, 627	Prostate cancer [296]	*DPH1*
Gain	17p13.3	25% (2/8)	PC9-477, 620	Prostate cancer [296]	*MNT*
Gain	17p13.3	25% (2/8)	PC9-532, 620	Prostate cancer [296]	*SMG6*
Gain	19p13.3	38% (3/8)	PC9-477, 532, 620	Prostate cancer [296]	*PTPRS*
Gain	19p13.3	50% (4/8)	PC9-20, 532, 620, 627	Prostate cancer [296]	*ZBTB7A*
Gain	19p13.3	100% (8/8)	PC9-12, 20, 477, 532, 588, 620, 627, 645	Prostate cancer [296]	*EEF2*
CNV: copy number variation; *Description of gene function and involvement in disease is further discussed in Appendix 23.					

A) Genome View (AMP: 17, GAIN: 0, LOSS: 0, DEL: 7, LOH: 0)
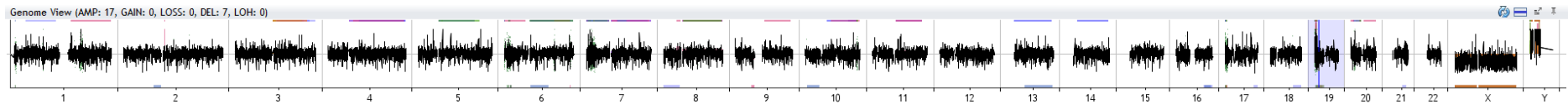
B) Gene View (chr19:3965696-3993066, Size=27.3 Kb)
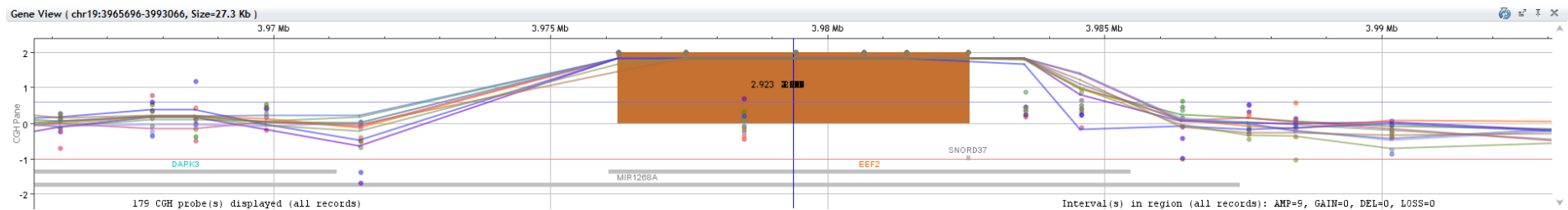
Figure 6.3 Visual representation of the recurrent 19p13.3 amplification identified by array comparative genomic hybridisation of tumour samples from PcTas9 cases.

A) Schematic of all amplifications and deletions across the entire genome in all samples combined; this was visualised using BlueFuse Multi Software (Illumina). Chromosomes are represented left to right, with the amplitude of loss and gain on the y axis. Regions considered to be significantly lost or gained are illustrated by the coloured lines above the chromosomal region. The q arm of chromosome 19 is amplified in all tumours as indicated by blue. **B)** A close-up view of 19p encompassing the *DAPK3* and *EEF2* genes (labelled). Each probe on the array is represented by a colour dot, and the coloured lines represent individual samples. The y axis is the CGH pane, with any alteration above or below 0 considered be an amplification or deletion, respectively. An amplification encompassing the beginning of *EEF2* all the way to the region upstream of *DAPK3* was evident in all eight samples.

6.3.4 Assessment of the chromosomal gain at 19p13.3 by gene expression analysis

The 19p13.3 chromosomal region was the most commonly altered region in the PcTas9 tumours. Three genes in this region were amplified, including *PTPRS* in 38% of tumours, *ZBTB7A* in 50% of tumours and *EEF2* in 100% of tumours. *EEF2* has previously been shown to be overexpressed in prostate tumours and is in a pathway that has recently been suggested as a therapeutic target for cancer [299]. To further investigate this amplification, gene expression analysis using RT-qPCR was undertaken. RNA was extracted from adjacent benign and malignant glands for 19 cases, and where limited tissue was available, in only tumour glands for 21 cases. These 40 tumours were from PcTas9 (n=17) and non-PcTas9 familial cases (n=16), and DVA sporadic PCa cases (n=7). *EEF2* expression was analysed in five regions across the gene, including 5'UTR/exon 2, exon 2/3, 4/5, 9/10 and 14/15 (Appendix 24). Significantly higher expression was observed in malignant compared to benign glands (n_{pairs}=19) in the regions of exon 2/3 (p=0.003), 4/5 (p=0.04) and 9/10 (p=0.004; Figure 6.4).
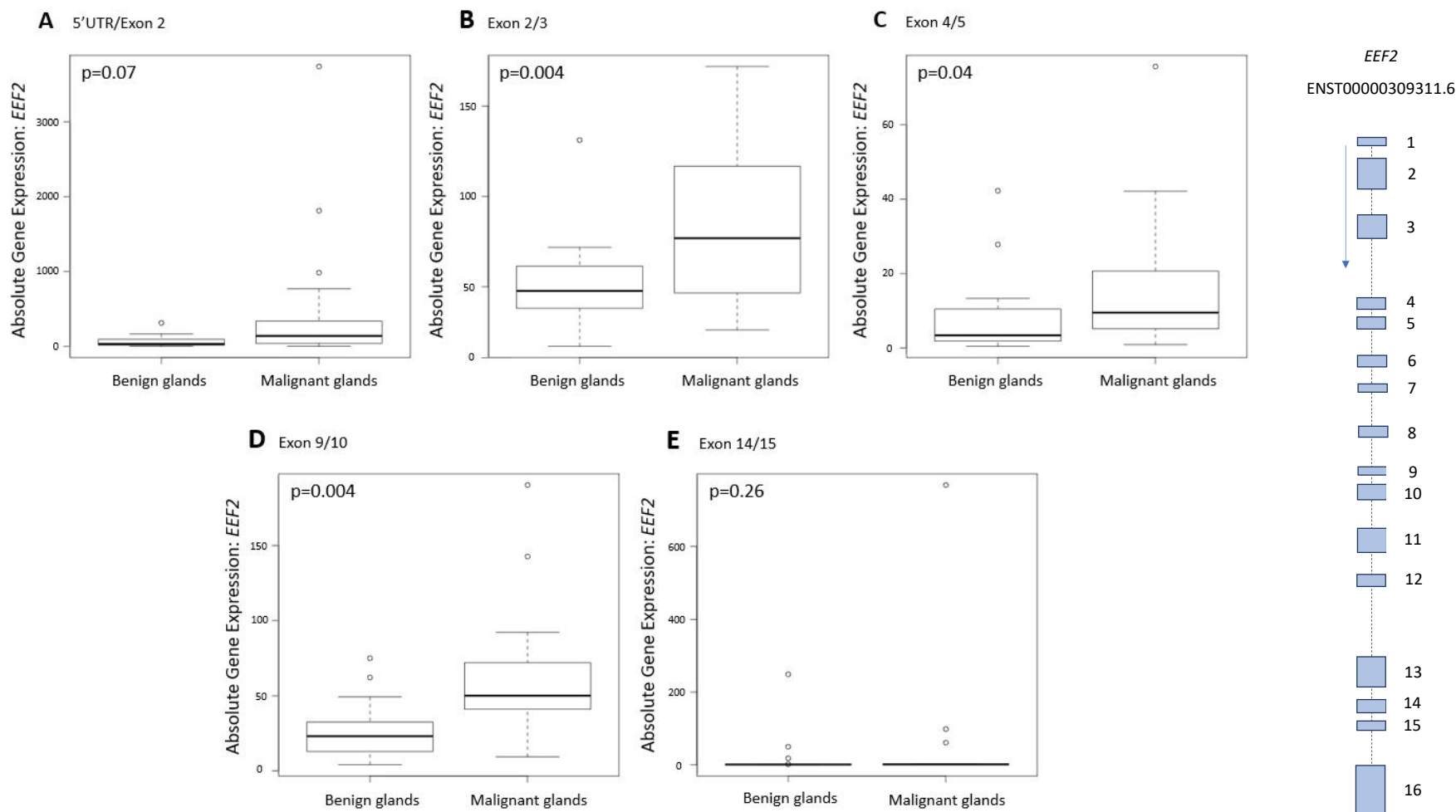
Figure 6.4 *EEF2* gene expression analysis in malignant and benign prostate glands.

EEF2 expression in five different regions of the gene was assessed in prostate tumours with matched malignant and benign glands (n_{pairs}=19). A schematic of the most commonly transcribed isoform of *EEF2* in the prostate is shown to the right. Absolute *EEF2* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. *EEF2* expression in malignant and benign glands was compared in each region using a paired Student's t-test. The spread of the data is represented by a box and whisker plot. Median expression is shown by the thick black line, the interquartile range (middle 50% of data set) is represented by the box, and the minimum and maximum values by the whiskers (dotted lines). Individual outliers are shown with dots.

When *EEF2* expression in malignant glands was compared across the patient groups, the regions of *EEF2* 5'UTR/exon 2 and exon 4/5 were expressed at a significantly higher level in PcTas9 tumours (n=18) compared to tumours from non-PcTas9 cases (n=23; p=0.02 and p=0.01, respectively; Figure 6.5).
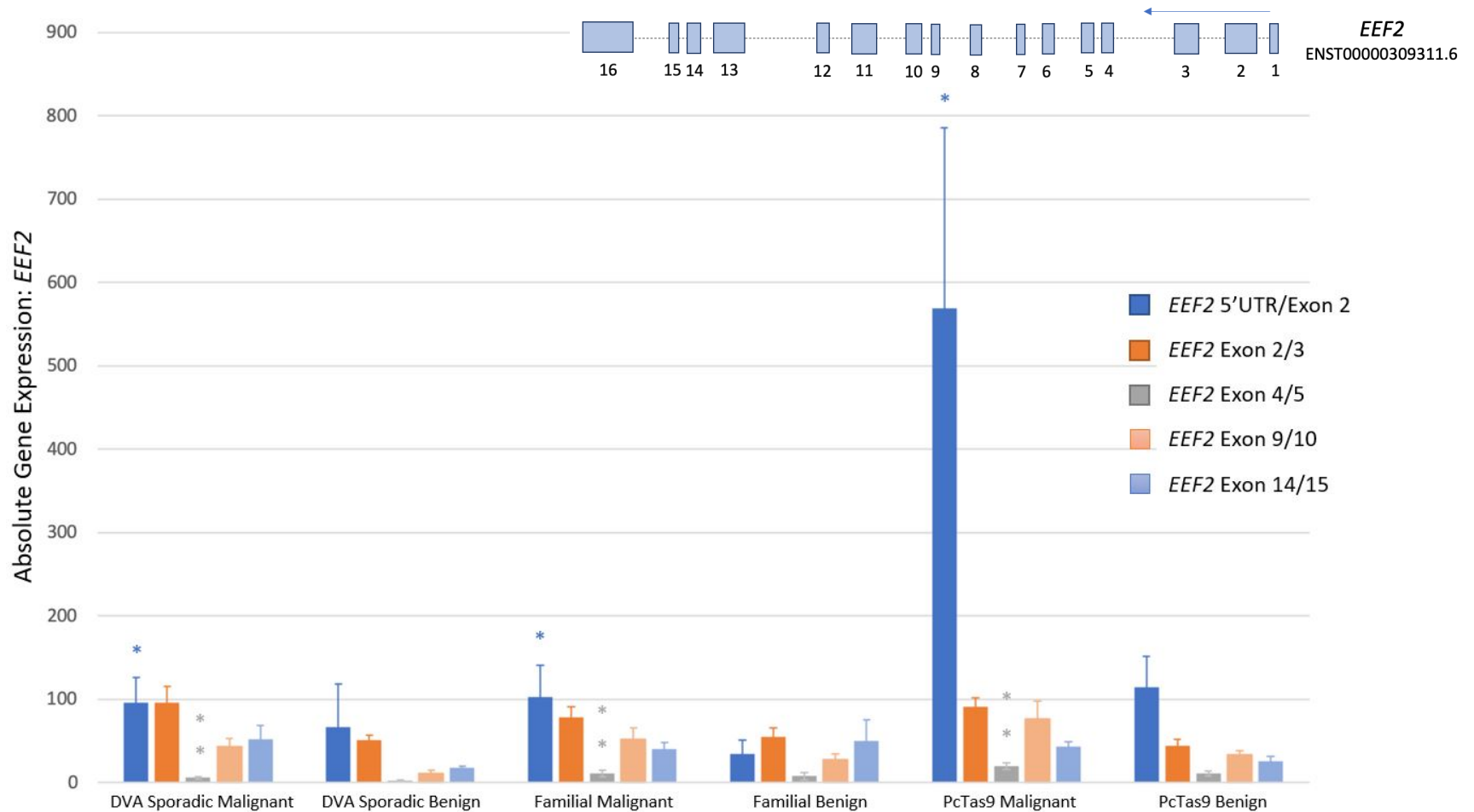
Figure 6.5 *EEF2* gene expression analysis in malignant and benign prostate glands from sporadic, familial and PcTas9 tumours.

EEF2 expression in five different regions of the gene was assessed in prostate tumours from three patient groups, sporadic (DVA), familial (PC) and PcTas9. A schematic of the most commonly transcribed isoform of *EEF2* in the prostate is shown at the top of the page. Absolute *EEF2* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. Shown here is the average absolute *EEF2* gene expression for each patient group for benign and malignant glands, with regions of *EEF2* depicted by different colours. Those considered to be significantly upregulated compared to other PcTas9 tumours are indicated by an *.

Initial observation of individual tumour expression revealed that not all PcTas9 tumours were overexpressing these two *EEF2* regions (Appendix 24). Random intercepts models to estimate and compare mean levels of absolute *EEF2* gene expression validated this finding and hence, the PcTas9 family was included in the model. Analysis of *EEF2* gene expression in the 5'UTR/exon 2 region, in only PcTas9 samples, revealed that the higher expression was driven by six samples (p=0.001), five of which showed amplification at 19p13.3 by aCGH analysis (PC9-12, 20, 532, 627 and 645; Figure 6.6). Notably, PC9-158 failed array QC, but had amplification of *EEF2* and this validated in our gene expression analysis, however the three other tumours with gain of *EEF2* on the array, PC9-447, PC9-588 and PC9-620 did not have significantly high 5'UTR/exon 2 expression. The six samples with significantly higher 5'UTR/exon 2 expression also had significantly higher expression of the exon 4/5 region compared to the other PcTas9 tumours (n=12, p=0.004).
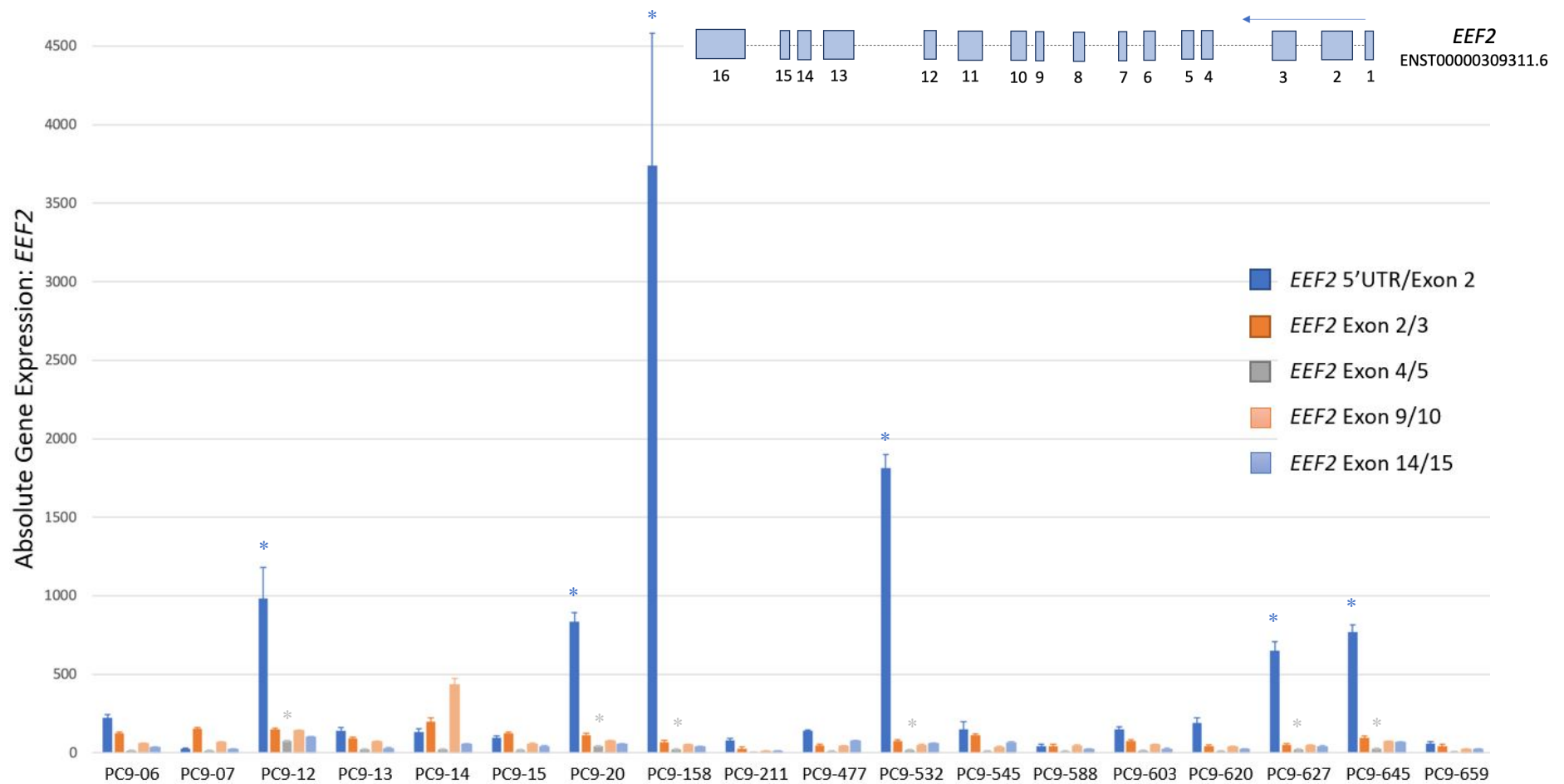
Figure 6.6 *EEF2* gene expression analysis in malignant prostate glands from PcTas9 tumours.

EEF2 expression in five different regions of the gene was assessed in prostate tumours from PcTas9 cases. A schematic of the most commonly transcribed isoform of *EEF2* in the prostate is shown at the top of the page. Absolute *EEF2* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. Individual PcTas9 malignant gland expression is shown here, with regions of *EEF2* depicted by different colours. Those considered to be significantly upregulated compared to other PcTas9 tumours are indicated by an *.

When *EEF2* expression was analysed in benign glands, there were no significant differences in gene expression across any of the regions between PcTas9 and non-Pctas9 tumours. Furthermore, analysis of *EEF2* 5'UTR/exon 2 expression in PcTas9 tumours with matched malignant and benign samples (n=7), indicated that three of the four tumours with overexpression of this region in malignant glands, also had very high 5'UTR/exon 2 malignant/benign expression ratios compared to tumours with no 'overexpression' (Table 6.6). These results suggest that *EEF2* overexpression is an anomaly of malignant glands only. Notably, three of the tumours with 5'UTR/exon 2 overexpression clustered within one specific branch of the PcTas9 pedigree (Figure 6.7).

Table 6.6 Malignant/benign *EEF2* 5'UTR/Exon 2 expression ratios in PcTas9 tumours.

Sample Identification	Amplification of *EEF2* on the aCGH	*EEF2* 5'UTR/Exon 2 overexpression in malignant glands	Malignant/Benign *EEF2* 5'UTR/Exon 2 Ratio
PC9-12	Yes	Yes	3.14
PC9-158	Yes (but did not pass QC)	Yes	112.51
PC9-477	Yes	No	1.34
PC9-532	Yes	Yes	21.68
PC9-588	Yes	No	0.29
PC9-620	Yes	No	3.01
PC9-645	Yes	Yes	15.46
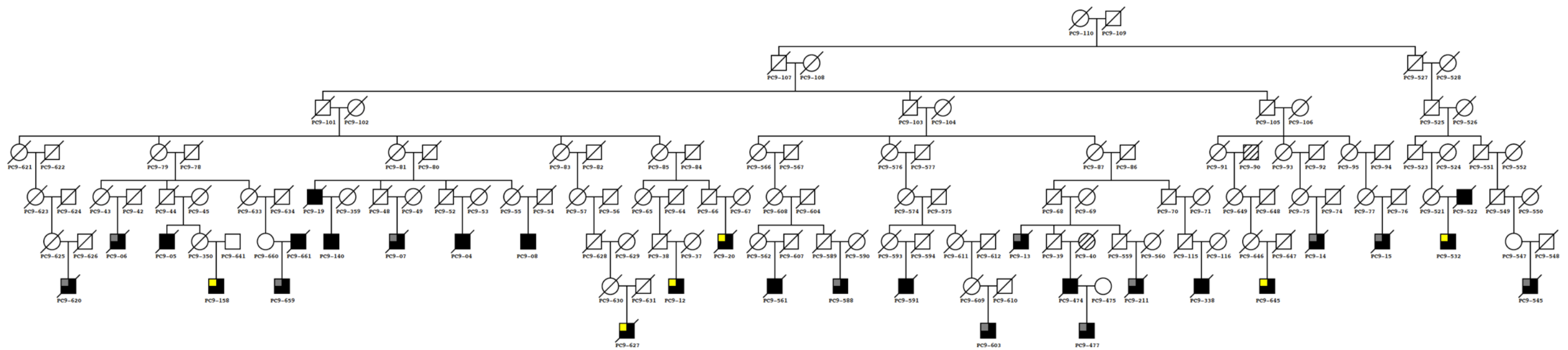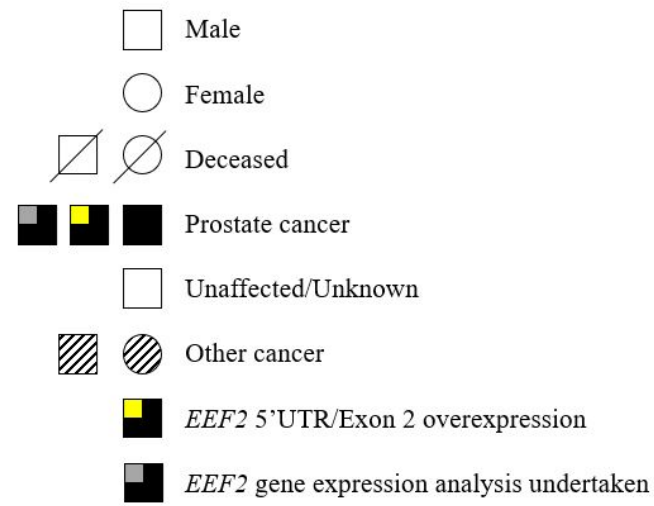
Figure 6.7 A condensed PcTas9 pedigree showing tumours with *EEF2* 5'UTR/Exon 2 overexpression in malignant glands.

This condensed PcTas9 pedigree indicates those tumours assessed for *EEF2* 5'UTR/exon 2 gene expression; tumours with *EEF2* 5'UTR/exon 2 overexpression in malignant glands are shown in yellow and those with expression similar to the rest of the dataset, in grey.

To determine whether *EEF2* overexpression was gene or region specific, expression of a neighbouring gene, *DAPK3* was determined. *DAPK3* expression was analysed by RT-qPCR in three regions across the gene, including exon 3/4, 4/5 and 7/8 (Appendix 25). Analysis of 14 paired malignant-benign samples found *DAPK3* expression to be similar in both gland types across all three regions (p=0.42, 0.48 and 0.52, respectively; Figure 6.8). However, in malignant glands only, *DAPK3* exon 3/4 expression was significantly different between PcTas9 (n=17) and non-PcTas9 samples (n=12; p=0.04; Figure 6.9). *DAPK3* expression in the non-PcTas9 malignant samples was approximately 3.7-fold higher than in the PcTas9 samples. When focusing on the six tumours with significantly higher *EEF2* 5'UTR/exon 2 expression, these were found to have lower average malignant *DAPK3* expression (n=6) compared to the remainder of the PcTas9 tumours (n=11), however this was not statistically significant (p=0.12). This also remained insignificant when just comparing *DAPK3* exon 3/4 expression (p=0.25), yet the six *EEF2* overexpressing tumours did on average have lower expression of this region. There was no detectable difference in *DAPK3* gene expression in any region between the benign glands of PcTas9 (n=7) and non-PcTas9 tumours (n=9; p=0.24, 0.38 and 0.46, respectively).
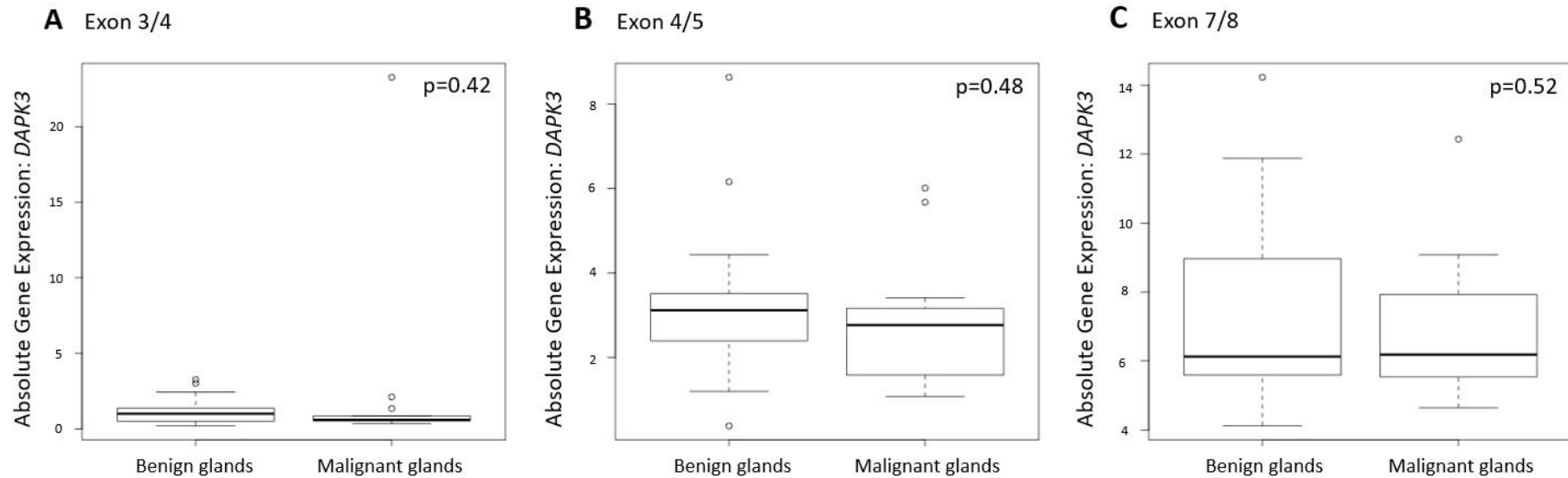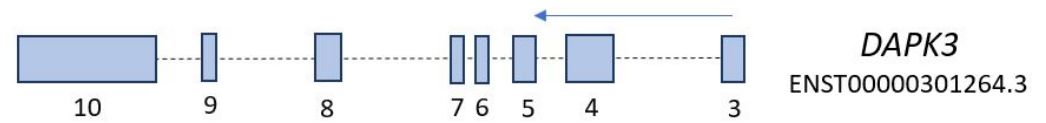
Figure 6.8 *DAPK3* gene expression analysis in malignant and benign prostate glands.

DAPK3 expression in three different regions of the gene was assessed in prostate tumours with matched malignant and benign glands (n_{pairs}=14). A schematic of the most commonly transcribed isoform of *DAPK3* in the prostate is shown at the top of the page. Absolute *DAPK3* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. *DAPK3* expression in malignant and benign glands in each region was compared using a paired Student's t-test. Median expression is shown by the thick black line, the interquartile range (middle 50% of data set) is represented by the box, and the minimum and maximum values by the whiskers (dotted lines). Individual outliers are shown with dots.
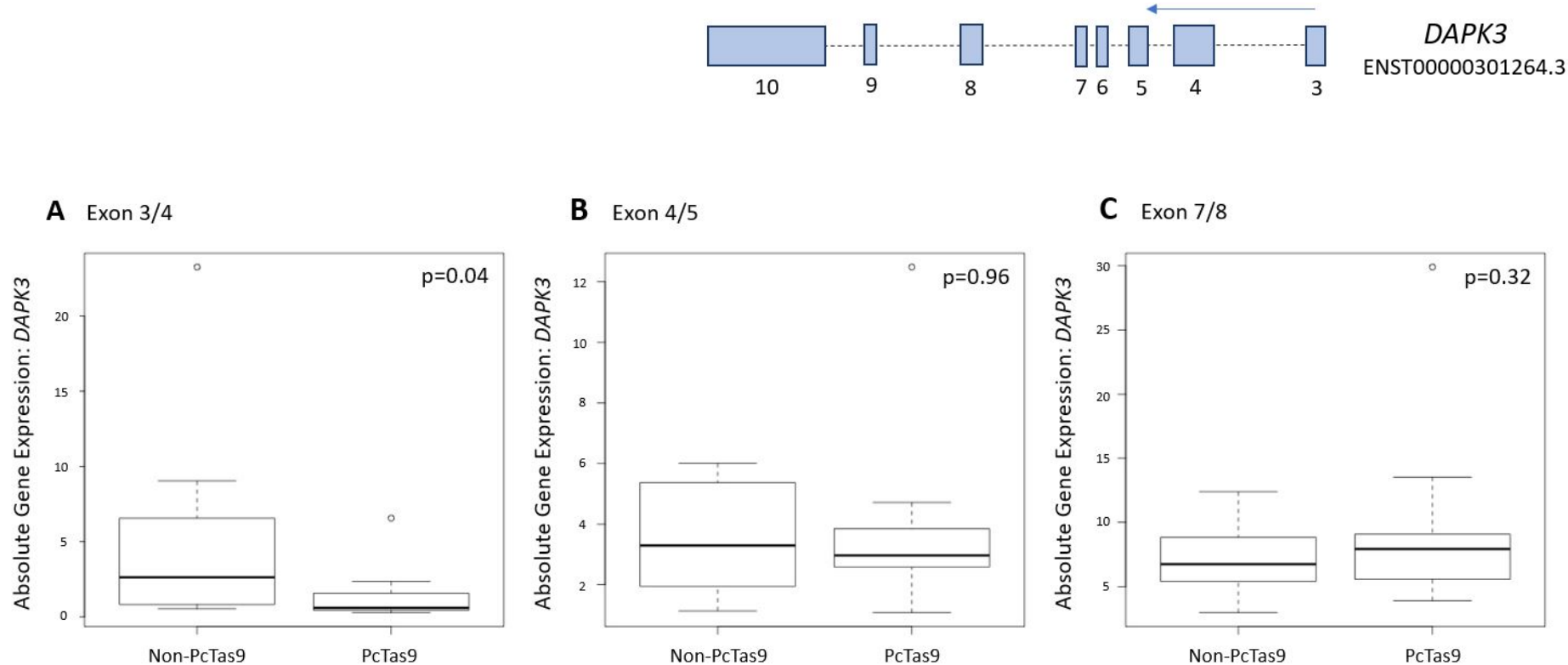
Figure 6.9 *DAPK3* gene expression analysis in malignant glands from non-PcTas9 cases compared to PcTas9 cases.

DAPK3 expression in three different regions of the gene was assessed in malignant prostate glands from two patient groups, non-PcTas9 (comprising sporadic and familial tumours; n=12) and PcTas9 (n=17). A schematic of the most commonly transcribed isoform of *DAPK3* in the prostate is shown at the top of the page. Absolute *DAPK3* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. *DAPK3* expression in malignant glands from non-PcTas9 and PcTas9 tumours in each region was compared using an unpaired Student's t-test. The spread of the data is represented by a box and whisker plot. Median expression is shown by the thick black line, the interquartile range (middle 50% of data set) is represented by the box, and the minimum and maximum values by the whiskers (dotted lines). Individual outliers are shown with dots.

6.3.5 Association of *EEF2* and *DAPK3* expression with clinical characteristics and tumour pathology

EEF2 gene expression was analysed as a mean of malignant gland expression per patient group (column 2 in Table 6.7) using an unpaired Student's t-test. While *EEF2* expression appeared to be higher in tumours with a lower GS ($\leq$7 (3+4), n=25) compared to a higher GS ($\geq$7 (4+3), n=15), this was not statistically significant (p=0.09). There was also no significant difference in average *EEF2* expression between patients with an early age of disease onset (<65 years, n=21) compared to those diagnosed $\geq$65 years of age (n=20; p=0.37).

In terms of *DAPK3* expression (column 4 in Table 6.7), similar levels of expression were observed between tumours with a low GS ($\leq$7 (3+4), n=17) and tumours with a high GS ($\geq$7 (4+3), n=11; p=0.72). There was also no detectable difference in average *DAPK3* expression between patients with an early age of disease onset (<65 years, n=11) compared to those diagnosed $\geq$65 years of age (n=18, p=0.70).

Table 6.7 Clinicopathological characteristics of FFPE prostate tumour samples assayed for EEF2 gene or protein expression and *DAPK3* gene expression .

Sample Identification	Malignant Gland Average *EEF2* Gene Expression	Malignant Gland EEF2 Protein Expression	Malignant Gland Average *DAPK3* Gene Expression	Age at Diagnosis	Tumour Grade[1]	Gleason Score[2]
DVA 67		0.70		61	-	6 (2+4)
DVA 157	58.88		7.55	66	-	7 (3+4)
DVA 167	16.67	1.60		53	PD	9 (5+4)
DVA 216	91.89	0.20		64	-	5 (3+2)

Blank cell= sample was not analysed; [1]Tumour grade obtained from pathology report; [2]Gleason Score obtained from pathology report; WD: well differentiated; MD: moderately differentiated; PD: poorly differentiated; W/MD: well-moderately differentiated; M/PD: moderately-poorly differentiated; -: information not present in original pathology report.

Sample Identification	Malignant Gland Average *EEF2* Gene Expression	Malignant Gland EEF2 Protein Expression	Malignant Gland Average *DAPK3* Gene Expression	Age at Diagnosis	Tumour Grade[1]	Gleason Score[2]
DVA 220	94.53	0	8.03	63	MD	6 (3+3)
DVA 302		2.70		65	W/MD	6 (3+3)
DVA 303		2.40		68	M/PD	7 (3+4)
DVA 402	86.94	2.00		52	MD	6 (3+3)
DVA 416	27.90	2.00		62	MD	6 (3+3)
DVA 422		1.60		60	M/PD	7 (3+4)
DVA 1002	32.61	1.80		61	WD	6 (3+3)
DVA 1006		0.70		67	-	6 (3+3)
DVA 1036		0.50		57	-	6 (3+3)
DVA 1050		0.60		63	-	5 (3+2)
DVA 1086		3.00		57	-	7 (4+3)
PC3-08		2.00		69	MD	6 (3+3)
PC3-31		0.80		54	-	5 (3+2)
PC4-03	95.98	1.40	4.56	80	M/PD	7 (4+3)
PC9-04		1.60		63	MD	6 (3+3)
PC9-06	90.43	1.60	6.61	79	-	-
PC9-07	56.53	0.50		71	PD	10 (5+5)
PC9-12	290.55	2.00	2.33	66	MD	6 (3+3)
PC9-13	70.97	1.40	4.77	83	-	9 (4+5)
PC9-14	169.14	1.60	4.97	79	MD	6 (3+3)
PC9-15	66.80	2.40	3.39	64	MD	5 (2+3)
PC9-20	224.60	0.70	3.36	76	PD	-
PC9-158	783.88	0.60	3.02	63	-	6 (3+3)
PC9-211	26.46	0.70	3.86	68	PD	9 (4+5)
PC9-338		2.00		63	-	6 (3+3)
PC9-474		1.80		74	PD	9 (4+5)
PC9-477	63.00		3.94	55	-	6 (3+3)
PC9-532	402.96	0.80	3.95	70	-	6 (3+3)
PC9-545	75.05	0.80	16.32	55	PD	-
PC9-561		2.00		63	MD	6 (3+3)

Blank cell= sample was not analysed; [1]Tumour grade obtained from pathology report; [2]Gleason Score obtained from pathology report; WD: well differentiated; MD: moderately differentiated; PD: poorly differentiated; W/MD: well-moderately differentiated; M/PD: moderately-poorly differentiated; -: information not present in original pathology report.

Sample Identification	Malignant Gland Average *EEF2* Gene Expression	Malignant Gland EEF2 Protein Expression	Malignant Gland Average *DAPK3* Gene Expression	Age at Diagnosis	Tumour Grade[1]	Gleason Score[2]
PC9-588	33.31	2.00	3.28	63	MD	6 (3+3)
PC9-603	62.82		3.79	73	MD	6 (3+3)
PC9-620	61.18	1.60	5.28	71	PD	9 (4+5)
PC9-627	161.46	1.40	1.89	65	-	7 (3+4)
PC9-645	206.05	0.10	2.97	60	-	7 (3+4)
PC9-659	30.68	1.60	4.25	65	-	9 (4+5)
PC9-951		1.00		80	WD	-
PC11-11	56.67	2.40	2.84	85	-	7 (3+4)
PC11-12	52.47			58	-	9 (4+5)
PC11-19		0		63	-	3 (2+1)
PC12-01	47.94	0	4.88	63	MD	6 (3+3)
PC12-03	29.25	0		62	WD	4 (2+2)
PC12-06	82.59	1.20	6.94	80	-	7 (3+4)
PC12-07	23.39	0.60	4.67	59	PD	9 (4+5)
PC12-08	1.95	1.80		73	-	6 (3+3)
PC12-09	24.68	0	3.53	68	-	6 (3+3)
PC19-02	20.29	1.20		50	-	6 (3+3)
PC22-17	49.94	0.80		56	MD	6 (3+3)
PC22-576	148.29	1.60	11.19	69	M/PD	7 (3+4)
PC23-02		0.50		78	MD	7 (3+4)
PC31-01	32.78	1.00		61	PD	10 (5+5)
PC60-01		1.60		58	WD	6 (3+3)
PC72-04	38.93	0.70	2.47	70	PD	9 (4+5)
PC72-06	20.42	0.70	3.25	62	-	8 (4+4)
PC213-991		2.00		68	-	9 (4+5)
PC3250-01	114.12	0.50	2.71	51	PD	9 (4+5)

Blank cell= sample was not analysed; [1]Tumour grade obtained from pathology report; [2]Gleason Score obtained from pathology report; WD: well differentiated; MD: moderately differentiated; PD: poorly differentiated; W/MD: well-moderately differentiated; M/PD: moderately-poorly differentiated; -: information not present in original pathology report.

6.3.6 Assessment of the chromosomal gain at 19p13.3 by protein expression analysis

IHC was performed on 56 FFPE prostate tumour samples to assess EEF2 protein expression. EEF2 staining intensity ranged from none (0) to strong (3) across the dataset, and the percentage of EEF2 positive nuclei ranged from approximately 5-100% (Figure 6.10; Appendix 24). Analysis of the quasi-continuous score (staining intensity x % of EEF2 positive nuclei) from 49 samples with paired malignant and benign glands, revealed increased EEF2 expression in malignant compared to benign glands (p=0.02; Figure 6.11). Analysis of malignant glands from non-PcTas9 tumours (n=36) and PcTas9 tumours (n=21) indicated no significant difference between the two patient groups (p=0.33, Figure 6.11). A similar result was observed for benign glands (p=0.57). For those PcTas9 tumours with significantly increased *EEF2* 5'UTR/exon 2 gene expression, there was no corresponding increase in protein expression compared to other PcTas9 tumours.

6.3.7 Association of EEF2 protein expression with clinical characteristics and tumour pathology

EEF2 protein expression was analysed as malignant gland expression per patient group using an unpaired Student's t-test (column 3 in Table 6.7). No correlation was observed between EEF2 protein expression in malignant glands and GS; tumours with a GS ≤7 (3+4) had similar average expression (n=38) compared to tumours with a GS ≥ 7(4+3) (p =0.47; Table 6.7). However, analysis of EEF2 expression and age at diagnosis revealed that expression was higher in tumours from men diagnosed 65 years of age and over (n=26) compared to those under 65 (n=30), however this was not statistically significant (p=0.07; Table 6.7).
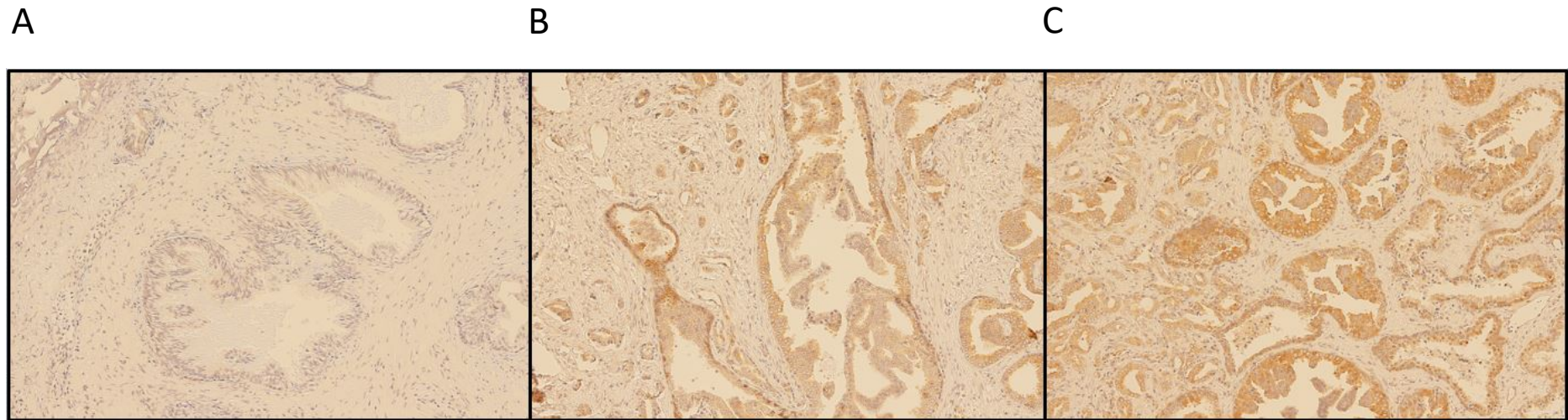
Figure 6.10 EEF2 protein expression in FFPE prostate tumour samples.

EEF2 protein expression was assessed in 56 prostate tumour specimens from the *Tasmanian Prostate Tissue Pathology Resource* to determine whether the amplification of *EEF2* was translated to the protein level. In short, IHC using an antibody targeting amino acid 31-80 of the EEF2 protein was utilised to assess protein expression. Staining intensity was scored as weak, moderate or strong. **A)** Weak staining of EEF2 in the plasma membrane and cytosol of benign prostate glands. **B)** Moderate staining of EEF2 in the plasma membrane and cytosol of benign prostate glands. **C)** Strong staining of EEF2 in the plasma membrane and cytosol of malignant prostate glands. Images were taken with an Olympus BX53 microscope, using the DP73 camera and software (x100).
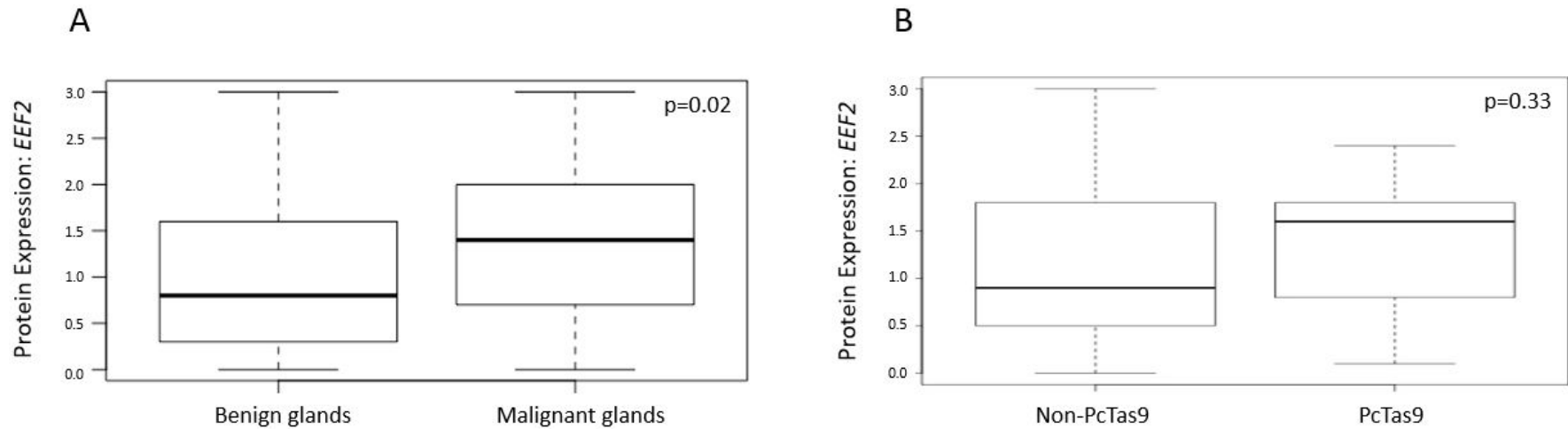
Figure 6.11 EEF2 protein expression analysis in malignant and benign prostate glands, and in malignant glands from non-PcTas9 and PcTas9 tumours.

EEF2 protein expression was calculated as a quasi-continuous score (staining intensity x % of EEF2 positive nuclei) for both malignant and benign glands in all samples. The spread of the data is represented by a box and whisker plot. Median expression is shown by the thick black line, the interquartile range (middle 50% of data set) is represented by the box, and the minimum and maximum values by the whiskers (dotted lines). **A)** EEF2 expression was assessed in prostate tumours with matched malignant and benign glands (n_{pairs}=49). EEF2 expression in malignant and benign glands was compared using a paired Student's t-test. **B)** EEF2 expression was assessed in malignant prostate glands from two patient groups, non-PcTas9 (comprising sporadic and familial tumours; n=36) and PcTas9 (n=21). EEF2 expression in malignant glands from non-PcTas9 and PcTas9 tumours was compared using an unpaired Student's t-test.

6.4   DISCUSSION

6.4.1   Overall findings

To date, only two CGH studies have investigated genetic changes in familial PCa tumours from high-risk families. To our knowledge, this study is the first to examine whether chromosomal alterations are shared by affected individuals within the one family. Plus, the first to validate regions of aberrations identified by aCGH using RT-qPCR. aCGH analysis of 12 prostate tumours from family PcTas9 identified two consistent regions of loss, including 1p36.21 and 19p13.3. Four regions of gain were also consistently observed across four or more tumours, including 6p23-p22.3, 6p24.2, 17p13.3 and 19p13.3 (Table 6.4). Notably, all eight tumours that passed QC were shown to harbor a gain at 19p13.3, a region which has been identified as a PCa susceptibility region by linkage analysis [300,301]. Three genes were amplified in this region, including *PTPRS, ZBTB7A* and *EEF2*, however a region of gain overlaid the *EEF2* gene in all eight tumours. Interestingly, EEF2 has been postulated as a potential biomarker of PCa [299]. Follow-up gene expression analysis of our entire *Tasmanian Prostate Tissue Pathology Resource* identified very high expression of the 5'UTR/exon 2 region of *EEF2* in PcTas9 malignant glands compared to other familial and sporadic cases. Further statistical analysis identified six PcTas9 tumours which were driving this overexpression; all of which showed amplification on the array (PC9-158 failed QC, but still showed *EEF2* aCGH amplification), thus further validating our results. Overall, the aim of this study was to determine if there was an inherited genetic predisposition to the tumour CNV changes identified in PcTas9. However, due to time and sample size limitations, we were only able to identify consistent regions of gain and loss in PcTas9 and validate one of these, an amplification of *EEF2*.

6.4.2   Potential effects of an *EEF2* amplification

Remarkably, eight PcTas9 tumours displayed an amplification of the *EEF2* gene at 19p13.3. Here, the *EEF2* amplification observed by aCGH was validated by RT-qPCR where five of these samples had increased expression at the gene level (plus PC9-158). *EEF2* is an essential factor for protein synthesis as it promotes the GTP-dependent translocation of the nascent protein chain from the A to the P-site of the ribosome [116]. It is overexpressed in a diverse range of cancer types, including PCa, and interestingly, has recently been suggested as a potential biomarker of PCa [299]. Given that *EEF2* mediates protein synthesis, which is one of the key characteristics of cancer cells, some studies have examined the contribution of *EEF2* to

tumourigenesis. A study by Nakamura *et al.* (2009) found that overexpression of *EEF2* in gastric cancer cell lines significantly enhanced cell growth through promotion of G2/M progression in the cell cycle, activated Akt and cdc2, and inactivated EEF2 kinase [302]. Overexpression of EEF2 in these cancer cells enhanced *in vivo* tumourigenicity in a mouse xenograft model, suggesting that overexpressed EEF2 promotes G2/M progression and enhances cancer cell growth *in vitro* and *in vivo* [302]. Such studies suggest a link between translational elongation and cell cycle mechanisms, and disruption of this link may lead to dysregulation and cancer promotion. Thus, the *EEF2* amplifications observed in our study may result in cell cycle alterations, leading to increased tumourigenesis.

6.4.3   Examining *EEF2* gene and protein expression in prostate tumours

The region of gain identified by aCGH analysis encompassed most of the *EEF2* gene, therefore, we aimed to determine exon-level expression across a number of *EEF2* exons to verify this result. At the gene level, significantly higher expression was observed in malignant compared to adjacent benign glands, in three out of the five regions assessed. The most significant finding from our study was that malignant glands from PcTas9 tumours had higher expression of *EEF2* in the 5'UTR/exon 2 and exon 4/5 regions compared to non-PcTas9 tumours. In fact, expression was driven by six PcTas9 tumours, all of which demonstrated amplification on the CGH array (PC9-158 failed QC, but showed aCGH amplification). Given this validation, it is possible to hypothesise that other samples with apparent high 5'UTR/exon 2 expression (>200) may also have amplification of this region. Notably, this included one other PcTas9 tumours that was not aCGH assayed (PC9-06), plus three other tumours from families PcTas12, PcTas22 and PcTas3250 (Appendix 24).

It has been reported that the EEF2 protein is highly expressed in human carcinoma tissue, but not in normal tissue [302,303]. Studies have reported EEF2 overexpression in ovarian [304] and breast cancer [305], and more recently in lung, gastric, colorectal and hepatocellular carcinoma tissue [302,303,306,307]. In fact, Nakamura and colleagues (2008) demonstrated that EEF2 was overexpressed in 92.9% of gastric and 91.7% of colorectal cancers [302]. To date, only two studies have assessed EEF2 expression levels in PCa, with lower percentages of overexpression compared to other cancers. Oji *et al.* (2014) examined four prostate samples, three of which overexpressed EEF2 [303]. More recently Zhang *et al.* (2018) examined 97 prostate tumours and found that 76.29% were EEF2 positive [299]. In our study, a significant increase in EEF2 expression in malignant glands compared to adjacent benign glands was observed (n_{pairs}=49,

p=0.02). In total, 49% of tumours had EEF2 overexpression, whereas 20% had comparable expression between malignant and benign glands. Overall, 87.72% of malignant and 94.74% of benign glands were EEF2 positive, thus, percentages were higher compared to the study by Zhang and colleagues (2018).

In terms of *EEF2* and clinical characteristics, overexpression has been shown to be associated with poor patient survival in ovarian cancer [308] and hormone receptor positive breast cancer [309]. A study by Shi *et al.* (2018) observed that EEF2 expression gradually increased with GS (more aggressive), and it correlated significantly with tumour grade (p=0.045) [309]. Zhang and colleagues (2018) observed a correlation between EEF2 protein expression and clinicopathological characteristics of PCa, in particular, the staining intensity of EEF2 was significantly associated with age, level of prostate-specific antigen and GS [299]. Our study found no significant difference in EEF2 gene or protein expression between tumours with a low and high GS (≤7 (3+4) *versus* ≥ 7(4+3)) or those diagnosed before or after 65 years of age.

6.4.4 Examining *DAPK3* gene expression in prostate tumours

To determine whether this amplification was a gene or region-specific anomaly, expression of a neighboring gene, *DAPK3* was also examined. *DAPK3* was not amplified on the array, therefore we wanted to validate this finding by RT-qPCR. *DAPK3* expression in the exon 3/4 region was determined to be 3.7-fold lower in PcTas9 tumour samples compared to non-PcTas9 tumours (p=0.04). The six PcTas9 tumours with significant *EEF2* 5'UTR/exon 2 overexpression had lower average *DAPK3* expression compared to the remaining PcTas9 tumours, however this was not statistically significant. *Death-associated protein kinase 3* (*DAPK3*) is involved in the regulation of apoptosis, autophagy, transcription and translation [116]. It has been reported that *DAPK3* is frequently methylated or mutated in many cancer types, resulting in a loss of tumour suppression via *DAPK3* [310]. A study by Chen *et al.* (2016) identified a link between low *DAPK3* expression and shorter overall survival rates in endometrial cancer (p=0.023) [311]. Das and colleagues (2016) examined *DAPK3* expression in 29 FFPE prostate samples and found decreased expression in samples of higher GS [312]. Here, we identified no significant difference in *DAPK3* expression between tumours with a GS ≤7 (3+4) and those ≥7 (4+3), nor was the mean GS any different between PcTas9 and non-PcTas9 tumours. Whilst significant *DAPK3* exon 3/4 loss was not apparent in the tumours with significant *EEF2* 5'UTR/exon 2 overexpression, it is possible that both alterations, together or

independently, influence carcinogenesis. Given that *DAPK3*'s main function is to regulate apoptosis, and *DAPK3* overexpressing cells exhibit extreme apoptotic-like morphology [312], loss of *DAPK3* may enable cancer cells to bypass apoptosis, thus giving them a selection advantage over other cells with normal *DAPK3* expression.

6.4.5 Other previously identified regions of loss and gain

Our study of a single PCa family has identified regions of loss and gain previously identified by other studies. The most commonly altered region of the PCa tumour genome is 7p21 and here, aCGH analysis revealed three tumours with gain of this region, however the breakpoints were not consistent across samples. PC9-477 had an amplification at 7p21.1, overlying the *histone deacetylase 9 (HDAC9)* gene, which is involved in cell cycle regulation and development [313]. This region has also been shown to harbor risk alleles to pancreatic cancer [314]. PC9-532 had an amplification at 7p21.3, which overlies the *islet cell autoantigen 1 (ICA1)* gene. Interestingly, one PcTas9 sample had an amplification at 7p21.2, which overlies the *ETV1* gene. *ETV1* is a well-known gene in PCa tumorigenesis and is often involved in gene fusion events at the tumour level [315]. This amplification could therefore be the result of a fusion event involving *ETV1* and an unknown 5' fusion partner.

The chromosomal region of 17p13.3 was amplified in five PcTas9 tumours, all of which represent different branches of the family. Gain of 17p was reported by Rokman *et al.* (2001) in their study of familial PCa, however, this region of gain has not been identified in any sporadic tumours, suggesting an association with familial prostate tumourigenesis [296]. A number of interesting genes are present in this region and play a role in transcriptional repression, initiation of transcription, the replication and maintenance of chromosome telomeres, and cell growth and differentiation. Of particular interest is the *diphthamide biosynthesis 1* (*DPH1)* gene, which was amplified in three out of the five tumours. *DPH1* is an enzyme involved in the biosynthesis of diphthamide, a modified histidine found only in *EEF2* [116]. The fact that we have found disruptions to two different genes in the same pathway highlights the potential role of this pathway in tumourigenesis.

All eight of the PcTas9 tumours showed gain at 19p13.3, comprising multiple branches of the family, including PC9-20 and his second cousin, PC9-12. The 19p region of amplification has previously been identified in tumours from familial PCa cases by Rokman and colleagues (2001), however has not been identified in sporadic tumours [296]. Aside from *EEF2*, there are a

number of other interesting genes underlying the three regions of 19p13.3 gain that play a role in; the clearance of misfolded proteins, protein synthesis, cellular processes, transcriptional repression, and malignant cell proliferation [116]. One interesting gene is *ZBTB7A*, which was amplified in four PcTas9 samples (50%). ZBTB7A is a zinc finger protein that is moderately expressed in the prostate. Functional studies of a transgenic mouse model overexpressing *Zbtb7a* in the prostate, found that ZBTB7A suppresses castration-resistant PCa, through repression of a *Soxa9*-dependent pathway for cellular senescence bypass and tumour invasion [316]. In fact, analysis of PCa samples revealed that men whose tumours had high levels of nonfunctional ZBTB7A cells responded poorly to androgen-deprivation therapy [317]. Given that ZBTB7A upregulation in gastric cancer cells promotes apoptosis and represses cell migration [318], the amplification identified in these four PCa samples may promote carcinogenesis by disrupting transcription or translation leading to downregulation of the gene.

6.4.6 Consistently observed regions of loss in the PcTas9 tumours

The 1p36.21 region of deletion (up to 1.18Mb) was found in three PcTas9 tumours, which encompass both branches of the family. This region of loss has never been observed in PCa tumours however, it has been linked to ovarian cancer. A study by Dimova *et al.* (2009) involved CGH analysis of 28 ovarian tumours and the 1p36 region was lost in 40% of tumours and associated with late-stage cancers [295]. This region of loss includes genes in the *PRAME* and *HNRNPCL* gene families. *Preferentially expressed antigen of melanoma (PRAME)* family members are expressed in many cancer types, but also function in reproductive tissues during development [116]. *Heterogeneous nuclear ribonucleoprotein C like (HNRNPCL)* genes encode for RNA binding proteins, which influence pre-mRNA splicing processes and alterations could lead to alternative transcripts [116]. Thus, this region harbours an extensive number of genes that could be important in PCa.

Three PcTas9 tumours had a deletion at 19p13.3 and notably, these cases also had loss of 1p36.21. The region of 19p13.3 has been extensively studied, with linkage studies of hereditary PCa identifying it as a PCa susceptibly region [300,301]. This region of loss has only been observed in familial and not sporadic prostate tumours. Of the 21 familial tumours investigated by Rokman *et al.* (2001) only a small number showed an alteration [296]. Present in this region are a number of interesting genes which play a role in the antigen presentation process, the generation of cytotoxic T cells, and the activation and development of T and B cells [116]. Particularly interesting is the *TINCR* long non-coding RNA (LIC00036), which has been

suggested to have altered expression in multiple human cancers [319,320]. In a recent study by Dong and colleagues (2018), low-expression of *TINCR* was observed in PCa and correlated with advanced clinical tumour stage, lymph node involvement, distant metastasis, high GS and poor prognosis in their cohort of 160 tumours [321].

6.4.7 Consistently observed regions of gain in the PcTas9 tumours

A total of five PcTas9 prostate tumours were shown to harbour an amplification at 6p23-p22.3. The recurrent gain has not previously been identified in sporadic or familial PCa studies, however an amplification at 6p22 has been identified in bladder cancer [297] and retinoblastoma tumours [298]. This region of gain encompasses the *jumonji and AT-rich interaction domain containing 2 (JARID2)* gene, which is a putative transcription factor that plays a role in DNA binding, nuclear localisation, transcriptional repression and recruitment of the Polycomb-repressive complex 2 [322-324]. Whilst no study has explored whether this gene has a role in PCa, *JARID2* has consistently been identified to play a role in the initiation, proliferation and maintenance of tumour cells in ovarian and bladder cancer [325,326]. Thus, *JARID2* may also have a role in PCa initiation and development, and further assessment to determine whether this gene is disrupted by the amplification is warranted.

Four PcTas9 tumours were shown to harbor a gain at 6p24.2, a region which overlays the *neural precursor cell expressed developmentally down-regulated protein 9 (NEDD9)*. *NEDD9* is frequently overexpressed in diverse cancer types and has been linked to tumorigenesis of many different malignancies, including PCa and is reasonably expressed in the normal prostate [116]. *NEDD9* is also highly conserved across species, is repressed by estrogen in breast cancer cells [327] and is induced by Wnt signaling in colon cancer [328]. Interestingly, the region of amplification of *NEDD9* encompasses only the small transcript (NM_006403) and upon further investigation using the GTEx portal (https://gtexportal.org/home/), this is the most highly expressed transcript in the prostate [137]. Therefore, this region of amplification and specifically, *NEDD9,* seems a fitting candidate for follow-up functional studies in our Tasmanian prostate tumour resource.

6.4.8 Somatic tumour variation and germline predisposition

There is no known observable difference in the histopathology of sporadic and familial PCa tumours, however it is interesting that not all chromosomal alterations are observed in both sporadic and familial tumours. The two most commonly observed losses in tumours of sporadic

PCa, 16q and 18q, were not commonly identified in the tumours from PcTas9 men. These results reflect those described by Rokman *et al.* (2001) [293]. The often-unique chromosomal alterations of familial tumours, such as those presented here and previously, suggest that germline variants may initiate different genetic pathways that then lead to distinct somatic alterations compared to sporadic tumours. In this study, each of the consistently observed regions of loss and gain contributed to PCa tumours across multiple branches of the PcTas9 pedigree. Given that previous literature suggests some may be unique to familial tumours and in PcTas9 they are shared by distantly related individuals strengthens the likelihood of the CNV being linked to underlying inherited genetic factors.

Further evidence for a link between inherited germline variants and somatic chromosomal alterations was presented in a study of breast cancer. It is hypothesised that the number and types of chromosomal alterations are influenced by underlying predisposition genes. In fact, *BRAC1*- and *BRCA2*-associated breast cancers have more CNVs per tumour compared to sporadic breast cancers, as described by Tirkkonen *et al.* (1997) [329]. More recently, Joosse and colleagues (2012) developed a test to identify *BRCA2*-mutated breast tumours, using aCGH profiles of 28 *BRCA2*-mutated and 28 sporadic breast tumours [330]. They subsequently tested 89 breast tumours from suspected breast cancer families, with unknown *BRCA1/2* mutation status and they were able to separate *BRCA1*-like, *BRCA2*-like and sporadic-like tumours using the tumours chromosomal profile [330]. This shows that specific germline mutations, such as *BRCA1* and *BRCA2*, predispose to some somatic tumour alterations. In terms of PCa, a large study of 539 prostate tumours found that a 7p14.3 germline variant positively selects for *SPOP* mutant PCa, as the variant accelerates the DNA damage phenotype [331]. Whilst the mechanism linking the 7p14.3 germline variant and the *SPOP* somatic mutation remains elusive, it was suggested that future studies should investigate the role of the allele in the emergence of *SPOP* somatic alterations [331]. Overall, the association of germline variants and tumour CNVs requires further investigation, as the number of studies in this area is small.

6.4.9   Clinical significance of this study

Results from the study presented here could lead to the clinical implementation of routine cytogenetic analysis for prostate tumour tissue. The knowledge of specific somatic tumour alterations could define particular disease phenotypes (i.e. indolent or aggressive) and potential response to treatment. For example, Zafarana and colleagues (2012) demonstrated that overexpression of 8q (*cMYC*) alone, or when combined with a *PTEN* loss were increasingly

prognostic for relapse after radiotherapy [332]. If future studies confirm that the chromosomal alterations identified in this study are associated with clinical outcomes, men could be tested for these somatic aberrations at diagnostic biopsy, when the disease is most curable.

6.4.10 Limitations of this study

This study has provided important insights into chromosomal aberrations at the tumour level in a large Tasmanian PCa family, but there are some limitations in the interpretation of this data. A significant limitation of aCGH analysis is that translocations and inversions cannot be identified [333,334]. This is because balanced chromosomal rearrangements do not result in any loss or gain, however there are other approaches to identify such alterations, which will be discussed in Chapter 7. Another significant limitation is admixture, or contamination of malignant with benign cells, which can skew results [334-336]. In terms of admixture, it is known that foci within the one tumour can be genetically very different [270], thus, nucleic acid extractions not macrodissected in parallel can result in very different genomic profiles, which makes interpretation of data much more complex. Here, three tumours, PC9-477, PC9-588 and PC9-620 had gain of *EEF2* on the array, however did not show *EEF2* overexpression in our gene expression analysis. This result is one such example of the potential effect of tumour heterogeneity. The FFPE nucleic acid samples were not co-extracted, nor macrodissected at the same time, therefore these results may represent the genomic profile of completely different tumour foci. Contamination of malignant samples with benign cells can also mask chromosomal gains and losses, thereby reducing the detection of true disease-associated genetic alterations [336]. Laser capture microdissection could deal with both of these issues, by almost guaranteeing a homogenous cell population for analysis. On another note, the nature of denatured chromosomes and the integration of fluorescent labels, can also cause the colour ratio signal to be spread over a larger region than the actual amplicon [334]. This could mean that the 19p13.3 amplification may not spread over the entirety of the *EEF2* gene as observed in the aCGH data. Instead, as reflected in the gene expression results, amplification may have been restricted to the 5'UTR/exon 2 region only. Although, another region (exon 4/5) was also significantly overexpressed in the six PcTas9 tumours that had overexpression of the 5'UTR/exon 2 region, yet the amplitude of overexpression was on average 60-fold lower than the 5'UTR/exon 2 region. Data from the GTEx Portal (https://gtexportal.org/home/) [137] suggests that these two regions have similar expression levels, therefore the discrepancies in levels of expression may be due to the chromosomal amplification or simply due to different primer efficiencies. To succumb this issue, the amplification break points could be accurately

mapped using PCR or where available, whole-genome data. The quality of DNA and RNA extracted from FFPE tissue is also fairly poor, therefore it is possible that this may have impacted our gene expression results. Overall, this study's sample size was quite small, thus, the concepts explored in this study should be assessed in a larger tissue cohort. Further studies will confirm the presence of the *EEF2* amplification and other chromosomal alterations in other *Tasmanian Prostate Tissue Pathology Resource* tumours and thus, whether they are genetically predisposed.

6.5   FUTURE DIRECTIONS

In summary, this study has identified chromosomal regions of deletion and amplification present in prostate tumours from PcTas9 men. Regions that were consistently deleted in three or more tumours included 1p36.21 and 19p13.3, whilst gains included 6p23-p22.3, 6p24.2, 17p13.3 and 19p13.3. The high resolution of aCGH compared to previous CGH analyses enabled us to identify genes underlying these regions. Of particular interest, this study highlighted chromosomal regions which may harbour genes involved in tumour development, including *TINCR, JARID2, NEDD9, DPH1, ZBTB7A,* and *EEF2*. *EEF2* was targeted for follow-up in this study due to the fact that 100% of PcTas9 tumours assayed showed amplification. Therefore, future work could involve assessing gene and protein expression of the other regions of loss and gain in the larger *Tasmanian Prostate Tissue Pathology Resource*. A particularly interesting candidate is *NEDD9,* because like *EEF2,* is frequently overexpressed in diverse cancer types and has been linked to tumorigenesis of many different malignancies [337-339]. Overall, the *EEF2* amplification was the most predominant alteration detected in tumours from PcTas9 cases, suggesting an inherited predisposition. However, single tumours from other Tasmanian families (PcTas12, 22 and 3250) also showed a similar pattern of overexpression, which would suggest that this phenomenon isn't restricted to PcTas9, and the amplification may in fact be more of a widespread occurrence in familial PCa. Thus, further aCGH analysis of tumours from other familial and sporadic PCa cases may provide additional insight into this and other chromosomal alterations. Further work will highlight the significance of differences between sporadic and familial tumours, plus, facilitate the investigation of the link between genetic predisposition and these tumour variations. In the future, the collection of additional tumours from newly diagnosed cases in PcTas9 will enable us to assess whether the *EEF2* amplification (or other tumour CNVs) clusters in closely related individuals in this large family. Plus, genome-wide germline genetic data from PcTas9 cases will permit us to perform

linkage analysis weighted on the presence/absence of the 19p13.3 amplification. This could lead to the identification of chromosomal regions and thus, inherited germline variants underpinning this amplification at the tumour level. Finally, the tumours from the other Tasmanian PCa families who were identified to have high *EEF2* 5'UTR/exon 2 expression should also be assessed by aCGH to validate this finding.

6.6   CONCLUSION

This study sought to identify CNVs in prostate tumours from a single Tasmanian family, PcTas9, with the overall aim to investigate underlying genetic drivers of these tumour events. The *EEF2* gene was consistently amplified in all eight tumours examined, and follow-up gene expression analysis revealed that six had significantly higher expression of the 5'UTR/exon 2 and exon 4/5 regions compared to other PcTas9 tumours. This is now one of very few studies to examine EEF2 protein expression in prostate tumours, however, to the best of our knowledge, the first to assess *EEF2* gene expression. Whilst study limitations restricted us from investigating whether germline variation predisposes this amplification, the recent generation of genome-wide germline data from this family will enable us to assess this hypothesis in the near future. Overall, given the known overexpression of EEF2 in cancer and the recent suggestion that it is an ideal therapeutic target, preliminary findings from this study are very promising.

CHAPTER 7 : GENE FUSIONS IN TASMANIAN PROSTATE TUMOURS

7.1   INTRODUCTION

Gene fusions are prominent in malignant tumours, with a total of 297 reported by the Catalogue of Somatic Mutations in Cancer (COSMIC) [340]. A gene fusion is a hybrid gene formed by the combination of two separate genes, with the regions of the genes fused together known as the fusion break points. Petrovics and colleagues (2005) identified one of the earliest genetic alterations in prostate tumours, the overexpression of the oncogene, *ERG*, which is a member of the large family of erythroblast transformation-specific (ETS) transcription factors [341]. It was subsequently found that in most cases ERG overexpression was driven by the fusion of the *ERG* gene (21q22.3) with *TMPRSS2* (21q22.2) [315]. *TMPRSS2* is an androgen-regulated gene that is preferentially expressed in the prostate and the fusion of the two genes results in the androgen-regulated overexpression of *ERG* [342]. Since this original study, many other studies have validated this recurrent fusion event in prostate tumours and have discovered additional *ETS* fusion events [343] (Table 7.1). It is now known that *ETS* genes are frequently involved in prostate gene fusions and they often result in the synthesis of chimeric proteins or altered expression of the ETS protein.

Table 7.1 ETS gene fusion partners involved in prostate cancer and their frequency.

ETS Gene	Fusion Partner(s)	Frequency [343]
ERG	*TMPRSS2, SCL45A3*	52%
ETV1	*TMPRSS2, SLC45A3, ACSL3, HERV-K, HERV-K17, FOXP1, EST14, chr14q13.3-14q21.1, C15orf21, HNRPA2B1, OR51E2*	7%
ETV4	*TMPRSS2, KLK2, CANT1, DDX5, UBTF*	1.5%
ETV5	*TMPRSS2, SLC45A3*	0.5%
FLI1	*SLC45A3*	0.5%

To date, the *TMPRSS2:ERG* fusion is the most common fusion event in prostate tumours, occurring in ~50% [315]. Normally, the *TMPRSS2* and *ERG* genes are located in close proximity (2.7Mb) to each other on chromosome 21 and are both transcribed in the reverse orientation. Fusion of *TMPRSS2:ERG* can occur by two mechanisms; firstly, the genomic region between the two genes can be lost by interstitial deletion, which occurs in approximately 60% of fusion positive tumours [344,345]. Secondly, less frequently, the fusion event can occur as a result of a complex genomic rearrangement, involving chromosome 21q22 and presumably other chromosomes [344,345]. Each mechanism can result in multiple fusion transcripts, in fact, there are over eight different *TMPRSS2:ERG* transcripts, the most common being the fusion of the first *TMPRSS2* exon(s) with exon 4 onwards of the *ERG* gene (Figure 7.1).
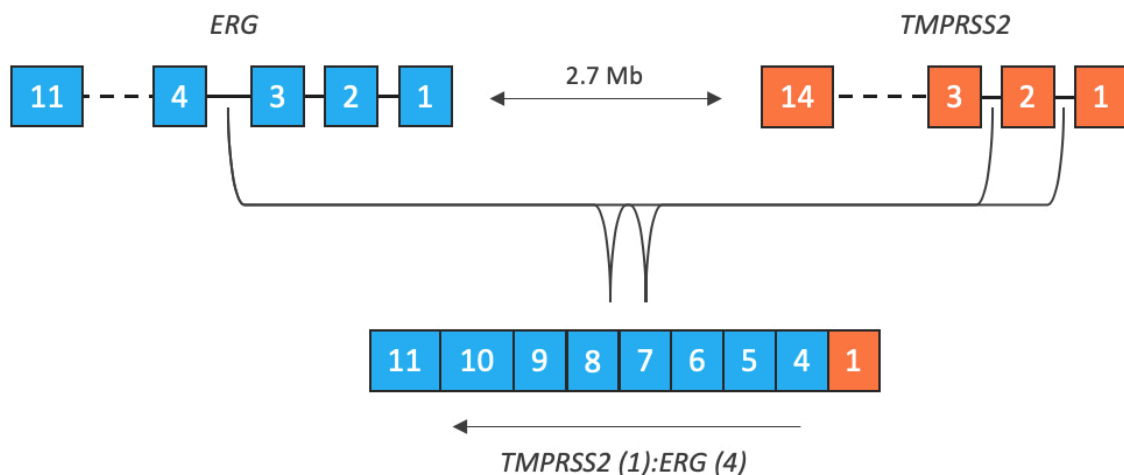
Figure 7.1 Schematic representation of the two most common *TMPRSS2:ERG* fusion transcripts. Exons 1-11 of *ERG* are shown in blue and exons 1-14 of *TMPRSS2* are shown in orange, with the schematic showing that they are in the same orientation, 2.7Mb apart. The most common *TMPRSS2:ERG* transcript includes exon 1 of *TMPRSS2* and exon 4 onwards of *ERG*, shown at the bottom of this figure (T1E4). The second most common fusion involves exon 1 and 2 of *TMPRSS2* and exon 4 onwards of *ERG*.

In the original study by Tomlins *et al.* (2005), a second recurrent gene fusion between another *ETS* family member, *ETV1* (7p21.2), and the *TMPRSS2* gene was discovered, with a total of 24.1% of prostate tumours harbouring the fusion event [315]. Currently there are over 10 different fusion partners of *ETV1* identified [315,346] and later studies have found a much lower frequency of events, with ~7% of all prostate tumours *ETV1* fusion positive [343]. *ETV1* gene fusions can

lead to overexpression of a truncated ETV1 protein that lacks the N-terminal topologically associating domain [315,346]. However, it can also be overexpressed as a full-length protein, due to translocation of the complete gene to a different genomic region [346].

In the last decade, our understanding of PCa development has changed radically with the discovery of *ETS* gene fusions. As mentioned previously (Chapter 1.3.5), *ETS*-rearrangements are used to subclassify PCa tumours, and recently they have been identified as potential novel urinary biomarkers for PCa diagnosis [347]. Tomlins and colleagues (2011) reported the use of a clinical-grade, transcription-mediated amplification assay to detect and stratify PCa tumours based on *TMPRSS2:ERG* fusion status [348]. Such studies demonstrate that urine-detected *TMPRSS2:ERG*, in combination with other PCa markers, enhances the utility of prostate-specific antigen (PSA) testing [348,349]. Several studies have evaluated the clinical significance of the *TMPRSS2:ERG* fusion event in prostate tumours and while some have demonstrated an association with advanced and invasive tumours with poor prognoses [350,351], others have shown that it is not a predictor of PCa recurrence or mortality [352,353]. Whilst the clinical consequence of *ETV1* fusion events resulting in overexpression of *ETV1* is not yet well understood, an *ETV1* expression signature was observed to be associated with aggressive PCa and poorer outcomes [354]. Given its role in testosterone production, the *ETV1* fusion may accelerate prostate carcinogenesis.

It is now estimated that approximately 50-60% of all PCa tumours harbor recurrent gene fusions [355]. Given the high frequency of these fusion events, and accumulating evidence from previous studies, they are unlikely due to chance. In fact, the *TMPRSS2:ERG* fusion is very consistent in its formation, and a high frequency suggests an underlying genetic predisposition [356]. Common PCa risk variants have been evaluated in cohorts of known *TMPRSS2:ERG* fusion positive (or *ERG* overexpression) and negative tumours. Penney *et al.* (2016) identified that six of 39 genome-wide association study PCa risk variants were significantly associated with *ERG* overexpression, in their cohort of 227 *ERG* positive and 260 negative tumours [357]. The most recent and largest study observed a significant difference between fusion positive and negative tumours for rs16901979 (8q24) and rs1859962 (17q24), which were enriched in fusion negative and positive tumours, respectively [358]. Interestingly, *TMPRSS2:ERG* has been identified more frequently in early-onset PCa, suggesting that the event may also be associated with familial PCa and potentially, rare germline variants [359,360]. In fact, Luedeke and colleagues (2009) studied familial and sporadic tumours and found a significant association of

TMPRSS2:ERG fusion-positive PCa with rare variants in *POL1* and *ESCO1*, both of which are DNA repair genes [361]. These findings suggest that tumours that develop the *TMPRSS2:ERG* fusion have a different germline predisposition from those that do not, and these genetic variations may influence fusion event occurrence.

Thus, this study hypothesises that germline variants may predispose some tumours to somatic alterations, such as gene fusions. To explore this theory, tumours from men belonging to a large Tasmanian PCa family, PcTas9 were assayed on the TruSight RNA Fusion Panel (Illumina) to identify gene fusions present in this family. TaqMan® expression assays and RT-qPCR gene expression analysis were then used to determine their frequency in the entire *Tasmanian Prostate Tissue Pathology Resource*. Ultimately, the overall aim was to investigate the relationship between identified fusion events and underlying genetic predisposition.

7.2   METHODS

7.2.1   TruSight RNA Fusion Panel

A total of 14 malignant RNA samples from PcTas9 cases were assayed on the TruSight RNA Fusion Panel, across two separate assays. This technology enables RNA from poor quality formalin-fixed paraffin embedded (FFPE) tumour samples to be assayed for 507 known cancer fusion genes, including *ETS* transcription factors, *ERG* and *ETV1*. Novel fusion partners can also be identified, as only one of the two genes involved in the fusion event must be present on the panel. This is because probes specific to the target RNA region bind appropriately and the fusion break point is sequenced. The TruSight RNA Fusion Capture chemistry is illustrated in Figure 7.2 and the libraries were prepared using 20-100ng of FFPE RNA (depending on RNA quality), as per the manufacturer's instructions. Targeted sequencing with deep coverage was performed on the Illumina MiSeq platform using the MiSeq® V2 300 Cycle Reagent Kit (Illumina), and data were analysed using the RNA-Seq Alignment workflow (Illumina) on BaseSpace. In short, raw fastq files were aligned to the hg19 reference genome using TopHat2, and each reference gene and transcript were given a FPKM (fragments per kilobase million) estimation using Cufflinks 2. Variants were called with the Isaac Variant caller and each fusion call was given a confidence score. This score (out of 1) is based on the FPKM, split read scores, paired read scores, break-end homology and, several other features. A score >0.5 meets all of the threshold filters (PASS) whereas, a score <0.5 is considered a low confidence fusion call (Low Fusion Rate), which may include true positive fusions, but expressed at lower levels.
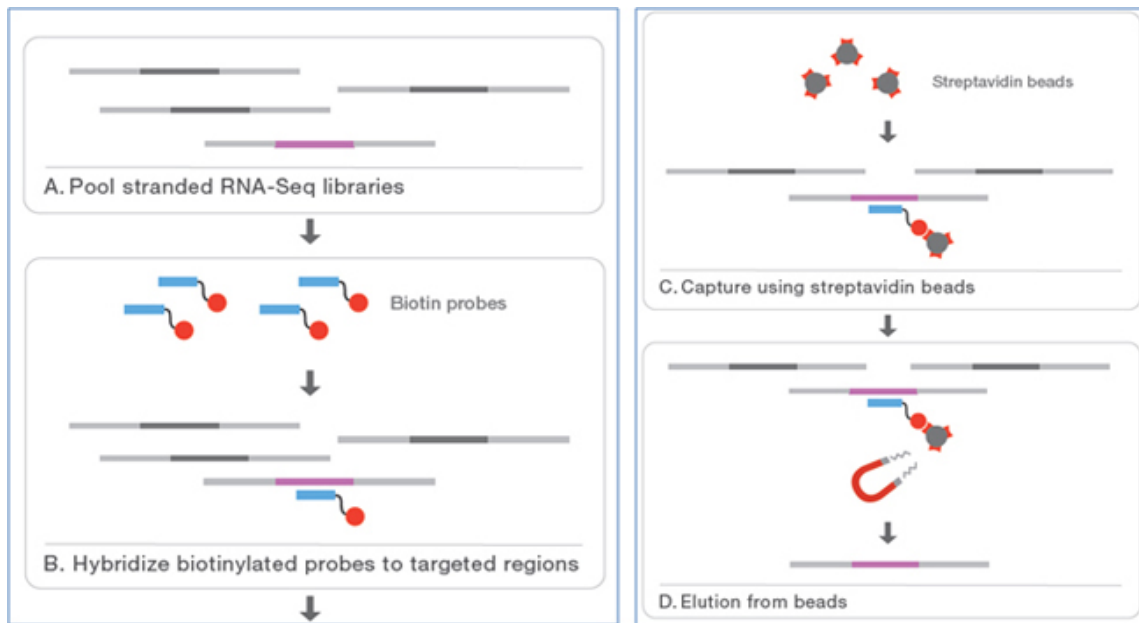
Figure 7.2 TruSight RNA Fusion Capture Chemistry.

The TruSight RNA Fusion Panel provides a simple, streamlined method for isolating targeted regions of interest from total RNA, including from FFPE tumour samples. This figure details the workflow and capture chemistry of the panel. Unique oligonucleotide indexes are added to each individual library. Once the RNA-Seq libraries are pooled they are hybridised to biotin-labelled probes specific for targeted RNA regions. These targets are captured by adding streptavidin beads that bind to the biotinylated probes. Magnetic beads are then used to remove the bound fragments efficiently from solution. Following amplification, the targeted library was clustered generated, followed by targeted sequencing with deep coverage on the Illumina MiSeq platform (Illumina, California, USA, 2019).

7.2.2 TaqMan® *TMPRSS2:ERG* Fusion Assays

In total, 56 *Tasmanian Prostate Tissue Pathology Resource* tumour samples were screened for two isoforms of the *TMPRSS2:ERG* fusion, including *TMPRSS2* (exon 1):*ERG* (exon 2) (T1E2) and *TMPRSS2* (exon 1):*ERG* (exon 4) (T1E4). This was performed using TaqMan® probes designed across the breakpoint of the fusion gene (Life Technologies; Appendix 26). Amplification was performed on 50ng of FFPE cDNA, in duplicate, as per the conditions in Appendix 1. Real-time quantitative (RT-qPCR) thermal cycling was conducted on the QuantStudio™ 3 Real-Time PCR System (Applied Biosystems) and quantification visualised using the QuantStudio™ Design and Analysis Software v1.5. Each qPCR run was conducted with a DNA-free NTC and each sample was run in duplicate for housekeeper, *β-Actin* (Life Technologies; Appendix 26). Samples that appeared to be fusion positive were confirmed by Sanger sequencing. In short, a forward primer was designed in the last included exon of *TMPRSS2* (1 or 2) and a reverse primer in any of the first few included exons of *ERG* (2, 3 or 4)*,* thus sequencing the fusion breakpoint. Sanger sequencing was conducted as previously described (Chapter 2.2.3; Appendix 27).

7.2.3 Quantification of *ETV1* gene expression

ETV1 (ENST00000405358.4) gene expression in prostate tissue samples was assessed by RT-qPCR analysis. Expression was normalised to the expression of two housekeeping genes, as discussed in Chapter 2.3. Briefly, three different regions of *ETV1* were amplified, including a region before the fusion breakpoint (exon 8/10) and two after (exon 16/17 and 21/22). RT-qPCR primers were designed to the most commonly transcribed isoform in the prostate (as per GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2; https://gtexportal.org/home/)) [137] and are displayed in Appendix 3. Absolute gene expression was compared between tumours from PcTas9 and non-PcTas9 cases. The non-PcTas9 patient group comprised DVA sporadic tumours from the *Tasmanian Prostate Cancer Case-Control Study* and other familial tumours from the *Tasmanian Familial Prostate Cancer Cohort*.

7.2.4 Quantification of ETV1 protein expression

Immunohistochemistry (IHC) was performed to quantify ETV1 protein expression in the prostate tissue samples, as previously described (Chapter 2.4; Appendix 5). Cytospins of HEK293 cells, and sections of colon and skin were used as positive controls. Negative controls included primary antibody only, secondary antibody only, and a mouse IgG$_1$ isotype control

(Dako). ETV1 protein expression was compared between tumours from PcTas9 and non-PcTas9 cases.

7.3 RESULTS

7.3.1 Gene fusion analysis of PcTas9 prostate tumour samples

In total, 26 PcTas9 and 30 non-PcTas9 FFPE prostate tissue samples were obtained for this study (described in Chapter 6.3.1). To investigate the prevalence of gene fusion events in the *Tasmanian Prostate Tissue Pathology Resource*, 14 PcTas9 tumour RNA samples were assayed on the TruSight RNA Fusion Panel (Table 7.2; Figure 7.3). Where sufficient RNA was available, one affected man from each branch of the family was selected for analysis.

Table 7.2 PcTas9 tumour samples chosen for the RNA Fusion Panel, including clinicopathological characteristics.

Sample Identification	Age at Diagnosis	Tissue Source	Tumour Grade[1]	Contemporary Gleason Score[2]
PC9-07	71	TURP	PD	9 (5+4)
PC9-12	66	RP	MD	6 (3+3)
PC9-13	83	TURP	-	9 (4+5)
PC9-14	79	TURP	MD	6 (3+3)
PC9-15	64	TURP	MD	5 (2+3)
PC9-20	76	TURP	PD	9 (4+5)
PC9-158	63	RP	-	6 (3+3)
PC9-211	68	TURP	PD	9 (4+5)
PC9-477	55	RP	-	6 (3+3)
PC9-588	63	RP	MD	6 (3+3)
PC9-603	73	RP	MD	6 (3+3)
PC9-627	65	RP	-	7 (3+4)
PC9-645	60	RP	-	7 (3+4)
PC9-659	65	RP	PD	9 (4+5)
RP: Radical prostatectomy; TURP: Transurethral resection of the prostate; [1]Tumour grade obtained from pathology report; [2]Contemporary Gleason Score from FFPE tissue block chosen for macrodissection of nucleic acids and IHC; MD: moderately differentiated; PD: poorly differentiated; -: information not present in original pathology report.				

Figure 7.3 A condensed PcTas9 pedigree showing tumours chosen for analysis on the TruSight RNA Fusion Panel.

This condensed version of PcTas9 indicates those PCa cases with available prostate tumour specimens (shown by red arrows) and their relationship. Tumours chosen for the RNA fusion panel are shown in yellow.

7.3.2 Identification of gene fusion events in tumours from PcTas9 men

Nine tumours showed evidence of one or more fusion event (Table 7.3). Notably, one tumour was shown to harbor three fusion events, including a known *TMPRSS2:ERG* fusion and two novel events involving the *ETS* fusion genes, *ETV4* and *FOXP1*. In total, four novel fusion genes in three different tumours were identified; *WHSC1L1:CNKSR3, SLC30A4:ETV1, C19orf48:ETV4* and *RYBP:FOXP1* (Table 7.3). The *WHSC1L1:CNKSR3* and *RYBP:FOXP1* fusions were considered low confidence fusion calls, however this result may indicate low expression of the fusion gene. In terms of known fusion events, one tumour was identified as *TMPRSS2:ETV1* positive and seven were *TMPRSS2:ERG* positive. Of the *TMPRSS2:ERG* fusion events, six involved exon 1 of *TMPRSS2* fused to exon 4 of *ERG* (T1E4), whereas one involved the fusion of exon 1 of *TMPRSS2* to exon 2 of *ERG* (T1E2). The presence of the *TMPRSS2:ERG* fusion transcripts at a frequency close to 50% was consistent with the literature [315], which suggested that the assay was working optimally (Table 7.3).

Table 7.3 Gene fusion events identified in the PcTas9 prostate tumour samples.

PcTas9 Identification	Gene 1 (Breakpoint)	Gene 1 Exon	Gene 2 (Breakpoint)	Gene 2 Exon	Score[1]	Filter[2]	Gene Fusion
PC9-07	No fusion detected						
PC9-12	*TMPRSS2* (chr21:42,880,007)	1	*ERG* (chr21:39,817,543)	4	0.942	PASS	*TMPRSS2:ERG*
	TMPRSS2 (chr21:42,880,007)	1	*ERG* (chr21:39,795,482)	5	0.562	Low Fusion Rate	*TMPRSS2:ERG*
PC9-13	*WHSC1L1* (chr8:38,205,113)	2	*CNKSR3* (chr6:154,762,378)	4	0.426	Low Fusion Rate	[#]*WHSC1L1:CNKSR3*[#] *
PC9-14	*TMPRSS2* (chr21:42,880,007)	1	*ERG* (chr21:39,817,543)	4	0.464	PASS	*TMPRSS2:ERG*
PC9-15	No fusion detected						
PC9-20	No fusion detected						
PC9-158	*SLC30A4* (chr15:45,803,402)	3	*ETV1* (chr7:13,978,871)	15	0.747	PASS	[#]*SLC30A4:ETV1* *
PC9-211	No fusion detected						
PC9-477	*TMPRSS2* (chr21:42,870,045)	2	*ERG* (chr21:39,817,543)	4	0.519	Low Fusion Rate	*TMPRSS2:ERG*
PC9-588	*TMPRSS2* (chr21:42,880,007)	1	*ERG* (chr21:39,817,543)	4	0.858	PASS	*TMPRSS2:ERG*
PC9-603	*TMPRSS2* (chr21:42,880,008)	1	*ETV1* (chr7:13,978,871)	3	0.474	Low Fusion Rate	*TMPRSS2:ETV1*
PC9-627	*TMPRSS2* (chr21:42,880,006)	1	*ERG* (chr21:39,956,867)	2	0.844	PASS	*TMPRSS2:ERG*
	TMPRSS2 (chr21:42,880,007)	1	*ERG* (chr21:39,817,543)	4	0.778	PASS	*TMPRSS2:ERG*
	C19orf48 (chr19:51,305,474)	3	*ETV4* (chr17:41,613,847)	4	0.616	PASS	[#]*C19orf48:ETV4* *
	RYBP (chr3:72,495,646)	1	*FOXP1* (chr3:71,090,682)	5	0.397	Low Fusion Rate	[#]*RYBP:FOXP1* *
PC9-645	*TMPRSS2* (chr21:42,880,007)	1	*ERG* (chr21:39,817,543)	4	0.911	PASS	*TMPRSS2:ERG*
PC9-659	No fusion detected						

[1]The confidence score (out of 1) is based on the FPKM, split read scores, paired read scores, break-end homology, and several other features. A score >0.5 meets all of the threshold filters (PASS[2]) whereas, a score <0.5 is considered a low confidence fusion call (Low Fusion Rate[2]), which may include true positive fusions, but expressed at lower levels; *Novel gene fusion; [#]Novel fusion partner.

7.3.3 The frequency of two *TMPRSS2:ERG* fusion transcripts in the *Tasmanian Prostate Tissue Pathology Resource*

Following the identification of two *TMPRSS2:ERG* fusion transcripts (T1E2 and T1E4) in PcTas9 tumours, the overall frequency in the *Tasmanian Prostate Tissue Pathology Resource* was determined. In total, 46 prostate tumours from 15 PcTas families, as well as eight sporadic cases (DVA) were screened for T1E2 and T1E4 by RT-qPCR. Overall, 17 tumours were observed to be *TMPRSS2:ERG* fusion positive (31.5%; Table 7.4). Five families were identified to have at least one case with a fusion positive tumour, four of which had two or more cases. Tumours from PcTas9 made up 33% of the available samples and had the highest number of fusion positive tumours, with ten out of 18 tumours fusion positive (56%; Figure 7.4). However, PcTas2 had the highest proportion of positive tumours (60%; Figure 7.5). Notably, the two *TMPRSS2:ERG* fusion transcripts were not detected in any of the eight sporadic cases.

Table 7.4 The total number of prostate tumours positive for *TMPRSS2:ERG*.

Family Identification	Number of PCa cases with tumour FFPE RNA	Number of *TMPRSS2:ERG* positive tumours
DVA Sporadic Cases	8	0
PcTas2	5	3 (60%)
PcTas3	2	0
PcTas4	1	0
PcTas9	18	10 (56%)
PcTas11	2	0
PcTas12	7	2 (29%)
PcTas19	1	0
PcTas22	2	0
PcTas23	1	0
PcTas31	1	0
PcTas60	1	0
PcTas72	2	1 (50%)
PcTas213	1	0
PcTas837	1	1 (100%)
PcTas3250	1	0
Entire Resource	**54**	**17 (31.5%)**
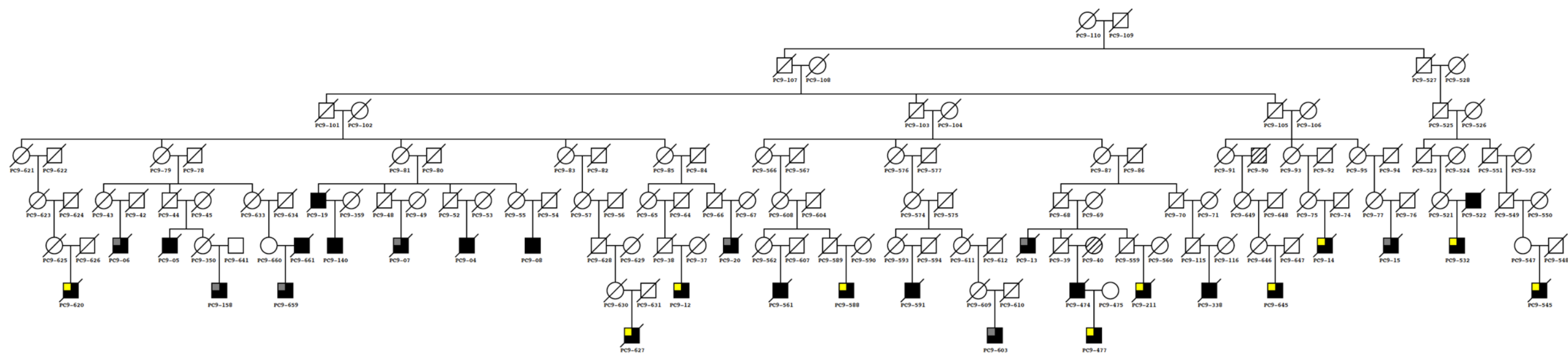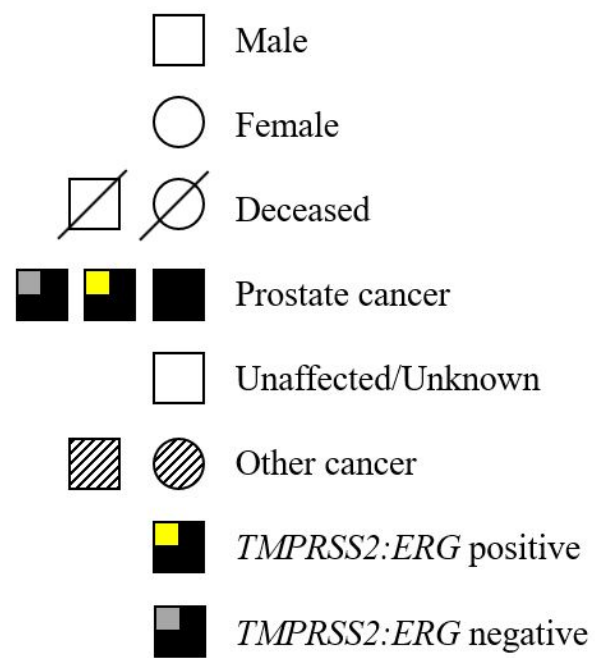
Figure 7.4 A condensed PcTas9 pedigree showing *TMPRSS2:ERG* fusion status.

This condensed PcTas9 pedigree indicates those tumours assessed for the two *TMPRSS2:ERG* fusion events; fusion positive are shown in yellow and fusion negative, in grey.
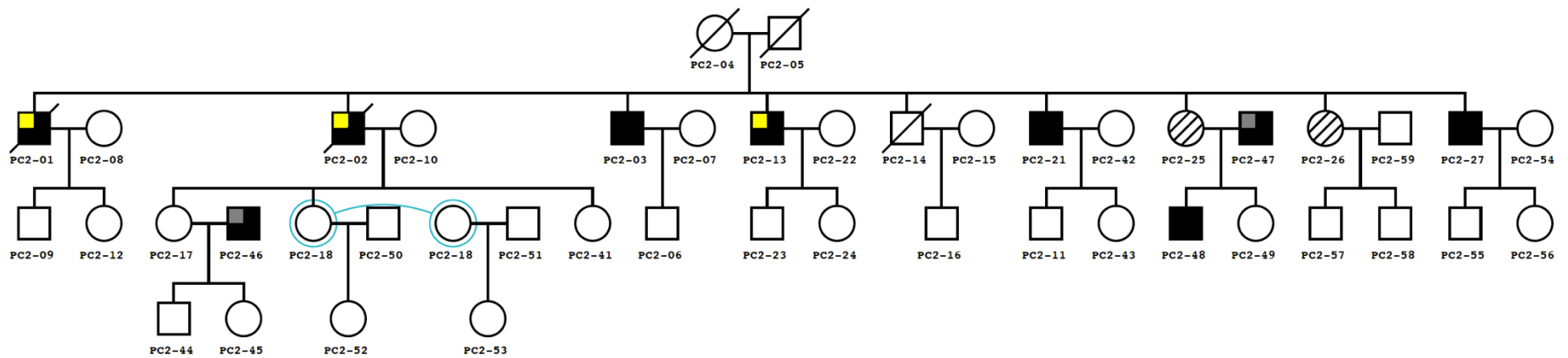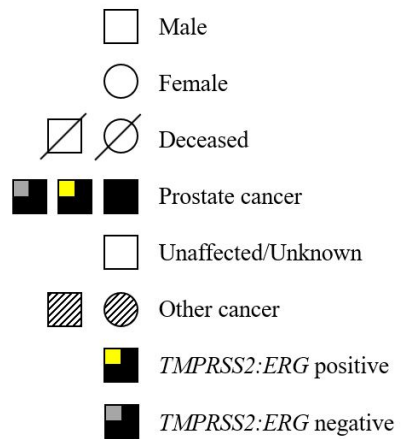
Figure 7.5 The PcTas2 pedigree showing *TMPRSS2:ERG* fusion status.

This condensed PcTas2 pedigree indicates those tumours assessed for the two *TMPRSS2:ERG* fusion events; fusion positive are shown in yellow and fusion negative, in grey.

7.3.4 Association of *TMPRSS2:ERG* with clinical characteristics and tumour pathology

The correlation between *TMPRSS2:ERG* fusion status and certain clinical characteristics, such as age at diagnosis, Gleason score (GS), age at death and cause of death was examined (Table 7.5 and Appendix 28). There was no difference in the age at diagnosis between *TMPRSS2:ERG* fusion positive (n=17) and negative tumours (n=37; p=0.91). Age at death was slightly younger for those with fusion positive compared to negative tumours however, this was not statistically significant (p=0.35). There was also no difference in GS (≤7 (3+4) *versus* ≥7 (4+3)) between tumours with *TMPRSS2:ERG* fusion positive (n=15) and negative status (n=33; p=0.78), nor cause of death (p=0.50; PCa *versus* non-cancer) between the two groups.

Table 7.5 Clinicopathological characteristics of *TMPRSS2:ERG* fusion positive tumours.

Sample Identification	*TMPRSS2:ERG* Transcript	Age at Diagnosis	Tumour Grade[1]	Contemporary Gleason Score[2]	Age at Death[3]	Cause of Death[3]
PC2-01	T1E4	62	PD	10 (5+5)	64	PCa
PC2-02	T1E2	53	-	5 (3+2)	75	Other
PC2-13	T1E4	54	-	4 (2+2)		
PC9-12	T1E4	66	MD	6 (3+3)		
PC9-14	T1E4	79	MD	6 (3+3)	82	Non-Cancer
PC9-211	T1E4	68	PD	9 (4+5)	70	PCa
PC9-477	T1E4	55	-	6 (3+3)		
PC9-532	T1E4	70	-	6 (3+3)		
PC9-545	T1E4	55	PD	-	55	PCa
PC9-588	T1E4	63	MD	6 (3+3)		
PC9-620	T1E4	71	PD	9 (4+5)	84	Non-Cancer
PC9-627	T1E2 & T1E4	65	-	7 (3+4)	68	Non-Cancer
PC9-645	T1E4	60	-	7 (3+4)		

[1]Tumour grade obtained from pathology report; [2]Contemporary Gleason Score from pathology report (if known) or FFPE tissue block chosen for microdissection of nucleic acids; MD: moderately differentiated; PD: poorly differentiated; -: information not present in original pathology report; [3]Age at death and cause of death information was obtained from the Tasmanian Cancer Registry (as at April 2019); PCa: Prostate Cancer; Other: Other cancer; *Clinical characteristics of *TMPRSS2:ERG* fusion negative tumours can be found in Appendix 28.

Sample Identification	*TMPRSS2:ERG* Transcript	Age at Diagnosis	Tumour Grade[1]	Contemporary Gleason Score[2]	Age at Death[3]	Cause of Death[3]
PC12-01	T1E4	63	MD	6 (3+3)	73	Non-Cancer
PC12-06	T1E4	80	-	7 (3+4)	84	Non-Cancer
PC72-04	T1E4	70	PD	9 (4+5)	82	PCa
PC837-04	T1E4	59	-	9 (4+5)		

[1]Tumour grade obtained from pathology report; [2]Contemporary Gleason Score from pathology report (if known) or FFPE tissue block chosen for microdissection of nucleic acids; MD: moderately differentiated; PD: poorly differentia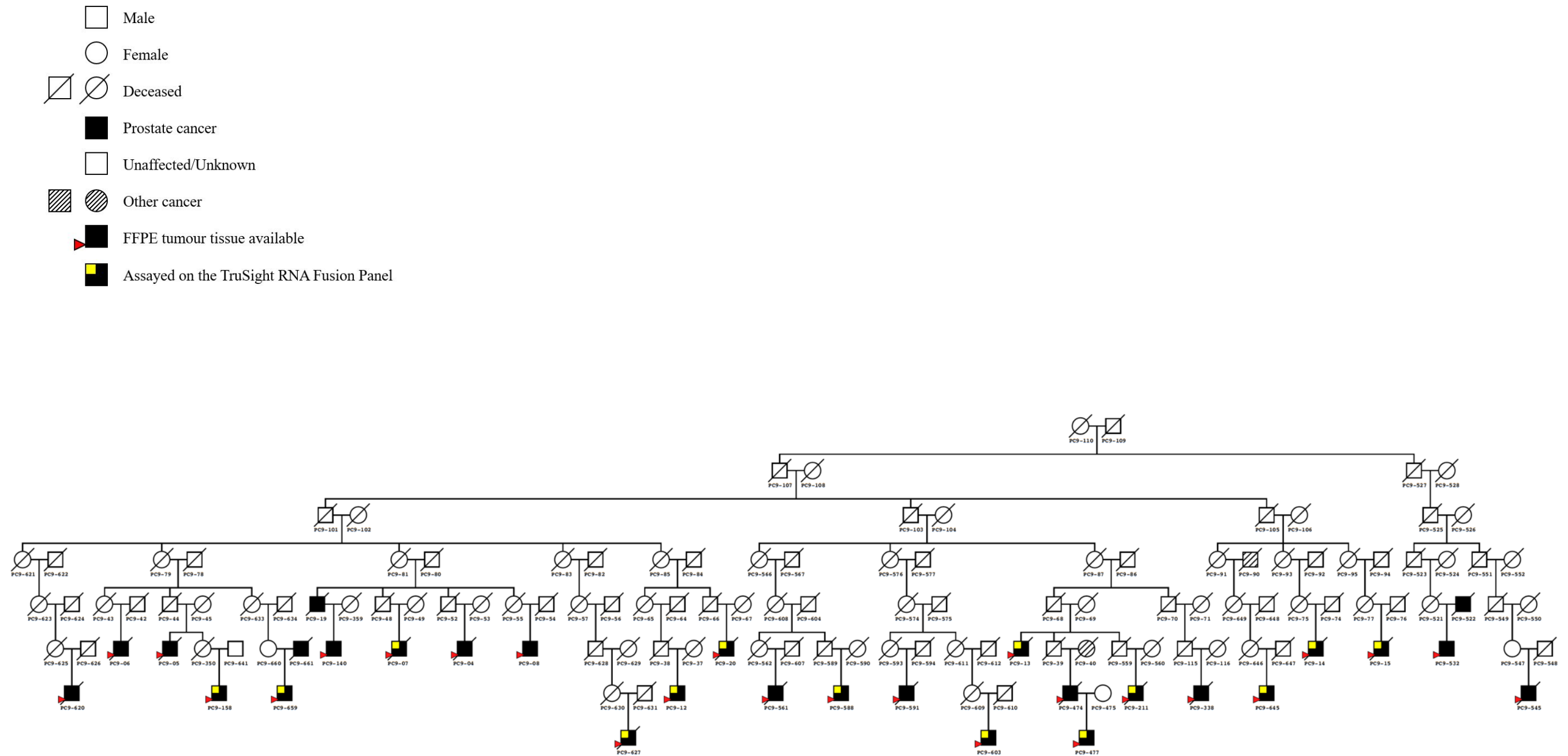ted; -: information not present in original pathology report; [3]Age at death and cause of death information was obtained from the Tasmanian Cancer Registry (as at April 2019); PCa: Prostate Cancer; Other: Other cancer; *Clinical characteristics of *TMPRSS2:ERG* fusion negative tumours can be found in Appendix 28.

7.3.5   The effect of *ETV1* fusion events on *ETV1* gene expression

Prostate tumours from two PcTas9 men were found to have an *ETV1* gene fusion, each with a different isoform. The tumour from individual PC9-158 had a *SLC30A4:ETV1* gene fusion, with *SLC30A4* identified as a novel 5' fusion partner. The known *TMPRSS2:ETV1* gene fusion was identified in PC9-603 at a low fusion rate, and it had the same *ETV1* breakpoint as PC9-158 (chr7:13,978,871). Prior literature suggests that *ETV1* fusions are fairly rare (~7%) and involve multiple 5' fusion partners [343], which suggests that targeted detection of these two fusions using TaqMan would likely uncover very few, if any additional carriers, plus miss other fusion events involving *ETV1*. Therefore, given that *ETS*-gene fusions often result in the overexpression of the *ETS* gene [341] and RT-qPCR is a more cost effective method for fusion detection, it was decided that *ETV1* gene expression would be examined to detect additional *ETV1* fusions in the *Tasmanian Prostate Tissue Pathology Resource*. This method would also determine the effect of the two already identified *ETV1* fusion events on *ETV1* expression.

RNA was extracted from malignant glands (n=28) and RT-qPCR was undertaken to determine the absolute expression of *ETV1* in three regions of the gene (exon 8/10, 16/17 and 21/22; Appendix 29). There was a borderline significant difference in expression of *ETV1* exon 8/10 between the malignant glands of PcTas9 tumours (n=17) compared to non-PcTas9 tumours (n=11; p=0.05; Figure 7.6). PcTas9 tumours had an overall lower level of *ETV1* exon 8/10 expression, however expression was generally very low for this amplified region. Across the remaining two *ETV1* regions, exon 16/17 and 21/22, no difference in expression was observed

between the malignant glands of PcTas9 and non-PcTas9 tumours (p=0.54 and 0.72, respectively; Figure 7.6). In benign prostate glands, there was no significant difference in *ETV1* expression across the three regions in PcTas9 (n=7) *versus* non-PcTas9 samples (n=8; p=0.15, 0.07 and 0.27, respectively).
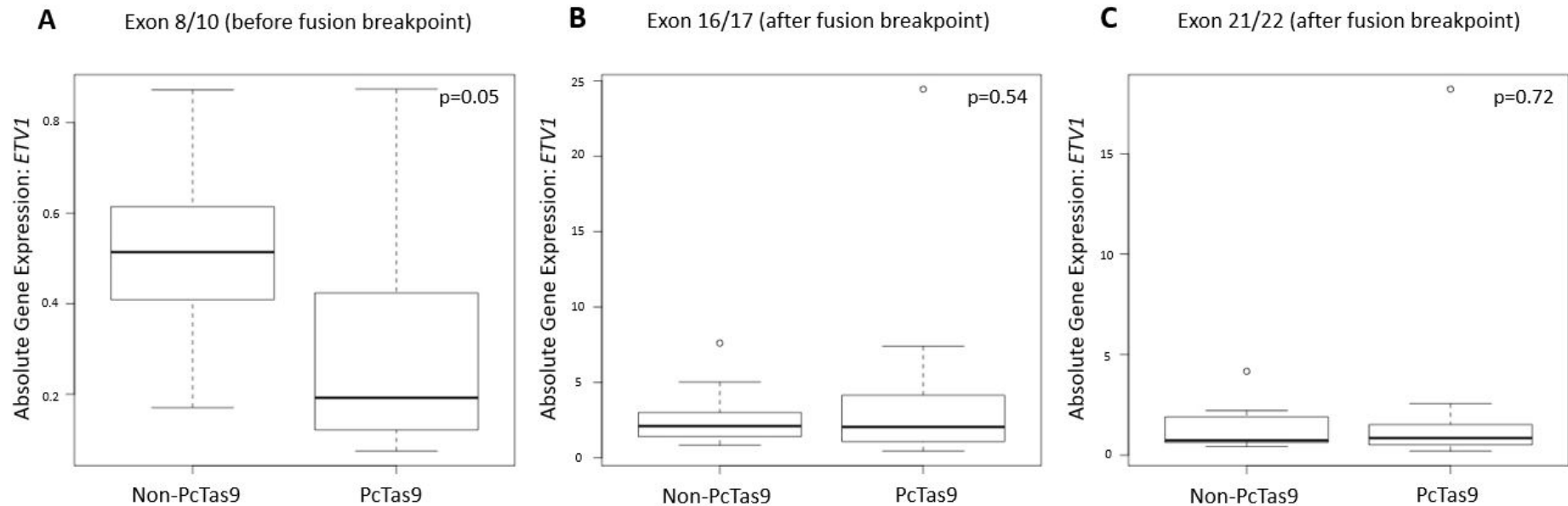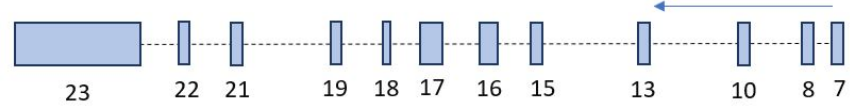
Figure 7.6 *ETV1* gene expression analysis in malignant prostate glands from non-PcTas9 cases compared to PcTas9 cases.

ETV1 expression in three different regions of the gene was assessed in malignant prostate glands from two patient groups, non-PcTas9 (comprising sporadic and familial tumours; n=11) and PcTas9 (n=17). A schematic of the most commonly transcribed isoform of *ETV1* in the prostate is shown at the top of the figure. Absolute *ETV1* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. *ETV1* expression in malignant glands from non-PcTas9 and PcTas9 tumours in each region was compared using an unpaired Student's t-test. The spread of the data is represented by a box and whisker plot. Median expression is shown by the thick black line, the interquartile range (middle 50% of data set) is represented by the box and the minimum and maximum values by the whiskers (dotted lines). Individual outliers are shown by dots.

The expression level of *ETV1* was then examined in the two fusion positive tumours compared to other tumours in the PcTas9 family (Figure 7.7). Individual PC9-158 had significantly increased *ETV1* expression in the two regions after the *SLC30A4:ETV1* fusion breakpoint (exon 16/17 and exon 21/22) compared to other PcTas9 tumours. Notably, the benign glands of PC9-158 had a low *ETV1* expression profile across all assessed regions, suggesting that increased expression is an anomaly unique to the malignant glands. PC9-603, who harbours a low fusion rate of the *TMPRSS2:ETV1* fusion, had a similar *ETV1* expression pattern to the other PcTas9 tumours.
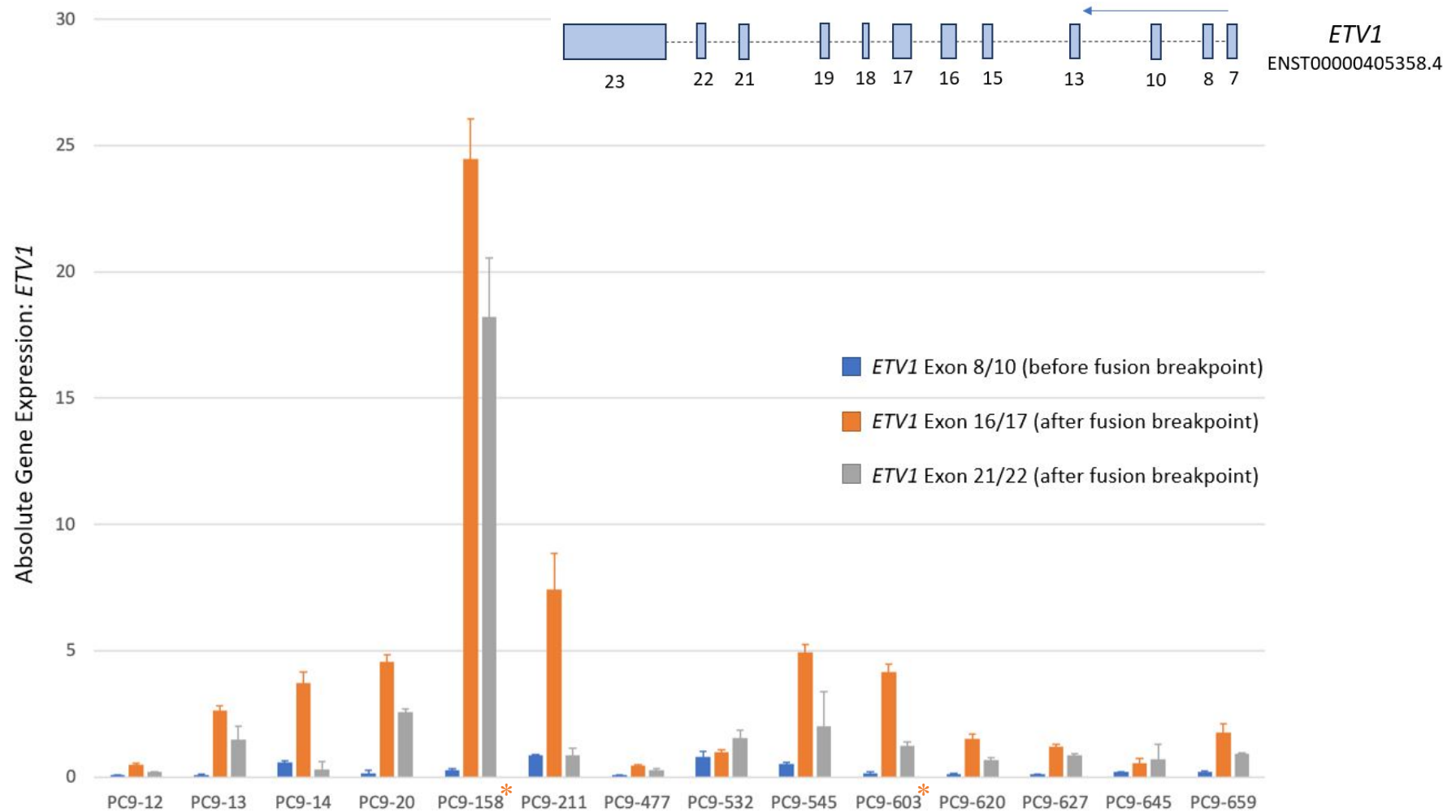
Figure 7.7 *ETV1* gene expression analysis in malignant prostate glands from PcTas9 tumours.

ETV1 expression in three different regions of the gene was assessed in prostate tumours from PcTas9 cases. A schematic of the most commonly transcribed isoform of *ETV1* in the prostate is shown at the top of the figure. Absolute *ETV1* gene expression was calculated for each sample by normalising to the expression of two housekeeping genes. Individual PcTas9 malignant gland expression is shown here, with regions of *ETV1* depicted by different colours. * *ETV1* fusion positive as idnetfiied by the RNA Fusion Panel.

Gasi and colleagues (2011) reported that a high ratio between *ETV1* expression at the 3' end (after the fusion breakpoint) *versus* the 5'end (before the fusion breakpoint) was indicative of a fusion transcript whereas, a ratio of 1:1 indicated expression of full-length *ETV1* [362]. Here, ratios of *ETV1* gene expression between exon 16/17:exon 8/10, exon 21/22:exon 8/10 and exon 21/22:exon 16:17 were determined in all prostate tumours, where data was available (Table 7.6). A high *ETV1* expression ratio between the 5' and 3' end was observed in PC9-158, and was also apparent in three other PcTas9 tumours, PC9-13, PC9-20 and PC9-603, however the ratios were not as high as PC9-158. Notably, PC9-603, the low rate fusion carrier, had a high *ETV1* exon 16/17:exon 8/10 ratio, comparable to the other two samples, yet previous data from the RNA Fusion Panel indicated that neither of these tumours had an *ETV1* fusion event (Table 7.3). Therefore, it is possible that PC9-13 and PC9-20 carry *ETV1* fusions but at a level too low to be detected by the fusion panel.

Table 7.6 Ratios of *ETV1* gene expression in regions before and after the fusion breakpoint.

Sample Identification	ETV1 Exon 16/17:8/10 Ratio	ETV1 Exon 21/22:8/10 Ratio	ETV1 Exon 21/22:16/17 Ratio
PC9-158	87.36	65.04	0.74
PC9-20	35.15	19.78	0.56
PC9-13	29.22	16.56	0.57
PC9-603	29.71	8.71	0.29
DVA 220	3.92	2.54	0.65
PC9-12	7.00	2.71	0.39
PC9-14	6.32	0.53	0.08
PC9-211	8.53	0.99	0.12
PC9-477	5.50	3.50	0.64
PC9-532	1.27	1.96	1.55
PC9-545	9.32	3.79	0.41
PC9-620	11.69	5.08	0.43
PC9-627	11.00	7.73	0.70
PC9-645	2.67	3.29	1.23
PC9-659	9.32	1.95	0.21
PC11-11	2.66	4.97	1.87
PC12-06	2.78	5.49	1.98
PC12-07	2.65	1.43	0.54
PC12-09	2.55	0.72	0.28
PC3250-01	2.85	1.32	0.46
High 'Exon 16/17:Exon 8/10' and 'Exon 21/22:Exon 8/10' ratios are indicative of a fusion gene. It is expected that the 'Exon 21/22:Exon 16/17 ratio will be ≤1.00.			

7.3.6   The effect of *ETV1* fusion events on ETV1 protein expression

An *ETV1* gene fusion can result in overexpression of a truncated or a full-length ETV1 protein [315,346]. IHC was undertaken on 56 FFPE prostate tumour samples to determine whether the *SLC30A4:ETV1* and *TMPRSS2:ETV1* fusions cause overexpression of the ETV1 protein. In addition, this assay could potentially identify additional *ETV1* fusion events in the wider *Tasmanian Prostate Tissue Pathology Resource*. ETV1 staining intensity ranged from negative (0) to moderate (2) across the dataset, and the percentage of ETV1 positive nuclei ranged from approximately 5-70% (Appendix 29). In total, only 24% (n=16) of the tissue samples were positive for the ETV1 protein; five samples had expression in benign glands only, nine in malignant glands only and two had expression in both benign and malignant glands (Appendix 29). There was no significant difference in ETV1 expression between paired malignant and benign glands (p=0.49). Notably, 17 samples were found to have weak-moderate staining of ETV1 in infiltrating inflammatory cells and nine of these did not express ETV1 in adjacent prostate glands (Figure 7.8). PC9-158, the *SLC30A4:ETV1* fusion positive tumour, had moderate expression of ETV1 in benign glands, but not in malignant glands. PC9-603, the carrier of the *TMPRSS2:ETV1* fusion, did not express ETV1 in either type of prostate gland. The two tumours with similar 5':3' *ETV1* ratios similar to PC9-603, did not express ETV1 and had moderate expression of ETV1 in their malignant glands, respectively.
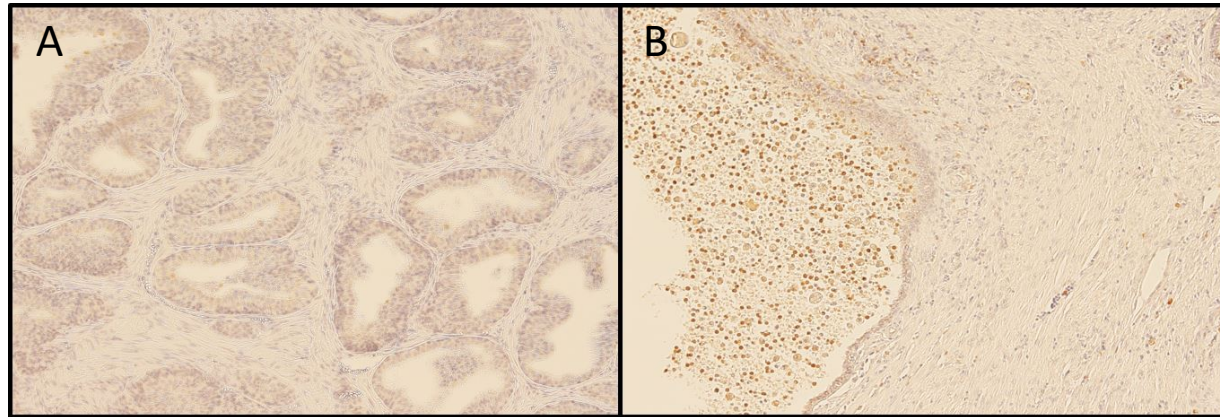
Figure 7.8 ETV1 protein expression in FFPE prostate tumour samples.

ETV1 protein expression was assessed in 56 prostate tumour specimens from the *Tasmanian Prostate Tissue Pathology Resource* to determine whether the *SLC30A4:ETV1* and *TMPRSS2:ETV1* fusions result in overexpression of the ETV1 protein. In short, IHC using an antibody targeting the 'middle region' of the ETV1 protein was utilised to assess protein expression. Staining intensity was scored as weak, moderate or strong. **A)** Weak staining of ETV1 in the nuclei of the prostate gland cells. **B)** Weak-moderate staining of ETV1 in inflammatory cells; identified in 26% of tissue samples. Images were taken with an Olympus BX53 microscope, using the DP73 camera and software (x100).

7.4 DISCUSSION

7.4.1 Overall findings

Here, 13 fusion events were observed in nine PcTas9 prostate tumours, including multiple events involving three known *ETS*-fusion transcripts and four novel fusion events. The novel events included two previously unobserved gene fusions, *WHSC1L1:CNKSR3* and *RYBP:FOXP1,* and two fusions involving novel partners of *ETV1* and *ETV4*; *SLC30A4* and *C19orf48,* respectively. Overall, the *TMPRSS2:ERG* fusion was identified in 31.5% of our prostate tumours and was more common in two Tasmanian families, PcTas2 and PcTas9.

7.4.2 *TMPRSS2:ERG* fusion events in Tasmanian prostate tumours

Fusion of the androgen-regulated promoter region of *TMPRSS2* with *ERG* is the most common *ETS* rearrangement in prostate tumours. Both of these genes lie within the 21q22.2 chromosomal region, which is a hot spot for rearrangement, thus multiple fusion transcripts have been identified in prostate tumours [344,345,363,364]. The transcripts identified in this study (T1E2 and T1E4) are likely to be caused by a 2.7Mb interstitial deletion or translocation of chromosome 21q22.2. The *TMPRSS2* gene encodes an androgen regulated, type II transmembrane-bound serine protease that is highly expressed in normal prostate tissue [365,366]. Normally *ERG* is lowly expressed in the prostate (as per the GTEx Portal; https://gtexportal.org/home/) [137], however a breakthrough study discovered that *ERG* was overexpressed in approximately 55.2% of prostate tumours and 20% of high-grade prostatic intraepithelial neoplasia lesions [315]. It was suggested that in the majority of tumours, *ERG* overexpression was driven by the fusion of *TMPRSS2* to *ERG*. Here, 31.5% of tumours in the *Tasmanian Prostate Tissue Pathology Resource* were *TMPRSS2:ERG* positive. The frequency of *ERG* fusions in our study is somewhat lower than initially reported by Tomlins *et al.* (2005), however subsequent studies have found similar frequencies to that reported here [367,368]. Tomlins and colleagues (2005) discussed that the frequency of *ETS* gene fusions in their study might have been overestimated, as fluorescence *in-situ* hybridisation assay (FISH) can also detect other *ERG* rearrangements [315].

The association between clinical characteristics and *TMPRSS2:ERG* fusion status has been well assessed in the literature, but with conflicting results. Some studies have demonstrated that *TMPRSS2:ERG* fusions are associated with an increased risk of more advanced and invasive PCa tumours with poor prognoses [341,369]. Demichelis and colleagues (2007) identified

a statistically significant association with fusion status and PCa specific death (cumulative incidence ratio 2.7, p<0.01, 95% CI=1.3-5.8) [350], suggesting that *TMPRSS2:ERG* may be used as a diagnostic and prognostic indicator of aggressive PCa in parallel with GS and prostate-specific antigen (PSA) level. However, other studies have found contrary results, in which fusion positive tumours were not associated with stage, GS, PSA-induced recurrence, progression, prognosis and/or disease aggressiveness. FitzGerald and colleagues (2008) found that *TMPRRS2:ERG* fusion positive tumours did not exhibit reduced PCa survival (hazard ratio=0.92; 95% CI=0.22-3.93) [352] and Gopalan *et al.* (2009) also found no difference in overall survival between the two subtypes [353]. Likewise, in our study, no difference in the age at diagnosis, GS, age at death or cause of death between fusion positive and negative cases was identified. Though, this result may be due to a small sample size, which may have limited the probability of finding an association.

The *TMPRSS2:ERG* fusion was identified in tumours from five Tasmanian families, two of which had more than 50% of assayed tumours with a fusion event. PcTas2 had the highest percentage of fusion events, and notably, the three positive tumours comprise an affected brother trio, and the negative tumours were from two unrelated, married-in cases (Figure 7.5). PcTas9 comprises the largest collection of FFPE prostate samples from a single family in the *Tasmanian Prostate Tissue Pathology Resource*, and had the highest number of fusion events. We have been able to determine fusion status from cases across the whole pedigree and have found fusion positive cases in several branches (Figure 7.4). However, as we were unable to source tumour tissue for every case, we were unable to determine clustering at a level of first, second or third-degree relatedness like PcTas2. Our hypothesis that there is an underlying genetic predisposition to the *TMPRSS2:ERG* fusion is supported by the presence of the fusion in three affected PcTas2 brothers and, further, by the fact that none of the sporadic PCa tumours were fusion positive. However, as discussed above this hypothesis wasn't able to be explored further due to a lack of germline genetic information for these families, as discussed further below (Chapter 7.5) and in Chapter 8.4.

7.4.3 *ETV1* fusion events in Tasmanian prostate tumours

Two *ETV1* fusion events (one novel) in two PcTas9 tumours were identified in our cohort. As far as we know, this is the first study to investigate the presence and prevalence of *ETV1* gene fusions in a familial tissue resource. *ETV1* is the second most common *ETS* gene involved in gene fusions, but unlike *ERG*, at least 10 different fusion partners have been identified to date

[315,346]. Notably, it is well documented that several as yet unidentified 5' fusion partners of *ETV1* exist [370] and this study has identified an additional one, *SLC30A4*. Interestingly, this gene is highly expressed in the prostate and throughout carcinogenesis, as documented in The Cancer Genome Atlas (TCGA) Research Network: https://www.cancer.gov/tcga. Other *SLC* genes, such as *SLC45A3* are commonly involved in gene fusion events, including PCa fusions [343]. *SLC45A3* is most commonly fused to *ERG* or *ETV1* in prostate tumours, displays similar tissue specificity as *TMPRSS2* and can induce androgens [370]. It is possible that *SLC30A4* has a similar role to *SLC45A3* and causes *ETV1* overexpression [346].

A study by Gasi and colleagues (2011) demonstrated that a high *ETV1* exon 11/12 to exon 1/4 ratio is indicative of a fusion event involving this gene, whereas a ratio of 1:1 indicated expression of a full-length *ETV1* mRNA [362]. A subsequent study also analysed exon-level expression of *ETV1* and identified four samples with differential expression between the 5' and 3' end, pinpointing fusion breakpoints before exons 4, 7 and 8, respectively [370]. These studies showed that exon-level expression analysis can be utilised to assess fusion status when one fusion partner is known. In our study, exon-level expression analysis of the *ETV1* gene revealed that PC9-158 had increased *ETV1* expression in the exon 16/17 and 21/22 regions compared to the exon 8/10 region. This tumour was observed to be *SLC30A4:ETV1* fusion positive on the RNA Fusion Panel. Unlike *ERG*, a characteristic of *ETV1* is that it can also be overexpressed in PCa as a full-length wild-type transcript, occurring in approximately half of the tumours assessed by Hermans and colleagues (2008) [346]. This study questioned whether overexpression of full-length *ETV1* is the result of genomic rearrangement of the complete *ETV1* locus [346]. However, gene fusion events can also change the amino acids at the N-terminus of ETV1, or result in N-terminal truncation [315,371]. Unfortunately, our study was unable to detect whether the fusion event/s resulted in overexpression of truncated or wild-type, full-length ETV1. In fact, in most instances, ETV1 protein expression was negative in our prostate tumour samples. Overall, this begs to question whether the ETV1 antibody used in this study is suitable for the detection of fusion events.

7.4.4   The identification of multiple *ETS* gene fusions in a single prostate tumour

The initial observation by Tomlins and colleagues (2005) that *ETS* rearrangements are mutually exclusive [315], was evident in the majority of tumours assessed with the RNA Fusion Panel in our study. However, one tumour was shown to harbor four different *ETS* gene fusions, including two transcripts of the *TMPRSS2:ERG* fusion and two novel fusion events,

C19orf48:ETV4 and *RYBP:FOXP1*. This finding confirms results presented by Clark *et al.* (2008), where *ERG* and *ETV1* fusion events were identified in two separate foci within the same tumour, indicating that *ETS* gene alterations can arise independently [372]. Another study examining *ETS* rearrangements, including *ERG, ETV1* and *ETV5* rearrangements in multifocal PCa, observed multiple *ETS* or 5' fusion partner rearrangements within one prostate gland, even occurring within the same nucleus [373].

The combination of fusion events may be biologically relevant. A study by Kluth and colleagues (2018) found that a deletion of chromosome 3p13 was twice as likely to occur in *TMPRSS2:ERG* fusion positive than negative tumours [374]. Notably, the individual in our study with multiple fusion events, is both *RYBP:FOXP1* (both genes are located on chr3p13) and *TMPRRS2:ERG* fusion positive. Whilst this study and those in the literature indicate that multiple *ETS* gene fusions can occur in a single prostate tumour [372,373], further investigations to determine the biological implications of this is important.

7.4.5   Clinical significance of this study

While we are still determining what causes *ETS* gene fusions at the tumour level, the most significant implication of these events is that they may provide novel therapeutic options. Recently, poly ADP-ribose polymerase (PARP) inhibitors have emerged as promising therapeutic candidates that target *ERG*. Just like *BRCA1/2* mutated tumours, *ETS* positive tumours are susceptible to PARP inhibition through the increased incidence of DNA double strand breaks [375]. The PARP1 inhibitor, olaparib, is approved for use in several countries for the treatment of breast, ovarian, fallopian tube and peritoneal cancer patients with an inherited *BRCA1* or *BRCA2* mutation [376].

Another recent study demonstrated that the small molecule inhibitor, YK-4-279, can also inhibit the biological activity of *ERG* and *ETV1* [377]. These small molecules do not significantly decrease ERG or ETV1 protein levels, instead they downregulate their targets, thus preventing protein-protein interactions [377]. An *in vivo* mouse xenograft model study by the same group demonstrated that *Etv1* fusion positive mice treated with YK-4-279 developed fewer tumours and were less likely to develop lung metastases compared to untreated *Etv1* fusion positive mice [378]. These studies provide promising evidence that *ETS*-based inhibitors may soon become an important tool in the treatment of PCa in *ETS* fusion-positive patients.

More specifically, this project may one day impact many Tasmanian PCa patients by improving their screening and treatment options. Screening options could include screening for *ETS* gene fusions in tumour samples using FISH, a cost-effective method that is routinely used in the clinic, and/or screening for underlying genetic variants associated with the development of somatic gene fusions. Luedeke and colleagues (2016) found that known PCa risk variants at 8q24 and 17q24 are differentially associated with *TMPRSS2:ERG* fusion status [379]. This suggests that subtype-specific risk variants could be ideal for stratifying PCa patients, in turn helping a clinician decide whether their patient may benefit from *ETS* therapies, such as *PARP* and *ETV1* small molecule inhibitors.

7.4.6 Limitations of this study

This study has provided important insights into the frequency and type of fusion events in prostate tumours from Tasmanian cases, but there are some limitations that should be raised. A small proportion of tumours from a single Tasmanian family were assayed on the RNA fusion panel, which may have restricted our opportunity to find a larger range of fusion events, given that some of these events may be caused by underlying genetic drivers. As only a proportion of tumours from PcTas9 were able to be sourced from pathology laboratories and subsequently assayed, it was hard to determine whether fusion events clustered in closely related PCa cases. Particularly limiting is that tumours from many affected men in the older generations of this family are not available, therefore this study relies on the collection of tumours from cases diagnosed within the last 10 years or so. Unfortunately, due to time constraints, while it was possible to explore the frequency of the *TMPRSS2:ERG* and *ETV1* fusions in our tumour resource, it was not possible to examine my hypothesis that germline variation predispose to these fusion events nor follow-up the additional novel fusion events that were identified.

With regards to the overall tumour resource, the sample size used in the gene and protein expression analyses was relatively small, which reduced our power for finding any additional tumours that overexpressed ETV1. The quality of DNA and RNA extracted from FFPE tissues is also fairly challenging to work with, therefore these findings require validation in larger FFPE cohorts or, if available, fresh frozen samples. Lastly, in the IHC experiment, due to a lack of information, it was impossible to determine where exactly the ETV1 antibody bound (specified as 'middle region' by ThermoFisher). Therefore, it is difficult to conclude that it is

able to detect the two *ETV1* fusion transcripts identified in this study. Notably, in most instances, ETV1 protein expression was negative in our prostate tumour samples.

7.4.7   Gene fusions and chromosomal alterations; comparison of Chapters 6 and 7

In total, 10 PcTas9 tumours were assayed on both the TruSight RNA Fusion Panel and the array Comparative Genomic Hybridisation (aCGH). As discussed earlier (7.1) gene fusions are often caused by chromosomal inversions, translocations, amplifications or deletions [380]. Thus, the *TMPRSS2:ERG* fusion may be the result of a deletion at 21q22 and an *ETV1* fusion positive tumour may have an amplification or deletion of the 7p21.2 chromosomal region. Six *TMPRSS2:ERG* positive tumours were also assayed on the aCGH, yet none had a 21q22 deletion. In terms of *ETV1*, PC9-158 was found to have a *SLC30A4:ETV1* fusion and although the sample did not pass aCGH quality control, there was an amplification seen across this region. On the contrary, PC9-20 had an amplification of the 7p21.2 region on the array, but no *ETV1* fusion was detected on the RNA Fusion Panel. As discussed in Chapter 7.3.5, PC9-20 had comparable *ETV1* expression ratios to a low rate *ETV1* fusion carrier, PC9-603, which may indicate that the *ETV1* fusion in this tumour is expressed at a level too low to be detected by the fusion panel. This is an assumption and it is possible that the 7p21.2 amplification in the PC9-20 tumour did not translate to a fusion event. Likewise, the *TMPRSS2:ERG* fusion is not always the result of a chromosomal deletion, however there may be other reasons as to why discrepant results were seen between the two methods.

One such explanation is tumour heterogeneity, a known phenomenon of PCa. It is common knowledge that PCa arises from multiple, independent clonal expansions [381-383] and as a result, 56-87% of all PCa cases of contemporary radical prostatectomies have multifocal disease [383]. Thus, heterogeneity is evident between prostate foci, but it can also vary at different depths from the same tumour area in a tissue block [384]. As the chromosomal aberrations and gene fusion events discussed in Chapters 6 & 7 are somatic changes that occur at the tumour level, it is possible that the DNA and RNA extracted from the FFPE samples represent different tumour foci as they were not co-extracted. Therefore, their genomic profile may appear different, which would explain why we see discrepant results between the fusion panel and aCGH analysis. The phenomenon of tumour heterogeneity was also apparent in our gene expression results discussed in Chapter 6, in which different PCa foci may have been assessed between aCGH analysis (FFPE DNA) compared to our RT-qPCR experiment (FFPE RNA). The availability of sufficient tumour tissue is always challenging and it is not always feasible

to extract both DNA or RNA, yet alone in parallel. However, the issue of tumour heterogeneity could be counteracted by the extraction of nucleic acids in parallel from the same tissue microdissection if possible.

7.5   FUTURE DIRECTIONS

This is the first study to identify the involvement of *WHSC1L1, CNKSR3, SLC30A4, C19orf48* and *RYBP* in a fusion event in PCa. Therefore, it is essential to screen larger prostate tissue cohorts for these fusion events. The literature suggests that both novel fusion events, *WHSC1L1:CNKSR3* and *RYBP:FOXP1*, could potentially be biologically relevant in PCa. Neither *WHSC1L1* or *CNKSR3* have previously been associated with PCa [340] however, *WHSC1L1* is highly expressed and *CNKSR3* is lowly expressed in the prostate [137], which is a typical 5' and 3' expression profile for a fusion gene. Notably, *CNKSR3* is an aldosterone-induced scaffold protein required for assembly of epithelial sodium channels, and sodium channels are abnormally expressed in malignant compared to matched benign tissue in a number of cancers [385]. The second novel fusion gene involved exon 1 of *RYBP,* a component of the Polycomb group multiprotein PRC1-like complex, which was fused to exon 5 of *FOXP1* [386]. Whilst *FOXP1* is a known partner in prostate tumour fusion events, it has only previously been identified as the 5' fusion partner of *ETV1,* causing transcriptional activation through AR-binding enhancers [346,387]. *FOXP1*, like other *FOX* transcription factors, plays an important role in the regulation of tissue- and cell-specific gene transcription during both development and adulthood [116]. This is the first PCa study to observe *FOXP1* as the 3' fusion partner, although, this phenomenon has previously been identified in a case of B-cell acute lymphoblastic leukemia, in which *PAX5* was fused to *FOXP1* [388]. Collaboration with national and international groups with access to prostate tissue samples, e.g. the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium, will enable us to determine the frequency of these two novel fusion events in other populations.

It is possible that the biological importance of the *RYBP:FOXP1* fusion may be due to the deletion of the region between the two genes, as deletion of the chromosome 3p13 region has been associated with poor prognosis and therapy resistance in PCa [389]. In terms of the *TMPRSS2:ERG* fusion, it can be formed due to a chromosomal translocation or an ~3Mb intrachromosomal deletion of 21q22.2. Linn *et al.* (2016) characterised two mouse models

representing *TMPRSS2:ERG* translocation and deletion events and found that mice lacking the interstitial region developed prostate tumours marked by poorer differentiation and epithelial-to-mesenchymal transition [390]. This study concluded that the loss of tumour suppressors in this region of deletion contributed to disease progression [390]. Therefore, it is possible that the deletion between *RYBP* and *FOXP1* similarly contributes to carcinogenesis. Investigation of this region failed to identify any compelling cancer-associated genes, despite loss of this region previously being associated with PCa as mentioned above [389]. Screening of additional prostate tumours by FISH could identify additional *RYBP:FOXP1* carriers and this would also determine whether the fusion was formed through deletion or translocation. Additionally, the other novel fusion events identified in this study, including *SLC30A4:ETV1* and *C19orf48:ETV4* could also be screened by FISH analysis in a larger prostate tissue cohort.

In this study, the *SLC30A4:ETV1* fusion resulted in increased exon-level expression of *ETV1* in the regions downstream of the breakpoint. Similarly, it would be valuable to determine the effect of the *C19orf48:ETV4* fusion on *ETV4* expression. Recently, RNA hybridisation has emerged as a useful tool for the *in-situ* detection of *ETV1, ETV4* and *ETV5* in FFPE prostate sections, therefore, *ETS* gene rearrangements could be assessed in independent tumour foci [391]. Thus, *ETV4* gene expression could be assessed in the entire *Tasmanian Prostate Tissue Pathology Resource* using this technique, potentially identifying additional *ETV4* fusion carriers. *ETV4* is the third most common *ETS* gene involved in gene fusions [371] and expression has been associated with a poor prognosis in PCa, including a correlation with GS (p=0.045) and pathological tumour stage (p=0.041) [392]. Thus, a fusion event involving *ETV4* could have detrimental effects on normal prostatic pathways and may contribute to the progression of disease.

As described earlier (Chapter 7.4.2), this study suggests that genetic susceptibility may increase the likelihood of some tumours developing the *TMPRSS2:ERG* fusion, as the event was more frequent in tumours from two Tasmanian PCa families, PcTas2 and PcTas9. Unfortunately, due to time constraints and a lack of genetic data for these families, this project was unable to test for an association between germline variants and somatic fusion events. Genome-wide germline genetic data from PcTas2 and 9 individuals would enable us to perform genome-wide linkage analysis based on *TMPRSS2:ERG* fusion status to replicate or identify novel loci associated with the fusion. This approach has been used by Hofer and colleagues (2009), who identified several loci on chromosomes 9, 18 and X that showed suggestive linkage to the

TMPRSS2:ERG fusion positive phenotype. This study assessed 75 patients from 36 German PCa families and found that 73% of fusion positive cases accumulated within 16 specific families [393]. Given that germline DNA is available for 54 Tasmanian cases with known *TMPRSS2:ERG* fusion status, another priority would be to replicate associations with known *TMPRSS2:ERG*-associated variants, including rare variants in *POLI* and *ESCO1* [394], as well as common GWAS variants [379,395]. Overall, targeted collection of additional prostate tumours from newly diagnosed cases in PcTas2 and 9 would assist in determining clustering, and assessing an underlying genetic predisposition to this fusion event.

As the specificity of the ETV1 antibody used in this study is unknown, an antibody targeting the region of ETV1 involved in the fusion event would be beneficial to determine whether ETV1 overexpression in this region is translated to the protein level. However, even more advantageous would be to assess protein expression in two different regions of ETV1 to determine whether there is an overexpression of truncated or full-length ETV1.

7.6   CONCLUSION

This study sought to identify gene fusion events in prostate tumours from a single Tasmanian PCa family, PcTas9, and explore the hypothesis that these somatic events are underpinned by inherited predisposition. Overall, we successfully identified the known *ERG* and *ETV1* fusions in our dataset, as well as four novel fusion events. Notably, the *TMPRSS2:ERG* fusion was more common in two families, PcTas2 and PcTas9, suggesting a germline genetic predisposition. However, due to time limitations we were unable to explore this further and test for associations with specific genetic loci or variants. In the future, the acquisition of genome-wide, germline genetic data and the collection of additional tumours from recently diagnosed familial cases will enable our group provide more insight into this area of research.

CHAPTER 8 : FINAL DISCUSSION

8.1 CONTRIBUTION OF RARE VARIANTS TO PROSTATE CANCER RISK IN A TASMANIAN RESOURCE

Genome-wide association studies (GWAS) of large prostate cancer (PCa) case-control cohorts have identified many common variants, however because of their frequency in the population, they are of limited use in the clinical setting. In recent years, interest has returned to rare variants, given only about one third of the genetic component of PCa risk has been described by common variants. GWAS are not powered to detect rare variants and instead, the combination of family studies and whole-genome sequencing (WGS) has been utilised to determine their contribution to cancer risk. However, this approach has been rarely applied to PCa. Rare variants are by definition rare in the population (MAF <2%) and although individually they may only have a marked effect on disease risk in a small proportion of patients, they provide important information about biological pathways that may be dysregulated in cancer.

The *Tasmanian Familial Prostate Cancer Study* cohorts provided an opportunity to examine rare variant contribution to risk in large PCa families with a dense aggregation of disease. This dissertation has detailed the utilisation of PCa families combined with WGS to identify potential risk variants, using both a targeted and agnostic approach. In total, 20 novel/rare variants were prioritised for validation and segregation analyses, and two of these variants were prioritised for functional assessment. Overall, the total genomic data obtained from sequencing the genomes of 33 individuals identified approximately 6,000 pathogenic rare variants (MAF <2%; CADD >15) in at least one affected family member.

Novel variants in *RND1* and *WNT1* were found to co-segregate with PCa in a single Tasmanian pedigree, PcTas22. A sporadic case was also identified as a carrier, yet we were unable to find a common ancestor with the other variant carriers from PcTas22. Given that they are previously undescribed, screening for these variants in additional familial and case-control cohorts is warranted to determine whether they contribute to PCa risk in other populations. Both *RND1* and *WNT1* are involved in carcinogenesis; *RND1* promotes the growth and migration of cancer cells [177] and high levels of *WNT1* is associated with advanced, metastatic PCa [184]. Even though these two variants appear to be private (to PcTas22 or the Tasmanian population), it is possible

that other variants in these genes contribute to PCa in other families in our resource and/or other populations.

Chapter 4 highlighted a previously undescribed association of an intronic *EZH2* variant with PCa risk. The *EZH2* variant was found to segregate with disease in PcTas12 and was identified in an additional PcTas family, as well as three sporadic PCa cases and one Tasmanian control. *EZH2* is a histone methyltransferase [192-194] and its expression is highly correlated with the progression of PCa [195,196], however the mechanism by which expression increases is currently unknown [196-199]. It is possible that rare variants such as the one identified in this study could contribute to increased *EZH2* expression during PCa progression, although no clear functional role for the *EZH2* variant could be identified here. As a next step, it is fundamental to determine whether this or other variants in *EZH2* are associated with risk in additional PCa cohorts, as without replication, this result would appear to be specific to our Tasmanian population or a false positive.

The previously identified *HOXB13* G84E variant [11] was found to contribute to PCa risk in the Tasmanian population. Here it was initially identified in two individuals from a single family, PcTas72 following examination of WGS data from 33 individuals. Later it was found to contribute to disease risk in six Tasmanian pedigrees and was also identified in three sporadic cases. In the original family, PcTas72, the G84E variant was only identified in two small branches of the large pedigree, demonstrating the heterogeneity of this disease. Thus, it is likely that a combination of a number of common and rare variants are contributing to disease risk in this and other families.

Rare variants in *CCL26, P2RX7* and *ATM* validated and segregated in their founder families, however they were not found to be significantly associated with PCa risk in the Tasmanian population by M_{QLS} analysis. Enrichment analysis of the *P2RX7* and *ATM* variants found higher carrier frequencies in familial compared to sporadic cases, suggesting that there may be a link with inherited PCa predisposition. Through collaboration with members of the International Consortium for Prostate Cancer Genetics (ICPCG) we will be able to assess the contribution of the rare variants discovered here to PCa risk in other populations, including the rare variants in *RND1, WNT1* and *EZH2.* The ICPCG consists of whole-exome sequencing data for over 500 PCa cases, the majority with a strong family history of disease and therefore, is an ideal cohort for replication analyses. Additionally, data from the Prostate Cancer

Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium, a case-control cohort, could be screened for rare variants that were replicated in the ICPCG cohort.

8.2 THE UTILISATION OF A FAMILY-BASED APPROACH TO RARE VARIANT DISCOVERY

This study involved a family-based approach to gene discovery and there are many strengths to this methodology. The *Tasmanian Familial Prostate Cancer Cohort* is an extensive, unique resource, which consists of multi-generational Tasmanian PCa families. These pedigrees are more genetically homogeneous than other populations and therefore, there is likely to be an enrichment of disease-causing genes within these extended families. This feature enhances statistical power for risk variant discovery, especially rare variants. Plus, families are assumed, to some extent, to have similar environmental exposures, which enables the direct association of genetic variants with PCa to be realised. An additional advantage is that WGS of family members allows for a more stringent quality control measure using Mendelian inheritance patterns. However, a major limitation of our WGS study was that we were only able to sequence a limited number of individuals per family due to the associated cost and availability of genetic material, which meant that we could only focus on a small cluster of disease in each family. Whilst sometimes challenging, we aimed to sequence distantly related, affected family members as these cases share a smaller proportion of their genomes, which narrows down the search for rare disease-causing variants. Because of the infrequency, we were also only able to WGS a limited number of unaffected older male relatives who we could use as controls to help during the filtering process. Recruitment of additional family members, particularly distantly related cases and unaffected, older, first-degree males relatives, and availability of funding would enable us to sequence additional genomes to aid in the prioritisation of disease-causing rare variants.

A strength of this study was that we were able to utilise the *Tasmanian Prostate Cancer Case-Control Study,* which enabled us to assess the impact of rare variants on PCa risk in the wider Tasmanian population. Genotyping of this cohort permitted us to determine whether the rare variants identified in familial cases also contributed to sporadic disease in Tasmania. If so, it is important to screen these variants in larger national and international cohorts.

Whilst this study successfully identified rare germline variants associated with PCa risk, the focus has been on the coding regions of the genome. It should be noted, however, that rare variants within introns, intergenic regions and regulatory elements (e.g. promoters, enhancers, silencers and insulators) are also likely to contribute to PCa susceptibility. Such variants can alter gene expression or result in cryptic splice sites, impacting gene function and influencing the development of disease. WGS of individuals from our valuable Tasmanian families was chosen over WES, to provide us with the opportunity to examine non-coding variation in future analyses. The *in silico* tools available for detecting non-coding variants with a functional impact are rapidly evolving, however there are still some significant challenges. Functional annotation of such variants is a huge task, with non-coding regions making up approximately 98% of the human genome. As discussed, linkage analysis in appropriate disease-enriched families will help us to narrow down regions of interest, including non-coding regions that can be examined with more insight and confidence in the near-future.

Overall, this study has provided evidence that the combination of Tasmanian PCa pedigrees and WGS can successfully identify rare disease-causing variants in known and novel cancer associated genes. The identification of the known *HOXB13* G84E variant in our PCa resource has shown that this variant also contributes to PCa risk in Tasmania, as it does to many other Caucasian PCa cohorts. The *ATM* variant, rs1800057 identified in PcTas4, was also recently identified in a large GWAS meta-analysis [9], following imputation to fine-map this region, which illustrates that the agnostic pipeline utilised in this study can successfully identify rare segregating variants.

8.3   EXAMINING THE FUNCTIONAL IMPACT OF RARE PROSTATE CANCER RISK VARIANTS

The identification of PCa risk variants requires replication in additional cohorts to provide further evidence for an association with PCa risk. Functional studies are also required to demonstrate how they play a role in disease initiation, yet this is often challenging. Chapter 4 detailed the assessment of the functional impact of the intronic *EZH2* variant, using an *in vitro* splicing assay, however no effect on splicing was demonstrated. *EZH2* expression was unable to be quantitated in our formalin-fixed paraffin embedded (FFPE) prostate tumours. Whilst this variant didn't appear to affect *EZH2*, there was some evidence to suggest that it may affect the expression of splicing factors associated with *EZH2*. Overall, analysis of the functional effect

of the *EZH2* intronic variant was challenging, partly because the function of untranslated regions of genes, including introns, intergenic regions and regulatory elements is not yet fully understood. This is further complicated by the fact that these regions may only be functional in specific cell types. *In silico* prediction tools of pathogenicity and deleteriousness of non-coding variants, in combination with datasets annotated with regulatory elements (such as the GTEx Portal; https://gtexportal.org/home/) [137] are now helping researchers gain a better understanding of their predicted functional effect before laboratory validation.

Whilst we and numerous other studies have replicated the association of the *HOXB13* G84E variant with PCa risk [396], no study has reported on the functional effect of this variant. Functional assessment of this variant suggested that it was rarely transcribed in G84E carrier prostate tissue (benign or malignant glands), nor did it have an effect on gene or protein expression (Chapter 5). Further analyses suggested that epigenetic mechanisms don't appear to account for the unbalanced allele transcription seen in G84E variant carriers. Therefore, future studies could focus on whether copy number variation at the *HOXB13* site or rapid targeted degradation of the variant mRNA transcript underpin the observed allelic imbalance. Given that *HOXB13* is essential for vertebrate embryonic development [116], it is possible that the G84E variant may affect the development of the normal prostate during embryonic development, when *HOXB13* expression levels are very high, and these changes may make the prostate susceptible to tumour development later in life.

Given the rarity of the *EZH2* and *HOXB13* variants it was challenging to collect a considerable sample size of prostate specimens from variant carriers. A small sample size results in reduced power and therefore lowers the likelihood of detecting real functional effects, which may explain why we didn't see any differences in expression between variant carriers and non-carriers. The quality of DNA and RNA extracted from FFPE tissues is also fairly poor, which makes functional assays challenging. Thus, the concepts explored in this study should be applied to a larger tissue cohort of *EZH2* and *HOXB13* carriers, which could be achieved in collaboration with other PCa groups with access to larger FFPE cohorts, or where possible fresh frozen cohorts, and this is underway.

8.4 EXPLORING GERMLINE VARIANT PREDISPOSITION TO SOMATIC TUMOUR ALTERATIONS

The *Tasmanian Familial Prostate Cancer Study* is a highly valuable resource as it is one of a limited number of cohorts comprised of large families with germline and tumour DNA available for multiples cases. This has allowed us to explore the relatively new and non-traditional hypothesis that there is an inherited predisposition to some somatic alterations in prostate tumours. Chapters 6 and 7 described chromosomal alterations identified in Tasmanian prostate tumours, including chromosomal amplifications and deletions (Chapter 6), and translocations resulting in gene fusions (Chapter 7).

To the best of our knowledge, this is the first study to assay tumours from a single family by array Comparative Genomic Hybridisation (aCGH). This study highlighted a novel amplification of *EEF2* (19p13.3) and follow-up gene expression analysis validated this finding in five out of eight tumours. Analysis of matched malignant and benign prostate glands suggested that *EEF2* overexpression is a feature of malignant glands only. Overall, *EEF2* mediates protein synthesis [116], a key characteristic of cancer cells, and overexpression of *EEF2* in cancer cell lines suggests that it significantly enhances cell growth through cell cycle progression [302]. Thus, whilst further assessment is required, it is possible that the *EEF2* amplification observed in our study may result in cell cycle alterations, leading to increased tumourigenesis. In fact, EEF2 overexpression has recently been suggested as an ideal therapeutic target [299]. Particularly interesting is that this amplification was consistently identified in multiple family members, suggesting an inherited predisposition. This hypothesis will be further investigated by utilising genome-wide, germline genetic data from these individuals.

Chapter 7 detailed the identification of the *TMPRSS2:ERG* fusion in our *Tasmanian Prostate Tissue Pathology Resource,* as well as the identification of novel fusion events in tumours, including *WHSC1L1:CNKSR3, RYBP:FOXP1, SLC30A4:ETV1* and *C19orf48:ETV4*. Given that this study is the first to describe the involvement of *WHSC1L1, CNKSR3, SLC30A4, C19orf48* and *RYBP* in a PCa fusion event, it is essential to screen larger prostate tissue cohorts to determine their frequency. Currently, *ETS*-fusion status is the major molecular subclassifier of localised PCa, yet it is currently still debated whether these and other fusion events are associated with poor clinical outcomes or not. Thus, in addition to those found previously, the

novel fusion events and novel *ETV1/4* fusion partners identified here must be investigated to determine whether all or only particular fusion events are associated with clinical outcomes, both good and poor. In the future, screening of *ETS* gene fusions in prostate tumours may provide us with valuable knowledge about the disease and its prognosis, which could inform targeted therapeutic options. This study has proven that there are a number of different prostate fusion events, therefore screening tools like the RNA Fusion Panel would be advantageous, however this is currently too expensive for routine clinical use.

It is apparent that the amplification of the *EEF2* gene is more common in tumours from PcTas9, with only a few individual tumours from other families showing similar expression patterns. The *TMPRSS2:ERG* fusion was also more frequent in tumours from this same Tasmanian PCa family. Appendix 30 shows the overlap of tumours with *EEF2* amplification and *TMPRSS2:ERG* fusion status; four tumours had both an *EEF2* amplification and were fusion positive, eight tumours had only one alteration and six tumours had neither alteration. Given that PCa is a complex disease, it is likely that there are multiple drivers of disease even within this one family.

Given the high frequency of *TMPRSS2:ERG* fusion events in these families, and accumulating evidence from previous studies, they are unlikely due to chance. The higher frequency of this fusion event in two Tasmanian PCa families suggests that there is an underlying genetic predisposition. This is further supported by the fact that none of the sporadic PCa tumours were fusion positive. It has been suggested that inherited germline variants in DNA repair genes can lead to increased chromosomal rearrangements, resulting in *TMPRSS2* fusing to *ERG* [356]. Whilst germline data was not available for all individuals, we have recently obtained GSA (Global Screening Array) SNP (single nucleotide polymorphism) array and WGS data from individuals from both PcTas2 and PcTas9. The SNP array data will enable us to perform genome-wide linkage analysis, based on *TMPRSS2:ERG* fusion status. The highlighted linkage loci can then be examined in the WGS data, which narrows the search for underlying germline genetic variants that may be associated with these somatic tumour events. We could also determine whether previously reported *TMPRSS2:ERG*-associated variants, including rare variants in *POLI* and *ESCO1* [394], as well as common GWAS variants [379,395] are present in cases in our Tasmanian cohort. It would also be interesting to test for association of the 63 common variants recently identified by a GWAS meta-analysis [9] with fusion status. Overall, collection of additional tumours from newly diagnosed cases in PcTas9 would assist in determining

clustering of the *TMPRSS2:ERG* fusion and *EEF2* amplification, which would enable us to further assess whether there is an underlying genetic predisposition to these somatic tumour events.

8.5 CLINICAL SIGNIFICANCE OF THIS STUDY

The research presented in this thesis has primarily been undertaken to advance our understanding of the underlying mechanisms of PCa risk and progression, with the longer-term goal of translating this knowledge in to the clinical setting. A Prostate Cancer Comprehensive Panel is currently offered to men with a family history of disease (Fulgent Genetics, CA, USA). This panel examines 12 genes associated with an increased risk for PCa, including *ATM, BRCA1, BRAC2, CHEK2, HOXB13* and *NBN*. This test is designed to identify germline pathogenic variants that may increase PCa risk. A positive result can prompt screening options for early detection and treatment of cancer, as well as encouraging the testing of other relatives. For our Tasmanian families carrying the *HOXB13* G84E, genetic testing has been offered free of charge, through the Tasmanian Genetic Counselling Service (with ethics approval). Given that variants in the 12 above-mentioned genes only explain a minor proportion of disease heritability, the identification of additional pathogenic variants and/or PCa predisposition genes will enable us to better inform men of their risk. For example, three previously undescribed associations with PCa were identified in this study and this knowledge could be disseminated to these families to inform their disease risk. If these findings are replicated and further investigation strengthens the argument that these genes are involved in cancer, the *RND1*, *WNT1* and *EZH2* genes may be included in PCa screening panels in the future. There is also currently a strong push to implement polygenic risk scores based on common variants in the clinical setting, yet with only one-third of genetic predisposition explained, this may be premature. Therefore, identification of rare germline risk variants will aid in the implementation of polygenic risk scores in to the clinic.

Routine cytogenetic analysis aids in the diagnosis of many cancers, particularly haematological malignancies, yet the clinical implementation of this has not yet been realised for PCa. The knowledge of specific somatic tumour alterations could define particular disease phenotypes and inform a man's response to treatment. If the recurrent *EEF2* amplification and/or the novel fusion events identified in this study were found to be associated with certain clinical outcomes, men could be tested for these at diagnostic biopsy, when the disease is most curable.

Knowledge of both inherited and somatic genetic alterations is now also informing treatment strategies. One promising therapeutic candidate which impairs tumorigenesis and cell invasion is the Poly-ADP ribose polymerase (PARP) inhibitors [397]. The PARP1 inhibitor, olaparib, is approved for use in several countries for the treatment of breast, ovarian, fallopian tube and peritoneal cancer patients with an inherited *BRCA1* or *BRCA2* mutation [376]. In the presence of a PARP inhibitor, a cell is PARP1 deficient, and together with a *BRCA1* or *BRCA2* variant, the cells cannot repair DNA damage effectively and die. PARP inhibitors also appear to be effective in the presence of deleterious variants in other DNA repair genes, including *ATM, CHEK2* and *PALB2* and clinical trials have been initiated in metastatic PCa patients. However, a study by Marshall and colleagues (2018) found that metastatic castration-resistant PCa with somatic *ATM* mutations responded poorly to PARP inhibitors, compared to those with *BRCA1* or *BRCA2* mutations, and concluded that alternative therapies should be explored for PCa cases with variants in *ATM* [398].

Somatic tumour events, such as *ERG*, *ETV1* and other *ETS* gene fusions may also benefit from PARP1 inhibitors as these tumours are susceptible to PARP inhibition through the increased incidence of DNA double strand breaks [375]. Another recent study demonstrated that a small molecule inhibitor, YK-4-279, has also been developed to inhibit the biological activity of ERG and ETV1 [399,400]. Overall, targeting DNA repair genes, ETS fusion proteins, or their binding partners, their DNA binding sites, or their downstream effectors provides multiple avenues through which tumour progression or metastasis can be effectively prevented.

8.6   FINAL CONCLUSION

In conclusion, this study has highlighted the success of combining Tasmanian PCa pedigrees with a dense aggregation of disease and WGS to narrow the search for rare disease-causing variants. Four novel/rare variants in *RND1, WNT1 EZH2* and *HOXB13* were found to be significantly associated with PCa risk in the Tasmanian population, three of which were previously undescribed. Despite the heterogenous nature of this disease, this study has also shown that some somatic alterations are shared by family members. The *EEF2* amplification in tumours from PcTas9 is a particularly interesting finding as EEF2 overexpression has recently been suggested as an ideal therapeutic target. Overall the findings of this study have highlighted genes and biological pathways that may be involved in PCa development in

Tasmania. The rare variants identified here could help explain some of the 'missing' PCa heritability, while our tumour work will lead to a better understanding of the link between germline variants and somatic events. Understanding the genetic determinants of disease development and somatic tumour variation will ultimately lead to better screening, diagnostic and therapeutic options for PCa patients.

CHAPTER 9 : **REFERENCES**

1   FitzGerald, LM *et al.* Impact of the G84E variant on HOXB13 gene and protein expression in formalin-fixed, paraffin-embedded prostate tumours. *Sci Rep* 7, 17778, (2017).

2   Thornton, T & McPeek, MS. Case-control association testing with related individuals: a more powerful quasi-likelihood score test. *Am J Hum Genet* 81, 321-337, (2007).

3   Field, MA *et al.* Reliably Detecting Clinically Important Variants Requires Both Combined Variant Calls and Optimized Filtering Strategies. *PLoS One* 10, e0143199, (2015).

4   Field, MA *et al.* Reducing the search space for causal genetic variants with VASP. *Bioinformatics* 31, 2377-2379, (2015).

5   Bray, F *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68, 394-424, (2018).

6   Welfare, AIoHa. Australia's health 2018. (Canberra: Australian Institute of Health and Welfare, 2018).

7   Stokes, BA, T.; Otahal, P.; Venn, A. Cancer in Tasmania: Incidence and Mortality 2016. *Menzies Institute for Medical Research Tasmania*, (October 2018).

8   Benafif, S *et al.* A Review of Prostate Cancer Genome-Wide Association Studies (GWAS). *Cancer Epidemiol Biomarkers Prev* 27, 845-857, (2018).

9   Schumacher, FR *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* 50, 928-936, (2018).

10   Mancuso, N *et al.* The contribution of rare variation to prostate cancer heritability. *Nat Genet* 48, 30-35, (2016).

11   Ewing, CM *et al.* Germline mutations in HOXB13 and prostate-cancer risk. *N Engl J Med* 366, 141-149, (2012).

12   Fitzgerald, LM *et al.* Germline missense variants in the BTNL2 gene are associated with prostate cancer susceptibility. *Cancer Epidemiol Biomarkers Prev* 22, 1520-1528, (2013).

13   Karyadi, DM *et al.* Whole exome sequencing in 75 high-risk families with validation and replication in independent case-control studies identifies TANGO2, OR5H14, and CHAD as new prostate cancer susceptibility genes. *Oncotarget* 8, 1495-1507, (2017).

14   Zuhlke, KA *et al.* Identification of a novel germline SPOP mutation in a family with hereditary prostate cancer. *Prostate* 74, 983-990, (2014).

15   Hsing, AW *et al.* International trends and patterns of prostate cancer incidence and mortality. *Int J Cancer* 85, 60-67, (2000).

16   Kvale, R *et al.* Interpreting trends in prostate cancer incidence and mortality in the five Nordic countries. *J Natl Cancer Inst* 99, 1881-1887, (2007).

17   Center, MM *et al.* International variation in prostate cancer incidence and mortality rates. *Eur Urol* 61, 1079-1092, (2012).

18   Epstein, MM *et al.* Dietary zinc and prostate cancer survival in a Swedish cohort. *Am J Clin Nutr* 93, 586-593, (2011).

19   Ferlay, J *et al.* Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer*, (2018).

20   McNeal, JE. Normal histology of the prostate. *Am J Surg Pathol* 12, 619-633, (1988).

21   McNeal, JE. The prostate gland: morphology and pathobiology. *Monograoh Urol* 9, 36-63, (1988).

22   Cunha, GRF, B.A.; Sugimura, Y.; Hom, Y.K. Keratinocyte growth factor as mediator of mesenchymal-epithelial interaction in the development of androgen target organs. *Cell Dev Biol* 7, 203-210, (1996).

23   Davey, RA & Grossmann, M. Androgen Receptor Structure, Function and Biology: From Bench to Bedside. *Clin Biochem Rev* 37, 3-15, (2016).

24   Kim, EH *et al.* Management of Benign Prostatic Hyperplasia. *Annu Rev Med* 67, 137-151, (2016).

25   Schulz, WA *et al.* Molecular biology of prostate cancer. *Mol Hum Reprod* 9, 437-448, (2003).

26   Shahbazi, S *et al.* Gene expression profile of FVII and AR in primary prostate cancer. *Cancer Biomark* 17, 353-358, (2016).

27   Carroll, P *et al.* Prostate-specific antigen best practice policy--part I: early detection and diagnosis of prostate cancer. *Urology* 57, 217-224, (2001).

28   Smith, RA *et al.* American Cancer Society guidelines for the early detection of cancer: update of early detection guidelines for prostate, colorectal, and endometrial cancers. Also: update 2001--testing for early lung cancer detection. *CA Cancer J Clin* 51, 38-75; quiz 77-80, (2001).

29   Descotes, JL. Diagnosis of prostate cancer. *Asian J Urol* 6, 129-136, (2019).

30   Andriole, GL, Jr. PSA screening and prostate cancer risk reduction. *Urol Oncol* 30, 936-937, (2012).

31   DeMarzo, AM *et al.* Pathological and molecular aspects of prostate cancer. *Lancet* 361, 955-964, (2003).

32   Polascik, TJ *et al.* Prostate specific antigen: a decade of discovery--what we have learned and where we are going. *J Urol* 162, 293-306, (1999).

33   Carter, HB *et al.* Estimation of prostatic growth using serial prostate-specific antigen measurements in men with and without prostate disease. *Cancer Res* 52, 3323-3328, (1992).

34   Carter, HB. Prostate cancers in men with low PSA levels--must we find them? *N Engl J Med* 350, 2292-2294, (2004).

35   Wong, H & Nilges, E. Living With "Man's Fate" Away From Home. *J Patient Exp* 3, 105-107, (2016).

36   Miller, LS *et al.* PSA divergence. A new parameter for the accurate longitudinal assessment of prostatic disease. *Am J Clin Oncol* 19, 217-222, (1996).

37   Wiklund, F. Prostate cancer genomics: can we distinguish between indolent and fatal disease using genetic markers? *Genome Med* 2, 45, (2010).

38   Panel, PCFoAaCCAPTGEA. Draft clinical practice guidelines for PSA testing and early management of test-detected prostate cancer. (2016).

39   Kalish, LA *et al.* Family history and the risk of prostate cancer. *Urology* 56, 803-806, (2000).

40   Epstein, JI *et al.* A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score. *Eur Urol* 69, 428-435, (2016).

41   Tomlins, SA *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310, 644-648, (2005).

42   Bismar, TA *et al.* Clinical utility of assessing PTEN and ERG protein expression in prostate cancer patients: a proposed method for risk stratification. *J Cancer Res Clin Oncol* 144, 2117-2125, (2018).

43   Cancer Genome Atlas Research, N. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* 163, 1011-1025, (2015).

44   Shoag, J & Barbieri, CE. Clinical variability and molecular heterogeneity in prostate cancer. *Asian J Androl* 18, 543-548, (2016).

45   Robinson, D *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* 161, 1215-1228, (2015).

46   Barbieri, CE *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet* 44, 685-689, (2012).

47   Liu, W *et al.* Identification of novel CHD1-associated collaborative alterations of genomic structure and functional assessment of CHD1 in prostate cancer. *Oncogene* 31, 3939-3948, (2012).

48   Brooks, JD *et al.* Evaluation of ERG and SPINK1 by Immunohistochemical Staining and Clinicopathological Outcomes in a Multi-Institutional Radical Prostatectomy Cohort of 1067 Patients. *PLoS One* 10, e0132343, (2015).

49   Tomlins, SA *et al.* The role of SPINK1 in ETS rearrangement-negative prostate cancers. *Cancer Cell* 13, 519-528, (2008).

50   Arora, K & Barbieri, CE. Molecular Subtypes of Prostate Cancer. *Curr Oncol Rep* 20, 58, (2018).

51   Kaffenberger, SD & Barbieri, CE. Molecular subtyping of prostate cancer. *Curr Opin Urol* 26, 213-218, (2016).

52   Hussein, S *et al.* Young-age prostate cancer. *J Clin Pathol* 68, 511-515, (2015).

53   Klotz, LH. PSA recurrence: definitions, PSA kinetics, and identifying patients at risk. *Can J Urol* 13 Suppl 2, 43-47, (2006).

54   Klotz, L. Active surveillance for low-risk prostate cancer. *F1000 Med Rep* 4, 16, (2012).

55   Nanus, DM. Decision to operate, radiate, or watch and wait: the prostate cancer dilemma. *Cancer Invest* 17, 374-375, (1999).

56   Johansson, E *et al.* Long-term quality-of-life outcomes after radical prostatectomy or watchful waiting: the Scandinavian Prostate Cancer Group-4 randomised trial. *Lancet Oncol* 12, 891-899, (2011).

57   Goy, BW *et al.* Treatment results of brachytherapy vs. external beam radiation therapy for intermediate-risk prostate cancer with 10-year followup. *Brachytherapy* 15, 687-694, (2016).

58   Wallis, CJD *et al.* Surgery Versus Radiotherapy for Clinically-localized Prostate Cancer: A Systematic Review and Meta-analysis. *Eur Urol* 70, 21-30, (2016).

59   Crawford, ED & Moul, JW. ADT risks and side effects in advanced prostate cancer: cardiovascular and acute renal injury. *Oncology (Williston Park)* 29, 55-58, 65-56, (2015).

60   Abrahamsson, PA. Potential benefits of intermittent androgen suppression therapy in the treatment of prostate cancer: a systematic review of the literature. *Eur Urol* 57, 49-59, (2010).

61   Crawford, ED. Epidemiology of prostate cancer. *Urology* 62, 3-12, (2003).

62   Gueye, SM *et al.* Clinical characteristics of prostate cancer in African Americans, American whites, and Senegalese men. *Urology* 61, 987-992, (2003).

63   Rebbeck, TR *et al.* Global patterns of prostate cancer incidence, aggressiveness, and mortality in men of african descent. *Prostate Cancer* 2013, 560857, (2013).

64   Carter, BS *et al.* Mendelian inheritance of familial prostate cancer. *Proc Natl Acad Sci U S A* 89, 3367-3371, (1992).

65   Eeles, RA *et al.* Familial prostate cancer: the evidence and the Cancer Research Campaign/British Prostate Group (CRC/BPG) UK Familial Prostate Cancer Study. *Br J Urol* 79 Suppl 1, 8-14, (1997).

66   Singh, G *et al.* Risk of seizures and neurocysticercosis in household family contacts of children with single enhancing lesions. *J Neurol Sci* 176, 131-135, (2000).

67   Steinberg, GD *et al.* Family history and the risk of prostate cancer. *Prostate* 17, 337-347, (1990).

68   Ahlbom, A *et al.* Cancer in twins: genetic and nongenetic familial risk factors. *J Natl Cancer Inst* 89, 287-293, (1997).

69   Goldgar, DE *et al.* Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst* 86, 1600-1608, (1994).

70   Gronberg, H *et al.* Studies of genetic factors in prostate cancer in a twin population. *J Urol* 152, 1484-1487; discussion 1487-1489, (1994).

71   Gronberg, H *et al.* Familial prostate cancer in Sweden. A nationwide register cohort study. *Cancer* 77, 138-143, (1996).

72   Kicinski, M *et al.* An epidemiological reappraisal of the familial aggregation of prostate cancer: a meta-analysis. *PLoS One* 6, e27130, (2011).

73   Negri, E *et al.* Family history of cancer and the risk of prostate cancer and benign prostatic hyperplasia. *Int J Cancer* 114, 648-652, (2005).

74   Hjelmborg, JB *et al.* The heritability of prostate cancer in the Nordic Twin Study of Cancer. *Cancer Epidemiol Biomarkers Prev* 23, 2303-2310, (2014).

75   Lichtenstein, P *et al.* Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 343, 78-85, (2000).

76   Gronberg, H *et al.* Segregation analysis of prostate cancer in Sweden: support for dominant inheritance. *Am J Epidemiol* 146, 552-557, (1997).

77   Schaid, DJ *et al.* Evidence for autosomal dominant inheritance of prostate cancer. *Am J Hum Genet* 62, 1425-1438, (1998).

78   Verhage, BA *et al.* Autosomal dominant inheritance of prostate cancer: a confirmatory study. *Urology* 57, 97-101, (2001).

79   Cui, J *et al.* Segregation analyses of 1,476 population-based Australian families affected by prostate cancer. *Am J Hum Genet* 68, 1207-1218, (2001).

80   Pakkanen, S *et al.* Segregation analysis of 1,546 prostate cancer families in Finland shows recessive inheritance. *Hum Genet* 121, 257-267, (2007).

81   Conlon, EM *et al.* Oligogenic segregation analysis of hereditary prostate cancer pedigrees: evidence for multiple loci affecting age at onset. *Int J Cancer* 105, 630-635, (2003).

82   MacInnis, RJ *et al.* Prostate cancer segregation analyses using 4390 families from UK and Australian population-based studies. *Genet Epidemiol* 34, 42-50, (2010).

83   Valeri, A *et al.* Early onset and familial predisposition to prostate cancer significantly enhance the probability for breast cancer in first degree relatives. *Int J Cancer* 86, 883-887, (2000).

84   Leongamornlert, D *et al.* Germline BRCA1 mutations increase prostate cancer risk. *Br J Cancer* 106, 1697-1701, (2012).

85   Kote-Jarai, Z *et al.* BRCA2 is a moderate penetrance gene contributing to young-onset prostate cancer: implications for genetic testing in prostate cancer patients. *Br J Cancer* 105, 1230-1234, (2011).

86   Montgomery, JS *et al.* The androgen receptor gene and its influence on the development and progression of prostate cancer. *J Pathol* 195, 138-146, (2001).

87   Chng, KR & Cheung, E. Sequencing the transcriptional network of androgen receptor in prostate cancer. *Cancer Lett* 340, 254-260, (2013).

88   Cybulski, C *et al.* NBS1 is a prostate cancer susceptibility gene. *Cancer Res* 64, 1215-1219, (2004).

89   Cybulski, C *et al.* CHEK2 is a multiorgan cancer susceptibility gene. *Am J Hum Genet* 75, 1131-1135, (2004).

90   Tischkowitz, M *et al.* Analysis of the gene coding for the BRCA2-interacting protein PALB2 in hereditary prostate cancer. *Prostate* 68, 675-678, (2008).

91   Erkko, H *et al.* A recurrent mutation in PALB2 in Finnish cancer families. *Nature* 446, 316-319, (2007).

92   Thompson, D *et al.* A multicenter study of cancer incidence in CHEK2 1100delC mutation carriers. *Cancer Epidemiol Biomarkers Prev* 15, 2542-2545, (2006).

93   Lange, EM *et al.* Linkage analysis of 153 prostate cancer families over a 30-cM region containing the putative susceptibility locus HPCX. *Clin Cancer Res* 5, 4013-4020, (1999).

94   Gronberg, H *et al.* Early age at diagnosis in families providing evidence of linkage to the hereditary prostate cancer locus (HPC1) on chromosome 1. *Cancer Res* 57, 4707-4709, (1997).

95   Hsieh, CL *et al.* A genome screen of families with multiple cases of prostate cancer: evidence of genetic heterogeneity. *Am J Hum Genet* 69, 148-158, (2001).

96   Smith, JR *et al.* Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. *Science* 274, 1371-1374, (1996).

97   Zheng, SL *et al.* Sequence variants of alpha-methylacyl-CoA racemase are associated with prostate cancer risk. *Cancer Res* 62, 6485-6488, (2002).

98   FitzGerald, LM *et al.* Sequence variants of alpha-methylacyl-CoA racemase are associated with prostate cancer risk: a replication study in an ethnically homogeneous population. *Prostate* 68, 1373-1379, (2008).

99   Langeberg, WJ *et al.* Genetic etiology of hereditary prostate cancer. *Front Biosci* 12, 4101-4110, (2007).

100   Karayi, MK *et al.* Current status of linkage studies in hereditary prostate cancer. *BJU Int* 86, 659-669, (2000).

101   Chung, CC *et al.* Genome-wide association studies in cancer--current and future directions. *Carcinogenesis* 31, 111-120, (2010).

102   Akbari, MR *et al.* Association between germline HOXB13 G84E mutation and risk of prostate cancer. *J Natl Cancer Inst* 104, 1260-1262, (2012).

103   Al Olama, AA *et al.* A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet* 46, 1103-1109, (2014).

104   Al Olama, AA *et al.* Multiple loci on 8q24 associated with prostate cancer susceptibility. *Nat Genet* 41, 1058-1060, (2009).

105   Amin Al Olama, A *et al.* A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease. *Hum Mol Genet* 22, 408-415, (2013).

106   Amundadottir, LT *et al.* A common variant associated with prostate cancer in European and African populations. *Nat Genet* 38, 652-658, (2006).

107   Duggan, D *et al.* Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J Natl Cancer Inst* 99, 1836-1844, (2007).

108   Eeles, RA *et al.* Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* 40, 316-321, (2008).

109   Gudmundsson, J *et al.* Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet* 41, 1122-1126, (2009).

110   Gudmundsson, J *et al.* Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet* 40, 281-283, (2008).

111   Gudmundsson, J *et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 39, 977-983, (2007).

112   Kote-Jarai, Z *et al.* Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat Genet* 43, 785-791, (2011).

113   Schumacher, FR *et al.* Genome-wide association study identifies new prostate cancer susceptibility loci. *Hum Mol Genet* 20, 3867-3875, (2011).

114   Takata, R *et al.* Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat Genet* 42, 751-754, (2010).

115   Thomas, G *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* 40, 310-315, (2008).

116   O'Leary, NA *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 44, D733-745, (2016).

117   Meyer, KB *et al.* A functional variant at a prostate cancer predisposition locus at 8q24 is associated with PVT1 expression. *PLoS Genet* 7, e1002165, (2011).

118   Thorisson, GA *et al.* The International HapMap Project Web site. *Genome Res* 15, 1592-1593, (2005).

119   Jin, G *et al.* Validation of prostate cancer risk-related loci identified from genome-wide association studies using family-based association analysis: evidence from the International Consortium for Prostate Cancer Genetics (ICPCG). *Hum Genet* 131, 1095-1103, (2012).

120   Teerlink, CC *et al.* Association analysis of 9,560 prostate cancer cases from the International Consortium of Prostate Cancer Genetics confirms the role of reported prostate cancer associated SNPs for familial disease. *Hum Genet* 133, 347-356, (2014).

121   Kuehn, BM. 1000 Genomes Project finds substantial genetic variation among populations. *JAMA* 308, 2322, 2325, (2012).

122   Cirulli, ET & Goldstein, DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 11, 415-425, (2010).

123   Manolio, TA *et al.* Finding the missing heritability of complex diseases. *Nature* 461, 747-753, (2009).

124   McCarthy, MI *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9, 356-369, (2008).

125   Koboldt, DC *et al.* The next-generation sequencing revolution and its impact on genomics. *Cell* 155, 27-38, (2013).

126   Zuhlke, KA *et al.* Identification of a novel NBN truncating mutation in a family with hereditary prostate cancer. *Fam Cancer* 11, 595-600, (2012).

127   Bonifaci, N *et al.* Exploring the link between germline and somatic genetic alterations in breast carcinogenesis. *PLoS One* 5, e14078, (2010).

128   Carter, H *et al.* Interaction Landscape of Inherited Polymorphisms with Somatic Events in Cancer. *Cancer Discov* 7, 410-423, (2017).

129   Carter, H & Ideker, T. Common genetic variation in the germline influences where and how tumors develop. *Mol Cell Oncol* 4, e1302905, (2017).

130   Mamidi, TKK *et al.* Interactions between Germline and Somatic Mutated Genes in Aggressive Prostate Cancer. *Prostate Cancer* 2019, 4047680, (2019).

131   Mirza, MR *et al.* Niraparib Maintenance Therapy in Platinum-Sensitive, Recurrent Ovarian Cancer. *N Engl J Med* 375, 2154-2164, (2016).

132   Rugo, HS *et al.* Adaptive Randomization of Veliparib-Carboplatin Treatment in Breast Cancer. *N Engl J Med* 375, 23-34, (2016).

133 Brenner, JC *et al.* Mechanistic rationale for inhibition of poly(ADP-ribose) polymerase in ETS gene fusion-positive prostate cancer. *Cancer Cell* 19, 664-678, (2011).

134 Koressaar, T & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289-1291, (2007).

135 Untergasser, A *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res* 40, e115, (2012).

136 Ye, J *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13, 134, (2012).

137 Consortium, GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45, 580-585, (2013).

138 FitzGerald, LM *et al.* Impact of the G84E variant on HOXB13 gene and protein expression in formalin-fixed, paraffin-embedded prostate tumours. *Sci Rep* 7, 17778, (2017).

139 Endo, A. A historical perspective on the discovery of statins. *Proc Jpn Acad Ser B Phys Biol Sci* 86, 484-493, (2010).

140 Goldstein, JL & Brown, MS. Familial hypercholesterolemia: identification of a defect in the regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity associated with overproduction of cholesterol. *Proc Natl Acad Sci U S A* 70, 2804-2808, (1973).

141 Stankiewicz, P & Lupski, JR. Structural variation in the human genome and its role in disease. *Annu Rev Med* 61, 437-455, (2010).

142 Guerreiro, RJ *et al.* Exome sequencing reveals an unexpected genetic cause of disease: NOTCH3 mutation in a Turkish family with Alzheimer's disease. *Neurobiol Aging* 33, 1008 e1017-1023, (2012).

143 Pottier, C *et al.* Amyloid-beta protein precursor gene expression in alzheimer's disease and other conditions. *J Alzheimers Dis* 28, 561-566, (2012).

144 Pottier, C *et al.* TREM2 R47H variant as a risk factor for early-onset Alzheimer's disease. *J Alzheimers Dis* 35, 45-49, (2013).

145 Chen, Z *et al.* The G84E mutation of HOXB13 is associated with increased risk for prostate cancer: results from the REDUCE trial. *Carcinogenesis* 34, 1260-1264, (2013).

146 Karlsson, R *et al.* A population-based assessment of germline HOXB13 G84E mutation and prostate cancer risk. *Eur Urol* 65, 169-176, (2014).

147 Kote-Jarai, Z *et al.* Prevalence of the HOXB13 G84E germline mutation in British men and correlation with prostate cancer risk, tumour characteristics and clinical outcomes. *Ann Oncol* 26, 756-761, (2015).

148 MacInnis, RJ *et al.* Population-based estimate of prostate cancer risk for carriers of the HOXB13 missense mutation G84E. *PLoS One* 8, e54727, (2013).

149 Stott-Miller, M *et al.* HOXB13 mutations in a population-based, case-control study of prostate cancer. *Prostate* 73, 634-641, (2013).

150 Xu, J *et al.* HOXB13 is a susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG). *Hum Genet* 132, 5-14, (2013).

151 Akbari, MR *et al.* Association between germline HOXB13 G84E mutation and risk of prostate cancer. *J Natl Cancer Inst* 104, 1260-1262, (2012).

152 Breyer, JP *et al.* Confirmation of the HOXB13 G84E germline mutation in familial prostate cancer. *Cancer Epidemiol Biomarkers Prev* 21, 1348-1353, (2012).

153 Karlsson, R *et al.* A population-based assessment of germline HOXB13 G84E mutation and prostate cancer risk. *Eur Urol* 65, 169-176, (2014).

154 Kote-Jarai, Z *et al.* Prevalence of the HOXB13 G84E germline mutation in British men and correlation with prostate cancer risk, tumour characteristics and clinical outcomes. *Ann Oncol* 26, 756-761, (2015).

155 Stott-Miller, M *et al.* HOXB13 mutations in a population-based, case-control study of prostate cancer. *Prostate* 73, 634-641, (2013).

156 Xu, J *et al.* HOXB13 is a susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG). *Hum Genet* 132, 5-14, (2013).

157 Shifman, S & Darvasi, A. The value of isolated populations. *Nat Genet* 28, 309-310, (2001).

158 Lek, M *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285-291, (2016).

159   Choi, Y *et al.* Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 7, e46688, (2012).

160   Adzhubei, IA *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* 7, 248-249, (2010).

161   Kircher, M *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46, 310-315, (2014).

162   McLaren, W *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* 17, 122, (2016).

163   Landrum, MJ *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 46, D1062-D1067, (2018).

164   Rentzsch, P *et al.* CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 47, D886-D894, (2019).

165   Lan, Q *et al.* CCL26 Participates in the PRL-3-Induced Promotion of Colorectal Cancer Invasion by Stimulating Tumor-Associated Macrophage Infiltration. *Mol Cancer Ther* 17, 276-289, (2018).

166   Lin, ZY *et al.* Cancer-associated fibroblasts up-regulate CCL2, CCL26, IL6 and LOXL2 genes related to promotion of cancer progression in hepatocellular carcinoma cells. *Biomed Pharmacother* 66, 525-529, (2012).

167   Schwarz, JM *et al.* MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 11, 361-362, (2014).

168   Venselaar, H *et al.* Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* 11, 548, (2010).

169   Di Virgilio, F. P2RX7: A receptor with a split personality in inflammation and cancer. *Mol Cell Oncol* 3, e1010937, (2016).

170   Di Virgilio, F *et al.* The P2X7 Receptor in Infection and Inflammation. *Immunity* 47, 15-31, (2017).

171   Ying, Y *et al.* Investigation into the association between P2RX7 gene polymorphisms and susceptibility to primary gout and hyperuricemia in a Chinese Han male population. *Rheumatol Int* 37, 571-578, (2017).

172   Metzker, ML. Sequencing technologies - the next generation. *Nat Rev Genet* 11, 31-46, (2010).

173   Nord, AS *et al.* Accurate and exact CNV identification from targeted high-throughput sequence data. *BMC Genomics* 12, 184, (2011).

174   Shirts, BH *et al.* Improving performance of multigene panels for genomic analysis of cancer predisposition. *Genet Med* 18, 974-981, (2016).

175   Walsh, T *et al.* Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc Natl Acad Sci U S A* 108, 18032-18037, (2011).

176   Walsh, T *et al.* Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc Natl Acad Sci U S A* 107, 12629-12633, (2010).

177   Xiang, G *et al.* RND1 is up-regulated in esophageal squamous cell carcinoma and promotes the growth and migration of cancer cells. *Tumour Biol* 37, 773-779, (2016).

178   Komatsu, H *et al.* Attenuated RND1 Expression Confers Malignant Phenotype and Predicts Poor Prognosis in Hepatocellular Carcinoma. *Ann Surg Oncol* 24, 850-859, (2017).

179   Boyrie, S *et al.* RND1 regulates migration of human glioblastoma stem-like cells according to their anatomical localization and defines a prognostic signature in glioblastoma. *Oncotarget* 9, 33788-33803, (2018).

180   Okada, T *et al.* The Rho GTPase Rnd1 suppresses mammary tumorigenesis and EMT by restraining Ras-MAPK signalling. *Nat Cell Biol* 17, 81-94, (2015).

181   Goessling, W *et al.* Genetic interaction of PGE2 and Wnt signaling regulates developmental specification of stem cells and regeneration. *Cell* 136, 1136-1147, (2009).

182   Komiya, Y & Habas, R. Wnt signal transduction pathways. *Organogenesis* 4, 68-75, (2008).

183   Logan, CY & Nusse, R. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* 20, 781-810, (2004).

184   Chen, G *et al.* Up-regulation of Wnt-1 and beta-catenin production in patients with advanced metastatic prostate carcinoma: potential pathogenetic and prognostic implications. *Cancer* 101, 1345-1356, (2004).

185   Slatkin, M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* 9, 477-485, (2008).

186   Kishore, S *et al.* Rapid generation of splicing reporters with pSpliceExpress. *Gene* 427, 104-110, (2008).

187   Oakford, PC *et al.* Transcriptional and epigenetic regulation of the GM-CSF promoter by RUNX1. *Leuk Res* 34, 1203-1213, (2010).

188   Cao, Q *et al.* Repression of E-cadherin by the polycomb group protein EZH2 in cancer. *Oncogene* 27, 7274-7284, (2008).

189   Gao, L *et al.* Higher expression levels of the HOXA9 gene, closely associated with MLL-PTD and EZH2 mutations, predict inferior outcome in acute myeloid leukemia. *Onco Targets Ther* 9, 711-722, (2016).

190   Beke, L *et al.* The gene encoding the prostatic tumor suppressor PSP94 is a target for repression by the Polycomb group protein EZH2. *Oncogene* 26, 4590-4595, (2007).

191   Desmet, FO *et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 37, e67, (2009).

192   Margueron, R & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* 469, 343-349, (2011).

193   Perdigoto, CN *et al.* Epigenetic regulation of skin: focus on the Polycomb complex. *Cell Mol Life Sci* 69, 2161-2172, (2012).

194   Vire, E *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439, 871-874, (2006).

195   Sauvageau, M & Sauvageau, G. Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. *Cell Stem Cell* 7, 299-313, (2010).

196   Varambally, S *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 419, 624-629, (2002).

197   Berezovska, OP *et al.* Essential role for activation of the Polycomb group (PcG) protein chromatin silencing pathway in metastatic prostate cancer. *Cell Cycle* 5, 1886-1901, (2006).

198   Hoffmann, MJ *et al.* Expression changes in EZH2, but not in BMI-1, SIRT1, DNMT1 or DNMT3B are associated with DNA methylation changes in prostate cancer. *Cancer Biol Ther* 6, 1403-1412, (2007).

199   Saramaki, OR *et al.* The gene for polycomb group protein enhancer of zeste homolog 2 (EZH2) is amplified in late-stage prostate cancer. *Genes Chromosomes Cancer* 45, 639-645, (2006).

200   Sanpaolo, E *et al.* EZH2 and ZFX oncogenes in malignant behaviour of parathyroid neoplasms. *Endocrine* 54, 55-59, (2016).

201   Bodor, C *et al.* EZH2 Y641 mutations in follicular lymphoma. *Leukemia* 25, 726-729, (2011).

202   Majer, CR *et al.* A687V EZH2 is a gain-of-function mutation found in lymphoma patients. *FEBS Lett* 586, 3448-3451, (2012).

203   Morin, RD *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* 42, 181-185, (2010).

204   Bracken, AP *et al.* EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J* 22, 5323-5335, (2003).

205   Bracken, AP *et al.* EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J* 22, 5323-5335, (2003).

206   Richter, GH *et al.* EZH2 is a mediator of EWS/FLI1 driven tumor growth and metastasis blocking endothelial and neuro-ectodermal differentiation. *Proc Natl Acad Sci U S A* 106, 5324-5329, (2009).

207   Thul, PJ *et al.* A subcellular map of the human proteome. *Science* 356, (2017).

208   Kunderfranco, P *et al.* ETS transcription factors control transcription of EZH2 and epigenetic silencing of the tumor suppressor gene Nkx3.1 in prostate cancer. *PLoS One* 5, e10547, (2010).

209   Zhao, JC *et al.* Cooperation between Polycomb and androgen receptor during oncogenic transformation. *Genome Res* 22, 322-331, (2012).

210   Yu, J *et al.* A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. *Cancer Res* 67, 10657-10663, (2007).

211   Chen, H *et al.* Down-regulation of human DAB2IP gene expression mediated by polycomb Ezh2 complex and histone deacetylase in prostate cancer. *J Biol Chem* 280, 22437-22444, (2005).

212   Koyanagi, M *et al.* EZH2 and histone 3 trimethyl lysine 27 associated with Il4 and Il13 gene silencing in Th1 cells. *J Biol Chem* 280, 31470-31477, (2005).

213   Batlle, E *et al.* The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat Cell Biol* 2, 84-89, (2000).

214   Fujita, Y *et al.* Hakai, a c-Cbl-like protein, ubiquitinates and induces endocytosis of the E-cadherin complex. *Nat Cell Biol* 4, 222-231, (2002).

215   Hajra, KM & Fearon, ER. Cadherin and catenin alterations in human cancer. *Genes Chromosomes Cancer* 34, 255-268, (2002).

216   Herranz, N *et al.* Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor. *Mol Cell Biol* 28, 4772-4781, (2008).

217   Rubin, MA *et al.* Common gene rearrangements in prostate cancer. *J Clin Oncol* 29, 3659-3668, (2011).

218   Deb, G *et al.* Multifaceted role of EZH2 in breast and prostate tumorigenesis: epigenetics and beyond. *Epigenetics* 8, 464-476, (2013).

219   Bjartell, AS *et al.* Association of cysteine-rich secretory protein 3 and beta-microseminoprotein with outcome after radical prostatectomy. *Clin Cancer Res* 13, 4130-4138, (2007).

220   Hanahan, D & Weinberg, RA. Hallmarks of cancer: the next generation. *Cell* 144, 646-674, (2011).

221   Srebrow, A & Kornblihtt, AR. The connection between splicing and cancer. *J Cell Sci* 119, 2635-2641, (2006).

222   Chen, K *et al.* Alternative Splicing of EZH2 pre-mRNA by SF3B3 Contributes to the Tumorigenic Potential of Renal Cancer. *Clin Cancer Res* 23, 3428-3441, (2017).

223   Grzenda, A *et al.* Functional characterization of EZH2beta reveals the increased complexity of EZH2 isoforms involved in the regulation of mammalian gene expression. *Epigenetics Chromatin* 6, 3, (2013).

224   Shiozawa, Y *et al.* Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. *Nat Commun* 9, 3649, (2018).

225   Daures, M *et al.* A new metabolic gene signature in prostate cancer regulated by JMJD3 and EZH2. *Oncotarget* 9, 23413-23425, (2018).

226   Wu, S *et al.* Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* 402, 832-835, (1999).

227   Thol, F *et al.* Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood* 119, 3578-3584, (2012).

228   Pacheco, TR *et al.* In vivo requirement of the small subunit of U2AF for recognition of a weak 3' splice site. *Mol Cell Biol* 26, 8183-8190, (2006).

229   Takayama, KI. Splicing Factors Have an Essential Role in Prostate Cancer Progression and Androgen Receptor Signaling. *Biomolecules* 9, (2019).

230   Kielkopf, CL *et al.* U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* 18, 1513-1526, (2004).

231   Kralovicova, J *et al.* Identification of U2AF(35)-dependent exons by RNA-Seq reveals a link between 3' splice-site organization and activity of U2AF-related proteins. *Nucleic Acids Res* 43, 3747-3763, (2015).

232   Kralovicova, J & Vorechovsky, I. Alternative splicing of U2AF1 reveals a shared repression mechanism for duplicated exons. *Nucleic Acids Res* 45, 417-434, (2017).

233   Anna, A & Monika, G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet* 59, 253-268, (2018).

234   Hocine, S *et al.* RNA processing and export. *Cold Spring Harb Perspect Biol* 2, a000752, (2010).

235   Agalliu, I *et al.* Germline mutations in the BRCA2 gene and susceptibility to hereditary prostate cancer. *Clin Cancer Res* 13, 839-843, (2007).

236   Cybulski, C *et al.* A novel founder CHEK2 mutation is associated with increased prostate cancer risk. *Cancer Res* 64, 2677-2679, (2004).

237   Dong, X *et al.* Mutations in CHEK2 associated with prostate cancer risk. *Am J Hum Genet* 72, 270-280, (2003).

238   Kote-Jarai, Z *et al.* A recurrent truncating germline mutation in the BRIP1/FANCJ gene and susceptibility to prostate cancer. *Br J Cancer* 100, 426-430, (2009).

239   Leongamornlert, DA *et al.* Germline DNA Repair Gene Mutations in Young-onset Prostate Cancer Cases in the UK: Evidence for a More Extensive Genetic Panel. *Eur Urol* 76, 329-337, (2019).

240   Ronen, A & Glickman, BW. Human DNA repair genes. *Environ Mol Mutagen* 37, 241-283, (2001).

241   Edwards, SM *et al.* Two percent of men with early-onset prostate cancer harbor germline mutations in the BRCA2 gene. *Am J Hum Genet* 72, 1-12, (2003).

242   Pritchard, CC *et al.* DNA-Repair Gene Mutations in Metastatic Prostate Cancer. *N Engl J Med* 375, 1804-1805, (2016).

243   Giri, VN *et al.* Role of Genetic Testing for Inherited Prostate Cancer Risk: Philadelphia Prostate Cancer Consensus Conference 2017. *J Clin Oncol* 36, 414-424, (2018).

244   Karyadi, DM *et al.* Whole exome sequencing in 75 high-risk families with validation and replication in independent case-control studies identifies TANGO2, OR5H14, and CHAD as new prostate cancer susceptibility genes. *Oncotarget* 8, 1495-1507, (2017).

245   Blankenberg, D *et al.* Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol* Chapter 19, Unit 19 10 11-21, (2010).

246   Giardine, B *et al.* Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* 15, 1451-1455, (2005).

247   Robinson, PN *et al.* Strategies for exome and genome sequence data analysis in disease-gene discovery projects. *Clin Genet* 80, 127-132, (2011).

248   Li, LC & Dahiya, R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18, 1427-1431, (2002).

249   Economides, KD & Capecchi, MR. Hoxb13 is required for normal differentiation and secretory function of the ventral prostate. *Development* 130, 2061-2069, (2003).

250   Norris, JD *et al.* The homeodomain protein HOXB13 regulates the cellular response to androgens. *Mol Cell* 36, 405-416, (2009).

251   Brechka, H *et al.* HOXB13 mutations and binding partners in prostate development and cancer: Function, clinical significance, and future directions. *Genes Dis* 4, 75-87, (2017).

252   MacInnis, RJ *et al.* Population-based estimate of prostate cancer risk for carriers of the HOXB13 missense mutation G84E. *PLoS One* 8, e54727, (2013).

253   Beebe-Dimmer, JL *et al.* The HOXB13 G84E Mutation Is Associated with an Increased Risk for Prostate Cancer and Other Malignancies. *Cancer Epidemiol Biomarkers Prev* 24, 1366-1372, (2015).

254   Chandrasekaran, G *et al.* Computational Modeling of complete HOXB13 protein for predicting the functional effect of SNPs and the associated role in hereditary prostate cancer. *Sci Rep* 7, 43830, (2017).

255   Kim, IJ *et al.* HOXB13 regulates the prostate-derived Ets factor: implications for prostate cancer cell invasion. *Int J Oncol* 45, 869-876, (2014).

256   Cardoso, M *et al.* Oncogenic mechanisms of HOXB13 missense mutations in prostate carcinogenesis. *Oncoscience* 3, 288-296, (2016).

257   Lotan, TL *et al.* Somatic molecular subtyping of prostate tumors from HOXB13 G84E carriers. *Oncotarget* 8, 22772-22782, (2017).

258   Benz, CC *et al.* Altered promoter usage characterizes monoallelic transcription arising with ERBB2 amplification in human breast cancers. *Genes Chromosomes Cancer* 45, 983-994, (2006).

259   Ghoshal, K *et al.* HOXB13, a target of DNMT3B, is methylated at an upstream CpG island, and functions as a tumor suppressor in primary colorectal tumors. *PLoS One* 5, e10338, (2010).

260   Acharya, CR *et al.* Mapping eQTL by leveraging multiple tissues and DNA methylation. *BMC Bioinformatics* 18, 455, (2017).

261   Beebe-Dimmer, JL *et al.* Prevalence of the HOXB13 G84E prostate cancer risk allele in men treated with radical prostatectomy. *BJU Int* 113, 830-835, (2014).

262   Smith, SC *et al.* HOXB13 G84E-related familial prostate cancers: a clinical, histologic, and molecular survey. *Am J Surg Pathol* 38, 615-626, (2014).

263   Storebjerg, TM *et al.* Prevalence of the HOXB13 G84E mutation in Danish men undergoing radical prostatectomy and its correlations with prostate cancer risk and aggressiveness. *BJU Int* 118, 646-653, (2016).

264   Zhan, J *et al.* HOXB13 networking with ABCG1/EZH2/Slug mediates metastasis and confers resistance to cisplatin in lung adenocarcinoma patients. *Theranostics* 9, 2084-2099, (2019).

265   Liu, Z *et al.* ATRA inhibits the proliferation of DU145 prostate cancer cells through reducing the methylation level of HOXB13 gene. *PLoS One* 7, e40943, (2012).

266   Xiong, Y *et al.* Long noncoding RNA HOXB13-AS1 regulates HOXB13 gene methylation by interacting with EZH2 in glioma. *Cancer Med* 7, 4718-4728, (2018).

267   Williams, TM *et al.* Range of HOX/TALE superclass associations and protein domain requirements for HOXA13:MEIS interaction. *Dev Biol* 277, 457-471, (2005).

268   Huang, Q *et al.* A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. *Nat Genet* 46, 126-135, (2014).

269   Yu, X *et al.* Foxa1 and Foxa2 interact with the androgen receptor to regulate prostate and epididymal genes differentially. *Ann N Y Acad Sci* 1061, 77-93, (2005).

270   Marusyk, A *et al.* Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer* 12, 323-334, (2012).

271   Yadav, SS *et al.* Intratumor heterogeneity in prostate cancer. *Urol Oncol* 36, 349-360, (2018).

272   Gray, JW *et al.* Molecular cytogenetics: diagnosis and prognostic assessment. *Curr Opin Biotechnol* 3, 623-631, (1992).

273   Gray, JW & Pinkel, D. Molecular cytogenetics in human cancer diagnosis. *Cancer* 69, 1536-1542, (1992).

274   Karan, D *et al.* Decreased androgen-responsive growth of human prostate cancer is associated with increased genetic alterations. *Clin Cancer Res* 7, 3472-3480, (2001).

275   Kim, JH *et al.* Integrative analysis of genomic aberrations associated with prostate cancer progression. *Cancer Res* 67, 8229-8239, (2007).

276   Kallioniemi, A *et al.* Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258, 818-821, (1992).

277   Pinkel, D *et al.* High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 20, 207-211, (1998).

278   Solinas-Toldo, S *et al.* Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 20, 399-407, (1997).

279   Ribeiro, FR *et al.* Comparison of chromosomal and array-based comparative genomic hybridization for the detection of genomic imbalances in primary prostate carcinomas. *Mol Cancer* 5, 33, (2006).

280   Knudson, AG, Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 68, 820-823, (1971).

281   Schulz, WA *et al.* Molecular biology of prostate cancer. *Mol Hum Reprod* 9, 437-448, (2003).

282   Sun, J *et al.* DNA copy number alterations in prostate cancers: a combined analysis of published CGH studies. *Prostate* 67, 692-700, (2007).

283   Matsuyama, H *et al.* Deletion mapping of chromosome 8p in prostate cancer by fluorescence in situ hybridization. *Oncogene* 9, 3071-3076, (1994).

284   Matsuyama, H *et al.* The role of chromosome 8p22 deletion for predicting disease progression and pathological staging in prostate cancer. *Aktuelle Urol* 34, 247-249, (2003).

285   Matsuyama, H *et al.* Deletion mapping of chromosome 8p in prostate cancer by fluorescence in situ hybridization. *Oncogene* 9, 3071-3076, (1994).

286   Jenkins, RB *et al.* Detection of c-myc oncogene amplification and chromosomal anomalies in metastatic prostatic carcinoma by fluorescence in situ hybridization. *Cancer Res* 57, 524-531, (1997).

287   Leube, B *et al.* Refined mapping of allele loss at chromosome 10q23-26 in prostate cancer. *Prostate* 50, 135-144, (2002).

288   Li, J *et al.* PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* 275, 1943-1947, (1997).

289   Abou-Ouf, H *et al.* Combined loss of TFF3 and PTEN is associated with lethal outcome and overall survival in men with prostate cancer. *J Cancer Res Clin Oncol* 145, 1751-1759, (2019).

290   Ye, X *et al.* Expression and significance of PTEN and Claudin-3 in prostate cancer. *Oncol Lett* 17, 5628-5634, (2019).

291   Sun, X *et al.* Frequent somatic mutations of the transcription factor ATBF1 in human prostate cancer. *Nat Genet* 37, 407-412, (2005).

292   Verhagen, PC *et al.* Microdissection, DOP-PCR, and comparative genomic hybridization of paraffin-embedded familial prostate cancers. *Cancer Genet Cytogenet* 122, 43-48, (2000).

293   Rokman, A *et al.* Genetic changes in familial prostate cancer by comparative genomic hybridization. *Prostate* 46, 233-239, (2001).

294   Fitzgerald, L. *Genetics of familial prostate cancer in Tasmania* Molecular Genetics thesis, Univserity of Tasmnia, (2007).

295   Dimova, I *et al.* Genomic markers for ovarian cancer at chromosomes 1, 8 and 17 revealed by array CGH analysis. *Tumori* 95, 357-366, (2009).

296   Rokman, A *et al.* Genetic changes in familial prostate cancer by comparative genomic hybridization. *Prostate* 46, 233-239, (2001).

297   Evans, AJ *et al.* Defining a 0.5-mb region of genomic gain on chromosome 6p22 in bladder cancer by quantitative-multiplex polymerase chain reaction. *Am J Pathol* 164, 285-293, (2004).

298   Orlic, M *et al.* Expression analysis of 6p22 genomic gain in retinoblastoma. *Genes Chromosomes Cancer* 45, 72-82, (2006).

299   Zhang, X *et al.* Eukaryotic Elongation Factor 2 (eEF2) is a Potential Biomarker of Prostate Cancer. *Pathol Oncol Res* 24, 885-890, (2018).

300   Hsieh, CL *et al.* A genome screen of families with multiple cases of prostate cancer: evidence of genetic heterogeneity. *Am J Hum Genet* 69, 148-158, (2001).

301   Wiklund, F *et al.* Genome-wide scan of Swedish families with hereditary prostate cancer: suggestive evidence of linkage at 5q11.2 and 19p13.3. *Prostate* 57, 290-297, (2003).

302   Nakamura, J *et al.* Overexpression of eukaryotic elongation factor eEF2 in gastrointestinal cancers and its involvement in G2/M progression in the cell cycle. *Int J Oncol* 34, 1181-1189, (2009).

303   Oji, Y *et al.* The translation elongation factor eEF2 is a novel tumorassociated antigen overexpressed in various types of cancers. *Int J Oncol* 44, 1461-1469, (2014).

304   Tash, JS *et al.* Gamendazole, an orally active indazole carboxylic acid male contraceptive agent, targets HSP90AB1 (HSP90BETA) and EEF1A1 (eEF1A), and stimulates Il1a transcription in rat Sertoli cells. *Biol Reprod* 78, 1139-1152, (2008).

305   Tomlinson, VA *et al.* Translation elongation factor eEF1A2 is a potential oncoprotein that is overexpressed in two-thirds of breast tumours. *BMC Cancer* 5, 113, (2005).

306   Sun, HG *et al.* Clinical value of eukaryotic elongation factor 2 (eEF2) in non-small cell lung cancer patients. *Asian Pac J Cancer Prev* 14, 6533-6535, (2014).

307   Tan, HY *et al.* Suppression of vascular endothelial growth factor via inactivation of eukaryotic elongation factor 2 by alkaloids in Coptidis rhizome in hepatocellular carcinoma. *Integr Cancer Ther* 13, 425-434, (2014).

308   Meric-Bernstam, F *et al.* Aberrations in translational regulation are associated with poor prognosis in hormone receptor-positive breast cancer. *Breast Cancer Res* 14, R138, (2012).

309   Shi, N *et al.* Eukaryotic elongation factors 2 promotes tumor cell proliferation and correlates with poor prognosis in ovarian cancer. *Tissue Cell* 53, 53-60, (2018).

310   Brognard, J *et al.* Cancer-associated loss-of-function mutations implicate DAPK3 as a tumor-suppressing kinase. *Cancer Res* 71, 3152-3161, (2011).

311   Chen, C & Guo, SQ. [Expressions of DAPK3 and c-Myc in endometrial cancer and their relationship with the patients' prognosis]. *Nan Fang Yi Ke Da Xue Xue Bao* 36, 863-869, (2016).

312   Das, TP *et al.* Activation of AKT negatively regulates the pro-apoptotic function of death-associated protein kinase 3 (DAPK3) in prostate cancer. *Cancer Lett* 377, 134-139, (2016).

313   Petrie, K *et al.* The histone deacetylase 9 gene encodes multiple protein isoforms. *J Biol Chem* 278, 16059-16072, (2003).

314   Childs, EJ *et al.* Association of Common Susceptibility Variants of Pancreatic Cancer in Higher-Risk Patients: A PACGENE Study. *Cancer Epidemiol Biomarkers Prev* 25, 1185-1191, (2016).

315   Tomlins, SA *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310, 644-648, (2005).

316   Wang, G *et al.* Zbtb7a suppresses prostate cancer through repression of a Sox9-dependent pathway for cellular senescence bypass and tumor invasion. *Nat Genet* 45, 739-746, (2013).

317   Razzak, M. Genetics: ZBTB7A suppresses castration-resistant prostate cancer. *Nat Rev Clin Oncol* 10, 427, (2013).

318   Sun, G *et al.* Upregulation of ZBTB7A exhibits a tumor suppressive role in gastric cancer cells. *Mol Med Rep* 17, 2635-2641, (2018).

319   Li, CY *et al.* Integrated analysis of long non-coding RNA competing interactions reveals the potential role in progression of human gastric cancer. *Int J Oncol* 48, 1965-1976, (2016).

320   Sarkar, D *et al.* Epigenetic regulation in human melanoma: past and future. *Epigenetics* 10, 103-121, (2015).

321   Dong, L *et al.* LncRNA TINCR is associated with clinical progression and serves as tumor suppressive role in prostate cancer. *Cancer Manag Res* 10, 2799-2807, (2018).

322   Klassen, SS & Rabkin, SW. Nitric oxide induces gene expression of Jumonji and retinoblastoma 2 protein while reducing expression of atrial natriuretic peptide precursor type B in cardiomyocytes. *Folia Biol (Praha)* 54, 65-70, (2008).

323   Pasini, D *et al.* JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* 464, 306-310, (2010).

324   Son, J *et al.* Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes Dev* 27, 2663-2677, (2013).

325   Cao, J *et al.* Knockdown of JARID2 inhibits the proliferation and invasion of ovarian cancer through the PI3K/Akt signaling pathway. *Mol Med Rep* 16, 3600-3605, (2017).

326   Wang, X *et al.* Jarid2 enhances the progression of bladder cancer through regulating PTEN/AKT signaling. *Life Sci*, (2019).

327   Hammerich-Hille, S *et al.* SAFB1 mediates repression of immune regulators and apoptotic genes in breast cancer cells. *J Biol Chem* 285, 3608-3616, (2010).

328   Li, Y *et al.* HEF1, a novel target of Wnt signaling, promotes colonic cell migration and cancer progression. *Oncogene* 30, 2633-2643, (2011).

329   Tirkkonen, M *et al.* Distinct somatic genetic changes associated with tumor progression in carriers of BRCA1 and BRCA2 germ-line mutations. *Cancer Res* 57, 1222-1227, (1997).

330   Joosse, SA *et al.* Prediction of BRCA2-association in hereditary breast carcinomas using array-CGH. *Breast Cancer Res Treat* 132, 379-389, (2012).

331   Romanel, A *et al.* Inherited determinants of early recurrent somatic mutations in prostate cancer. *Nat Commun* 8, 48, (2017).

332   Zafarana, G *et al.* Copy number alterations of c-MYC and PTEN are prognostic factors for relapse after prostate cancer radiotherapy. *Cancer* 118, 4053-4062, (2012).

333   Carter, NP. Cytogenetic analysis by chromosome painting. *Cytometry* 18, 2-10, (1994).

334   Kallioniemi, OP *et al.* Comparative genomic hybridization: a rapid new method for detecting and mapping DNA amplification in tumors. *Semin Cancer Biol* 4, 41-46, (1993).

335   Hemminki, A *et al.* Localization of a susceptibility locus for Peutz-Jeghers syndrome to 19p using comparative genomic hybridization and targeted linkage analysis. *Nat Genet* 15, 87-90, (1997).

336   Kim, SH *et al.* Genetic alterations in microdissected prostate cancer cells by comparative genomic hybridization. *Prostate Cancer Prostatic Dis* 3, 110-114, (2000).

337   Gu, Y *et al.* NEDD9 overexpression predicts poor prognosis in solid cancers: a meta-analysis. *Onco Targets Ther* 12, 4213-4222, (2019).

338   Sima, N *et al.* The overexpression of scaffolding protein NEDD9 promotes migration and invasion in cervical cancer via tyrosine phosphorylated FAK and SRC. *PLoS One* 8, e74594, (2013).

339   Wang, J *et al.* Overexpression of NEDD9 in renal cell carcinoma is associated with tumor migration and invasion. *Oncol Lett* 14, 8021-8027, (2017).

340   Bamford, S *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer* 91, 355-358, (2004).

341   Petrovics, G *et al.* Frequent overexpression of ETS-related gene-1 (ERG1) in prostate cancer transcriptome. *Oncogene* 24, 3847-3852, (2005).

342   Tomlins, SA *et al.* Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia* 10, 177-188, (2008).

343   Paulo, P *et al.* FLI1 is a novel ETS transcription factor involved in gene fusions in prostate cancer. *Genes Chromosomes Cancer* 51, 240-249, (2012).

344   Hermans, KG *et al.* 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* 66, 10658-10663, (2006).

345   Perner, S *et al.* TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. *Cancer Res* 66, 8337-8341, (2006).

346   Hermans, KG *et al.* Truncated ETV1, fused to novel tissue-specific genes, and full-length ETV1 in prostate cancer. *Cancer Res* 68, 7541-7549, (2008).

347   Newcomb, LF *et al.* Performance of PCA3 and TMPRSS2:ERG urinary biomarkers in prediction of biopsy outcome in the Canary Prostate Active Surveillance Study (PASS). *Prostate Cancer Prostatic Dis*, (2019).

348   Tomlins, SA *et al.* Urine TMPRSS2:ERG fusion transcript stratifies prostate cancer risk in men with elevated serum PSA. *Sci Transl Med* 3, 94ra72, (2011).

349   Newcomb, LF *et al.* Performance of PCA3 and TMPRSS2:ERG urinary biomarkers in prediction of biopsy outcome in the Canary Prostate Active Surveillance Study (PASS). *Prostate Cancer Prostatic Dis*, (2019).

350   Demichelis, F *et al.* TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene* 26, 4596-4599, (2007).

351   Attard, G *et al.* Duplication of the fusion of TMPRSS2 to ERG sequences identifies fatal human prostate cancer. *Oncogene* 27, 253-263, (2008).

352   FitzGerald, LM *et al.* Association of TMPRSS2-ERG gene fusion with clinical characteristics and outcomes: results from a population-based study of prostate cancer. *BMC Cancer* 8, 230, (2008).

353   Gopalan, A *et al.* TMPRSS2-ERG gene fusion is not associated with outcome in patients treated by prostatectomy. *Cancer Res* 69, 1400-1406, (2009).

354   Baena, E *et al.* ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. *Genes Dev* 27, 683-698, (2013).

355   Attard, G *et al.* Prostate cancer. *Lancet* 387, 70-82, (2016).

356   Perner, S *et al.* TMPRSS2-ERG fusion prostate cancer: an early molecular event associated with invasion. *Am J Surg Pathol* 31, 882-888, (2007).

357   Penney, KL *et al.* Association of Prostate Cancer Risk Variants with TMPRSS2:ERG Status: Evidence for Distinct Molecular Subtypes. *Cancer Epidemiol Biomarkers Prev* 25, 745-749, (2016).

358   Luedeke, M *et al.* Prostate cancer risk regions at 8q24 and 17q24 are differentially associated with somatic TMPRSS2:ERG fusion status. *Hum Mol Genet* 25, 5490-5499, (2016).

359   Schaefer, G *et al.* Distinct ERG rearrangement prevalence in prostate cancer: higher frequency in young age and in low PSA prostate cancer. *Prostate Cancer Prostatic Dis* 16, 132-138, (2013).

360   Weischenfeldt, J *et al.* Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. *Cancer Cell* 23, 159-170, (2013).

361   Luedeke, M *et al.* Predisposition for TMPRSS2-ERG fusion in prostate cancer by variants in DNA repair genes. *Cancer Epidemiol Biomarkers Prev* 18, 3030-3035, (2009).

362   Gasi, D *et al.* Overexpression of full-length ETV1 transcripts in clinical prostate cancer due to gene translocation. *PLoS One* 6, e16332, (2011).

363   Teixeira, MR. Chromosome mechanisms giving rise to the TMPRSS2-ERG fusion oncogene in prostate cancer and HGPIN lesions. *Am J Surg Pathol* 32, 642-644; author reply 644, (2008).

364 Yoshimoto, M *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* 8, 465-469, (2006).

365 Vaarala, MH *et al.* 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* 94, 705-710, (2001).

366 Vaarala, MH *et al.* Expression of transmembrane serine protease TMPRSS2 in mouse and human tissues. *J Pathol* 193, 134-140, (2001).

367 Soller, MJ *et al.* Confirmation of the high frequency of the TMPRSS2/ERG fusion gene in prostate cancer. *Genes Chromosomes Cancer* 45, 717-719, (2006).

368 Winnes, M *et al.* Molecular genetic analyses of the TMPRSS2-ERG and TMPRSS2-ETV1 gene fusions in 50 cases of prostate cancer. *Oncol Rep* 17, 1033-1036, (2007).

369 Mehra, R *et al.* Comprehensive assessment of TMPRSS2 and ETS family gene aberrations in clinically localized prostate cancer. *Mod Pathol* 20, 538-544, (2007).

370 Barros-Silva, JD *et al.* Novel 5' fusion partners of ETV1 and ETV4 in prostate cancer. *Neoplasia* 15, 720-726, (2013).

371 Tomlins, SA *et al.* TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. *Cancer Res* 66, 3396-3400, (2006).

372 Clark, J *et al.* Complex patterns of ETS gene alteration arise during cancer development in the human prostate. *Oncogene* 27, 1993-2003, (2008).

373 Svensson, MA *et al.* Testing mutual exclusivity of ETS rearranged prostate cancer. *Lab Invest* 91, 404-412, (2011).

374 Kluth, M *et al.* Deletion of 3p13 is a late event linked to progression of TMPRSS2:ERG fusion prostate cancer. *Cancer Manag Res* 10, 5909-5917, (2018).

375 Brenner, JC *et al.* Mechanistic rationale for inhibition of poly(ADP-ribose) polymerase in ETS gene fusion-positive prostate cancer. *Cancer Cell* 19, 664-678, (2011).

376 Caulfield, SE *et al.* Olaparib: A Novel Therapy for Metastatic Breast Cancer in Patients With a BRCA1/2 Mutation. *J Adv Pract Oncol* 10, 167-174, (2019).

377 Rahim, S *et al.* YK-4-279 inhibits ERG and ETV1 mediated prostate cancer cell invasion. *PLoS One* 6, e19343, (2011).

378 Rahim, S *et al.* A small molecule inhibitor of ETV1, YK-4-279, prevents prostate cancer growth and metastasis in a mouse xenograft model. *PLoS One* 9, e114260, (2014).

379 Luedeke, M *et al.* Prostate cancer risk regions at 8q24 and 17q24 are differentially associated with somatic TMPRSS2:ERG fusion status. *Hum Mol Genet* 25, 5490-5499, (2016).

380 Petrovics, G *et al.* Frequent overexpression of ETS-related gene-1 (ERG1) in prostate cancer transcriptome. *Oncogene* 24, 3847-3852, (2005).

381 Boutros, PC *et al.* Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat Genet* 47, 736-745, (2015).

382 Cooper, CS *et al.* Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat Genet* 47, 367-372, (2015).

383 Karavitakis, M *et al.* Tumor focality in prostate cancer: implications for focal therapy. *Nat Rev Clin Oncol* 8, 48-55, (2011).

384 Cyll, K *et al.* Tumour heterogeneity poses a significant challenge to cancer biomarker research. *Br J Cancer* 117, 367-375, (2017).

385 Besson, P *et al.* How do voltage-gated sodium channels enhance migration and invasiveness in cancer cells? *Biochim Biophys Acta* 1848, 2493-2501, (2015).

386 Chen, D *et al.* RYBP stabilizes p53 by modulating MDM2. *EMBO Rep* 10, 166-172, (2009).

387 Takayama, K *et al.* Integrative analysis of FOXP1 function reveals a tumor-suppressive effect in prostate cancer. *Mol Endocrinol* 28, 2012-2024, (2014).

388 Put, N *et al.* FOXP1 and PAX5 are rare but recurrent translocations partners in acute lymphoblastic leukemia. *Cancer Genet* 204, 462-464, (2011).

389 Krohn, A *et al.* Recurrent deletion of 3p13 targets multiple tumour suppressor genes and defines a distinct subgroup of aggressive ERG fusion-positive prostate cancers. *J Pathol* 231, 130-141, (2013).

390   Linn, DE *et al.* Deletion of Interstitial Genes between TMPRSS2 and ERG Promotes Prostate Cancer Progression. *Cancer Res* 76, 1869-1881, (2016).

391   Kunju, LP *et al.* Novel RNA hybridization method for the in situ detection of ETV1, ETV4, and ETV5 gene fusions in prostate cancer. *Appl Immunohistochem Mol Morphol* 22, e32-40, (2014).

392   Qi, M *et al.* Overexpression of ETV4 is associated with poor prognosis in prostate cancer: involvement of uPA/uPAR and MMPs. *Tumour Biol* 36, 3565-3572, (2015).

393   Hofer, MD *et al.* Genome-wide linkage analysis of TMPRSS2-ERG fusion in familial prostate cancer. *Cancer Res* 69, 640-646, (2009).

394   Luedeke, M *et al.* Predisposition for TMPRSS2-ERG fusion in prostate cancer by variants in DNA repair genes. *Cancer Epidemiol Biomarkers Prev* 18, 3030-3035, (2009).

395   Penney, KL *et al.* Association of Prostate Cancer Risk Variants with TMPRSS2:ERG Status: Evidence for Distinct Molecular Subtypes. *Cancer Epidemiol Biomarkers Prev* 25, 745-749, (2016).

396   Ewing, CM *et al.* Germline mutations in HOXB13 and prostate-cancer risk. *N Engl J Med* 366, 141-149, (2012).

397   Wiggans, AJ *et al.* Poly(ADP-ribose) polymerase (PARP) inhibitors for the treatment of ovarian cancer. *Cochrane Database Syst Rev*, CD007929, (2015).

398   Marshall, CH *et al.* Differential Response to Olaparib Treatment Among Men with Metastatic Castration-resistant Prostate Cancer Harboring BRCA1 or BRCA2 Versus ATM Mutations. *Eur Urol* 76, 452-458, (2019).

399   Rahim, S *et al.* YK-4-279 inhibits ERG and ETV1 mediated prostate cancer cell invasion. *PLoS One* 6, e19343, (2011).

400   Rahim, S *et al.* A small molecule inhibitor of ETV1, YK-4-279, prevents prostate cancer growth and metastasis in a mouse xenograft model. *PLoS One* 9, e114260, (2014).

401   Mori, R *et al.* Both beta-actin and GAPDH are useful reference genes for normalization of quantitative RT-PCR in human FFPE tissue samples of prostate cancer. *Prostate* 68, 1555-1560, (2008).

402   Ma, XJ *et al.* A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 5, 607-616, (2004).

403   Uhlen, M *et al.* Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419, (2015).

404   Scott, A *et al.* Functional analysis of synonymous substitutions predicted to affect splicing of the CFTR gene. *J Cyst Fibros* 11, 511-517, (2012).

405   Listerman, I *et al.* The major reverse transcriptase-incompetent splice variant of the human telomerase protein inhibits telomerase activity but protects from apoptosis. *Cancer Res* 73, 2817-2828, (2013).

406   Liu, Z *et al.* ATRA inhibits the proliferation of DU145 prostate cancer cells through reducing the methylation level of HOXB13 gene. *PLoS One* 7, e40943, (2012).

407   Geli, J *et al.* Deletions and altered expression of the RIZ1 tumour suppressor gene in 1p36 in pheochromocytomas and abdominal paragangliomas. *Int J Oncol* 26, 1385-1391, (2005).

408   Yang, G & Hurlin, PJ. MNT and Emerging Concepts of MNT-MYC Antagonism. *Genes (Basel)* 8, (2017).

409   Redon, S *et al.* Protein RNA and protein protein interactions mediate association of human EST1A/SMG6 with telomerase. *Nucleic Acids Res* 35, 7011-7022, (2007).

410   Luo, W *et al.* Characteristics of genomic alterations of lung adenocarcinoma in young never-smokers. *Int J Cancer*, (2018).

411   Lu, B *et al.* Detection of TMPRSS2-ERG fusion gene expression in prostate cancer specimens by a novel assay using branched DNA. *Urology* 74, 1156-1161, (2009).

CHAPTER 10 : **APPENDICES**

APPENDIX 1

Thermal cycling conditions.

- **cDNA synthesis (Chapter 2.1.7)**

    20uL reactions, per sample:

    > 4.0μL 5X VILO^{TM} Reaction Mix (Invitrogen)
    >
    > 2.0μL 10X SuperScript^{TM} Enzyme Mix (Invitrogen)
    >
    > XμL FFPE RNA
    >
    > H$_2$O to 20uL final volume
    >
    > > X – Variable, up to 2.5μg

    Thermal cycling conditions:

    > 25 °C – 10 minutes
    >
    > 42 °C – 120 minutes
    >
    > 85 °C – 5 minutes

- **Amplification of DNA for Sanger sequencing (Chapter 2.2)**

    10μL reactions, per sample:

    > 5.0μL MyTaq^{TM} HS Mastermix (Bioline)
    >
    > 0.8μL forward primer at 10μM (Sigma-Aldrich)
    >
    > 0.8μL reverse primer at 10μM (Signma-Aldrich)
    >
    > 2.4μL H$_2$O
    >
    > 1.0μL DNA at 10ng/μL

    Thermal cycling conditions:

    > 95°C – 1 minute
    >
    > 95°C – 10 seconds ⎤
    >
    > X°C – 10 seconds   ⎬ 40 cycles
    >
    > 72°C – 20 seconds * ⎦
    >
    > 4°C – ∞
    >
    > > X – Annealing temperature is primer pair specific (Appendix 2)
    > >
    > > * Extension time increased to 30 seconds for larger fragments

- **Quantification of gene expression by RT-qPCR (Chapter 2.3)**

  10μL reactions, per sample:

  >5.0μL SensiFAST^{TM} SYBR$^®$ No-Rox Mastermix (Bioline) or

  >5.0 μL PowerUp^{TM} SYBR$^®$ Green Mastermix (ThermoFisher Scientific)

  >0.3μL forward primer at 10μM (Sigma-Aldrich)

  >0.3μL reverse primer at 10μM (Sigma-Aldrich)

  >3.4μL H$_2$O

  >1.0μL FFPE cDNA at 50ng/μL

  Thermal cycling was conducted on the Rotor Gene 6000 (Qiagen) when using SensiFAST^{TM} SYBR$^®$ (Chapter 5), as per the following conditions:

  >95°C – 3 minutes

  >95°C – 10 seconds

  >60°C – 10 seconds   40 cycles

  >72°C – 10 seconds

  Thermal cycling was conducted on the QuantStudio^{TM} 3 Real-Time PCR System when using PowerUp^{TM} SYBR Green (Chapters 4, 6 and 7), as per the following conditions:

  >50°C – 2 minutes

  >95°C – 2 minutes

  >95°C – 1 second

  >60°C – 20 seconds   40 cycles

- **Big Dye Terminator sequencing reaction (Chapter 3.2.3)**
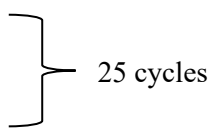
  10μL reactions, per sample:

  >0.25μL BigDye$^®$ Terminator v3.1 Ready Reaction Mix

  >1.75μL 5X Sequencing Buffer

  >1.6μL primer at 3.3μM (forward or reverse) (Sigma-Aldrich)

  >5.4μL H$_2$O

  >~ 1.0μL AMPure purified PCR product *

  >>*Variable depending on concentration and size of the PCR fragment

Thermal cycling conditions:

96°C – 1 minute

96°C – 10 seconds ⎱
50°C – 5 seconds   ⎰ 25 cycles
60°C – 1 minute 15 seconds

4°C - ∞
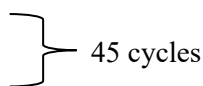
- **TaqMan® SNP genotyping (Chapter 3.2.4)**

8μL reactions, per sample:

4.0μL SensiFAST™ Probe No-Rox Mastermix (Bioline)

0.1μL 40x TaqMan® SNP genotyping probe (Applied Biosystems; Appendix 6)

2.9μL H_2O

1.0μL gDNA at 10ng/μL

Thermal cycling was conducted on the LightCycler® 480 system (Roche), as per the following conditions:

95°C – 10 minutes

95°C – 15 seconds ⎱
60°C – 1 minute   ⎰ 45 cycles
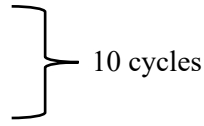
- **Allele-specific next-generation sequencing (Chapter 5.2.4)**

10μL reaction, per sample:

5μL Phusion® (GeneSearch)

1μL forward tag (Integrated DNA Technologies; Appendix 13)

1μL reverse tag (Integrated DNA Technologies; Appendix 13)

1μL H_2O

Thermal cycling was conducted on the Mastercycler® nexus (Eppendorf), as per the following conditions:

98°C – 5 seconds   ⎱
98°C – 20 seconds   ⎰ 10 cycles
72°C – 20 seconds

- **Allele-specific methylation analysis (Chapter 5.2.6)**

  10μL reaction, per sample:

  > 5.0μL MyTaq^{TM} HS Mastermix (Bioline)
  >
  > 0.8μL forward primer at 10μM (Sigma Aldrich)
  >
  > 0.8μL reverse primer at 10μM (Sigma Aldrich)
  >
  > 2.0μL Q solution (Qiagen)
  >
  > 0.4μL H_2O
  >
  > 1.0μL bisulphite-converted DNA @ 25ng

  Thermal cycling conditions:

  > 95°C – 2 minutes
  >
  > 95°C – 10 seconds
  >
  > 56°C – 10 seconds   45 cycles
  >
  > 64°C – 30 seconds
  >
  > 4°C - ∞

- **TaqMan® *TMPRSS2:ERG* expression assay (Chapter 7.2.2)**

  10μL reactions, per sample:

  > 5.0μL TaqMan® Fast Advanced Mastermix (ThermoFisher Scientific)
  >
  > 0.5μL TaqMan® expression probe (Applied Biosystems; Appendix 26)
  >
  > 2.5μL H_2O
  >
  > 2.0μL FFPE cDNA at 50ng/μL

  Thermal cycling was conducted on the QuantStudio^{TM} 3 Real-Time PCR System, as per the following conditions:

  > 50°C – 2 minutes
  >
  > 95°C – 2 minutes
  >
  > 95°C – 1 second
  >
  > 60°C – 20 seconds   40 cycles

APPENDIX 2

Primers designed for Sanger sequencing of prioritised rare variants and their optimal annealing temperatures.

All sequencing primers were designed using Primer3 [135] or PrimerBLAST [136].

	Gene	Variant	Allele Change	Forward Primer (5'-3')	Reverse Primer (5'-3')	Product size (bp)	Optimal annealing temperature (°C)
Genomic DNA	*CCL26*	rs41463245	C > T	CATCCCAAGGCTCATCCTG	CTGCTTCTGTTCCCAACCAC	500	64
	P2RX7	rs28360447	G > A	ATGATGTCCCTCCTGGAGAA	ATGGCCCTCCCAGAGATACT	354	62
	NDE1	rs113493697	C > T	CTCCCAAAGTGCTGGCATTA	GCTCTGAGCCTGATGCAAAT	366	60
	CLDN4	Novel	A > G	CTGGTCTGCTCACACTTGCT	AGAGAGGCTGAAGGCTGCTG	969	66
	ATM	rs1800057	C > G	TGGCAAGGTGAGTATGTTGG	TACTGCCATCTGCAGCATTC	526	64
	SSH3	rs373641394	G > A	CAATGATGATGCAGCAGAGG	AGCAGGGTCACTGGGATATG	336	64
	IRS1	rs41265094	C > G	GGCCAGACAAGTAGCCAGAC	TCTTCCTCTTCCACCAGCAG	316	64
	CRIP2	rs375691223	C > T	CTCCCTCCACAGGAGTGAAC	GATTCGGACACGCAGACAC	320	64
	KMT2C	rs76844681	C > T	GGAGTCAAAGAGGAAGGTAAGAAA	TACATAGGGCCGTGGGTCT	337	64
	RHPN2	Novel	A > G	ACTCAACCCCAAACCTGATG	GAGGGCACTTCTCTCCCTCT	315	64
	HSD3B1	rs4986952	G > T	TTTTTGGTTCTAGAATTTCACATCA	TGCCCTTCTTTGTGATCCTT	443	66
	NAT10	rs72910804	A > G	CCCTCTGTCCTTTCTGCTGT	AGGGGACTCTCAAAGGGAAG	380	66
	RND1	Novel	C > A	GGGCCATATTTCAAGCTGTC	CTCATGGGCAGGAAAATGAT	393	58

	Gene	Variant	Allele Change	Forward Primer (5'-3')	Reverse Primer (5'-3')	Product size (bp)	Optimal annealing temperature (°C)
	WNT1	Novel	G > A	GGAGAGGGCAGTGTCTGG	CGGGCGACGAGCTGTTAC	410	66
	CHEK2	rs200432447	G > C	CCAGGTTCCATCAGGTTTTT	TGAGATGGGAGAGAAACAGATG	369	62
	ITGAD	rs147321998	C > T	ATGTGAGGGTGCCAGGACT	CTGAAGGAGATGCAGGCTGA	314	60
	EZH2	rs78589034	G > A	CTGGGATTGCAGGAGTCG	TTTGTCCCCAGTCCATTTTC	365	60
	EPS8	rs78763451	C > T	ATGCAGTCTGTGCCCTTATG	GACTAGAGAAGAGCCAGGGAGTT	493	64
	TIA1	rs115611153	T > C	CGCTTTACATAAGAGGCCCTA	TGATGGCCCTGTGTGTTTT	355	62
	HOXB13	rs138213197	C > G	CACAACGGTCCCTCTTGTCT	GTTCAGCGGACGTAAGCG	696	62
FFPE DNA	*EZH2*	rs78589034	G > A	CAGATGGTGCCAGCAATAGA	TGAAGCTGTGTGCCCAATTA	170	60
	HOXB13	rs138213197	C > G	CCGGATAGAAGGCAAACTCA	GCTGATGCCTGCTGTCAACT	272	62

APPENDIX 3

Primers designed for gene expression analysis in prostate tissue specimens by RT-qPCR.

This table also details the most commonly transcribed isoform in the prostate, for each gene and the median TPM (transcripts per million) expression of 152 prostate samples from the GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2; https://gtexportal.org/home/) [137].

Gene	Transcript	GTEx TPM Expression [137]	Forward Primer (5'-3')	Reverse Primer (5'-3')	Product size (bp)
β-Actin	ENST00000331789.5	3095	GAGCGCGGATACAGCTT [401]	TCCTTAATGTCACGCACGATTT [401]	59
GAPDH	ENST00000396859.1	786.1	CAACGGATTTGGTCGTATTGG [401]	GCAACAATATCCACTTTACCAGAGTTAA [401]	72
EZH2 Exon 17	ENST00000492143.1	3.220	AAGCACAGTGCAACACCAAG	AGCGGCTCCACAAGTAAGAC	86
EZH2 Exon 4/5	ENST00000492143.1	3.220	GCGACTGAGACAGCTCAAGA	CCAAAATTTTCTGACGATTGGAACT	80
EZH2 Exon 8/9	ENST00000492143.1	3.220	ATGGGAAAGTACACGGGGATAG	GGCATTCACCAACTCCACAAAA	71
EZH2 Exon 12/14	ENST00000492143.1	3.220	GGACCACAGTGTTACCAGCA	TTGGTGGGGTCTTTATCCGC	82
EZH2 Exon 14/16	ENST00000492143.1	3.220	GAGGAAACACCGGTTGTGGG	TGTAAACATGGTTAGAGGAGCCG	77
EZH2 Exon 17/18	ENST00000492143.1	3.220	TATTCAGCGGGGCTCCAAAA	GATAAAAATCCCCCAGCCTGC	70
EZH2 Exon 20/21	ENST00000492143.1	3.220	TTCGGTAAATCCAAACTGCTATGC	CCAGTCTGGATGGCTCTCTTG	90
GTEx TPM: Median transcripts per million expression of 152 prostate samples from the GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2) [137].					

Gene	Transcript	GTEx TPM Expression [137]	Forward Primer (5'-3')	Reverse Primer (5'-3')	Product size (bp)
CDH1	ENST00000261769.5	58.33	AAGGGGTCTGTCATGGAAGG	GGTGTTCACATCATCGTCCG	84
MSMB	ENST00000358559.2	1160	TGATCTTTGCCACCTTCGTGA	ACAGGTGTAGAAACATCCTGGTT	99
HOXA9	ENST00000343483.6	26.19	ATCCCAATAACCCAGCAGCC	TTTGTATAGGGGCACCGCTT	70
SF3B1	ENST00000424674.1	145.1	TTGTTGGTCGTATTGCTGACA	TCAAAGCAAATCCTCATCCACTC	70
SF3B3	ENST00000565990.2	34.10	GCATCCTTGTGCCATTCACG	TCAGACCGCAGGTGCATTTC	73
U2AF1	ENST00000291552.4	71.93	TGTGGAGATGCAGGAACACT	ACTTCCCCATACTTCTCCTCC	75
HOXB13	ENST00000290295.7	114.5	TTCATCCTGACAGTGGCAATAATC [402]	CTAGATAGAAAATATGAGGCTAACGATCAT [402]	77
EEF2 5'UTR/Exon 2	ENST00000309311.6	1213	CGACTCGCTTCTTTCGGTTC	CGGATCTGGTCTACCGTGAAG	88
EEF2 Exon 2/3	ENST00000309311.6	1213	AGACACGCTTCACTGATACCC	AGGGAGATGGCAGTTGACTTG	73
EEF2 Exon 4/5	ENST00000309311.6	1213	ATCATCTCCACCTACGGCGA	CGGTACCGAGGACAGGATCG	73
EEF2 Exon 9/10	ENST00000309311.6	1213	GAGGACCTCTACCTGAAGCC	CCACAAGGCACATCCTCGAT	83
EEF2 Exon 14/15	ENST00000309311.6	1213	AAGGCCTATCTGCCCGTCAA	AAGGCCTATCTGCCCGTCAA	89
DAPK3 Exon 3/4	ENST00000301264.3	68.91	ATGTCCACGTTCAGGCAGG	CTTCCGCACGATCGCAAAC	87
DAPK3 Exon 4/5	ENST00000301264.3	68.91	GCGTTCACTACCTGCACTCTA	ACGTTCTTGTCCAGCAGCAT	79
GTEx TPM: Median transcripts per million expression of 152 prostate samples from the GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2) [137].					

Gene	Transcript	GTEx TPM Expression [137]	Forward Primer (5'-3')	Reverse Primer (5'-3')	Product size (bp)
DAPK3 Exon 7/8	ENST00000301264.3	68.91	CTATATCCTCCTGAGCGGTGC	TTCACGGCTGAGATGTTGGT	78
ETV1 Exon 8/10	ENST00000405358.4	2.885	AACAGAGATCTGGCTCATGATTC	CTTCTGCAAGCCATGTTTCCTG	76
ETV1 Exon 16/17	ENST00000405358.4	2.885	GATAGCAGCTACCCCATGGAC	TCGTCGGCAAAGGAGGAAAG	79
ETV1 Exon 20/21	ENST00000405358.4	2.885	GACTGGTCGAGGCATGGAAT	TTTCTGAATGCCCCAACGTC	70
GTEx TPM: Median transcripts per million expression of 152 prostate samples from the GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2) [137].					

APPENDIX 4

Standard curves for each RT-qPCR primer pair.

The lines of best fit were used to calculate the copy number of each gene in each sample and can be used to calculate the PCR efficiency.

β- *Actin* – SensiFAST™ SYBR® No-Rox

y = -3.2387x + 33.569
R² = 0.99998

β- *Actin* – PowerUP™ SYBR® Green

y = -3.06x + 39.293
R² = 0.9961

GAPDH – SensiFAST™ SYBR® No-Rox

y = -3.154x + 32.945
R² = 0.99883

GAPDH – PowerUP™ SYBR® Green

$y = -3.1559x + 42.259$
$R^2 = 0.9981$

EZH2 Exon 17 – PowerUP™ SYBR® Green

$y = -3.487x + 36.609$
$R^2 = 0.99928$

EZH2 Exon 4/5 – PowerUP™ SYBR® Green

$y = -2.9427x + 43.987$
$R^2 = 0.9871$

EZH2 Exon 8/9 – PowerUP™ SYBR® Green

$y = -2.9032x + 46.641$
$R^2 = 0.9885$

260

EZH2 Exon 12/14 – PowerUP™ SYBR® Green

y = -3.7622x + 49.966
R² = 0.9826

EZH2 Exon 14/16 – PowerUP™ SYBR® Green

y = -2.9355x + 42.651
R² = 0.9899

EZH2 Exon 17/18 – PowerUP™ SYBR® Green

y = -2.6718x + 40.814
R² = 0.9931

EZH2 Exon 20/21 – PowerUP™ SYBR® Green

y = -2.1785x + 42.616
R² = 0.9845

CDH1 – PowerUP™ SYBR® Green

$y = -2.8253x + 35.039$
$R^2 = 0.9735$

MSMB – PowerUP™ SYBR® Green

$y = -3.382x + 33.506$
$R^2 = 0.9887$

HOXA9 – PowerUP™ SYBR® Green

$y = -3.304x + 34.004$
$R^2 = 0.9993$

SF3B1 – PowerUP™ SYBR® Green

$y = -3.0263x + 41.844$
$R^2 = 0.9968$

SF3B3 – PowerUP™ SYBR® Green

$y = -2.506x + 28.25$
$R^2 = 0.9813$

U2AF1 – PowerUP™ SYBR® Green

$y = -2.5179x + 44.256$
$R^2 = 0.9714$

HOXB13 – SensiFAST™ SYBR® No-Rox

$y = -3.339x + 35.321$
$R^2 = 0.99654$

EEF2 5'UTR/Exon 2 – PowerUP™ SYBR® Green

$y = -2.2907x + 40.164$
$R^2 = 0.9924$

EEF2 Exon 2/3 – PowerUP™ SYBR® Green

y = -3.9153x + 50.471
R² = 0.9839

EEF2 Exon 4/5 – PowerUP™ SYBR® Green

y = -2.278x + 40.822
R² = 0.9976

EEF2 Exon 9/10 – PowerUP™ SYBR® Green

y = -3.1199x + 46.174
R² = 0.9804

EEF2 Exon 14/15 – PowerUP™ SYBR® Green

y = -3.381x + 47.617
R² = 0.9995

264

DAPK3 Exon 3/4 – PowerUP™ SYBR® Green

$y = -3.8966x + 48.642$
$R^2 = 0.9698$

DAPK3 Exon 4/5 – PowerUP™ SYBR® Green

$y = -4.5888x + 52.498$
$R^2 = 0.9865$

DAPK3 Exon 7/8 – PowerUP™ SYBR® Green

$y = -3.402x + 49.091$
$R^2 = 0.9707$

ETV1 Exon 8/10 – PowerUP™ SYBR® Green

$y = -2.9222x + 44.474$
$R^2 = 0.9729$

ETV1 Exon 16/17 – PowerUP™ SYBR® Green

y = -4.559x + 52.721
R² = 0.9834

ETV1 Exon 20/21 – PowerUP™ SYBR® Green

y = -3.5763x + 48.728
R² = 0.9942

266

APPENDIX 5

Primary antibodies used for protein expression analyses.

This table details the chosen primary antibody and the expression of the protein in the prostate, as per the Human Protein Atlas (https://www.proteinatlas.org) [403].

Protein	Antibody	Working Dilution	Immunogen	Expression at the protein level [403]
EZH2	ab186006 (abcam)	1:150	Amino acid 696-746	Nuclear expression in the testis, lymphoid tissues and gastrointestinal tract. Not detected in the prostate, as reported by the Human Protein Atlas, however 4 of 11 PCa patients had moderate/strong staining of EZH2.
HOXB13	sc-28333 (Santa Cruz)	1:50	Amino acid 1-284	Nuclear expression in the prostate and gastrointestinal tract. Highly expressed in the prostate, as reported by the Human Protein Atlas and 10 of 12 PCa patients had moderate/strong staining of HOXB13.
ETV1	PA5-41484 (ThermoFisher)	1:150	'Middle region'	Localised to the nucleoplasm. Expression of the protein in any tissue is not reported by the Human Protein Atlas.
EEF2	SAB4500695 (Sigma-Aldrich)	1:150	Amino acid 31-80	Cytoplasmic and membranous expression in most tissues. Highly expressed in the prostate, as reported by the Human Protein Atlas and 10 of 11 PCa patients had moderate/strong staining of EEF2.

APPENDIX 6

TaqMan® SNP genotyping assay identification numbers (Applied Biosystems).

Gene	Assay Identification
CCL26	C_86323013_10
P2RX7	C_59964848_10
ATM	C_45273750_10
RND1	Custom Designed Probe
WNT1	Custom Designed Probe
ITGAD	C_164249153_10
EZH2	C_64633016_10
HOXB13	C_164436492_10

APPENDIX 7

Primers designed for the *EZH2 in vitro* splicing assay.

	Forward Primer (5'-3')	Reverse Primer (5'-3')	Amplicon Size (bp)
EZH2 insert (exon 16-19)	AGAGCACCTTGCTGAACGAT	CTGTCAACAGCAGGGTGAGA	3,181
EZH2 insert with attB1 and attB2 attachment sites [186]	GGGG-ACAAGTTTCTACAAAAAAGCAGGCT-AGAGCACCTTGCTGAACGAT	GGGG-ACCACTTTGTACAAGAAAGCTGGGT-CTGTCAACAGCAGGGTGAGA	3,239
Rat insulin exon 2 (forward) [404] and 3(reverse) [405]	CCTGCTCATCCTCTGGGAGC	ATGCTGGTGCAGCACTGAT	253; 717*
*The amplicon size is dependent on whether the insert is present or not. If present, the product would be 717bp and if absent, the product would be 253bp.			

APPENDIX 8

EZH2 gene expression analysis in prostate cancer cell lines and prostate needle biopsy samples (raw data).

Sample Identification	Absolute *EZH2* Gene Expression						
	Exon 17	Exon 4/5	Exon 8/9	Exon 12/14	Exon 14/16	Exon 17/18	Exon 20/21
PC3	0.06	0.44	11.08	1.21	0.15	0.16	18.33
22Rv1	0.22	4.07	32.24	3.84	0.17	0.25	39.16
LNCaP	0.57	0.12	8.18	0.80	0.10	0.18	34.94
PT0001 Right	0.05	0.01	1.11	0.23	0.01	0.03	9.73
PT0001 Left	0.02	0.04	0.75	0.36	0.03	0.01	0.60
PT0002 Right	0.09	0.02	1.68	0.43	0.03	0.06	17.86
PT0002 Left	0.11	0.02	0.74	0.26	0.02	0.02	5.80
PT0003 Right	0.03	0.01	0.32	0.13	0.01	0.01	1.67
PT0003 Left	0.06	0.07	1.91	0.40	0.03	0.04	4.92
PT0018 Right	0.03	0.02	0.91	0.36	0.02	0.03	1.45
PT0018 Left	0.03	0.03	0.80	0.42	0.02	0.02	1.89

APPENDIX 9

EZH2 target gene and splicing factor expression analysis in prostate cancer cell lines and prostate needle biopsy samples (raw data).

Sample Identification	Absolute Gene Expression					
	CDH1	*HOXA9*	*MSMB*	*SF3B1*	*SF3B3*	*U2AF1*
PC3	4.74	0.04	0.01	0.35	1.06	140.67
22Rv1	0.30	0.01	0.03	0.31	0.61	28.69
LNCaP	1.07	0.01	0.04	0.29	1.23	59.00
PT0001 Right	2.06	0.03	0.15	0.73	1.07	48.82
PT0001 Left	1.31	0.02	0.02	0.40	0.63	23.52
PT0002 Right	1.58	0.16	0.02	1.09	0.62	55.74
PT0002 Left	2.71	0.06	0.11	0.81	0.51	35.32
PT0003 Right	3.80	0.02	0.52	0.73	0.80	104.43
PT0003 Left	4.10	0.04	0.32	0.73	1.08	75.31
PT0018 Right	2.89	0.02	0.12	0.56	0.53	16.73

APPENDIX 10

CDH1, MSMB and *U2AF1* expression analysis in FFPE prostate tissue samples (raw data).

	Sample Identification	Tissue Cell Type	*CDH1* Absolute Gene Expression	*MSMB* Absolute Gene Expression	*U2AF1* Absolute Gene Expression
EZH2 variant non-carrier	PC4-03	Malignant	7.17	1.30	101.18
		Benign	6.58	1.80	20.54
	PC11-11	Malignant	5.93	0.23	32.34
		Benign	1.71	0.43	18.45
	PC12-07	Malignant	1.37	0.02	6.42
	PC19-02	Malignant	19.39	4.20	18.94
	PC60-01	Malignant	14.96	1.26	19.11
		Benign	3.86	1.45	16.82
	PC72-04	Malignant	1.24	0.14	22.34
		Benign	201.94	8.30	11.01
	PC72-06	Malignant	7.99	3.04	8.12
		Benign	22.86	3.08	37.24
	PC3250-01	Malignant	1.21	1.86	76.92
		Benign	16.73	21.54	29.61
	DVA 216	Malignant	8.07	18.33	18.95
		Benign	4373.71	11.17	27.34
	DVA 402	Malignant	8.09	18.17	12.54
		Benign	2.87	1.70	37.24
	DVA 1002	Malignant	46.19	126.73	14.84
		Benign	548.53	208.53	-
EZH2 variant carrier	PC12-01	Malignant	4.80	2.24	1514.76
		Benign	35.58	1.35	6.32
	PC12-03	Malignant	397.72	5.25	2.25
		Benign	0.95	1.88	5.64
	PC12-06	Malignant	19.68	14.10	24.50
	PC12-08	Malignant	6.83	8.00	12.51
	PC12-09	Malignant	2.52	0.10	6.19
		Benign	2.02	0.28	13.94
	PC12-132	Malignant	12.91	1.90	23.50
		Benign	46.94	17.63	18.03
	DVA 416	Malignant	134.76	17.18	-
		Benign	162.64	0.41	11.44

APPENDIX 11

Gene panel of cancer predisposition and DNA repair genes.

Genes include known prostate, breast and ovarian cancer predisposition genes and DNA repair genes commonly disrupted in cancer (from the BROCA gene panel) [172-176]. WGS data from five Tasmanian prostate cancer families was examined for rare variants in genes from this panel.

Gene	
AMACR	NBN
AR	NBS1
ATM	NKX3-1
ATR	OR5H14
BRCA1	PALB2
BRAC2	PMS2
BRIP1	PRSS1
BTNL2	PTEN
CDH1	RBFOX1
CDKN2A	RAD51C
CHAD	RAD51D
CHEK2	RNASEL
ELAC1	SLX4
ELAC2	SPOP
ESR1	STK11
ESR2	TANGO2
HOXB13	TP53
MSR1	XRCC2

APPENDIX 12

Primers designed for *HOXB13* allele-specific next-generation sequencing and methylation analysis.

	Method	Forward Primer (5'-3')	Reverse Primer (5'-3')	Amplicon Size (bp)
Allele-Specific Next-Generation Sequencing	*HOXB13* rs138213197	GGACACTCGGCAGGAGTAGTA	GCTGATGCCTGCTGTCAACT	224 with Illumina Adaptors
	HOXB13 rs9900627	GGGAACCTACCAGCCTATGG	GTTCTGTTCTCCCTGGCAAC	215 with Illumina Adaptors
	Illumina Adaptors	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	-
Methylation Analysis	Upstream CpG Island (Product 1 in Figure 5.7)	TTCTCCCAACTAAAACAAACTCTAT	GTAAAGGTTATAGGTTGTTTGTGGG	254
	HOXB13 Promoter/Exon 1 CpG Island [406] (Product 2 in Figure 5.7)	ACTTATTCTCTCTCTCTCTCT	CCTTAACTCCATCCAAAATAAC	314
	HOXB13 Allele-Specific Methylation Analysis (Product 3 in Figure 5.7)	TTAATTATGTTTTTTTGGATTTGTTAGGT	ACTACCTAAACACAAAATTTCAAC	175

APPENDIX 13

Primer sequences of the forward and reverse tags used to barcode PCR products for the allele-specific next-generation sequencing assays.

Forward and reverse tags were designed by our collaborator, Andrea Polanowski (Australian Antarctic Division), and purchased from Integrated DNA Technologies.

Forward Tag Identification	Primer Sequence (5'–3')	Reverse Tag Identification	Primer Sequence (5'-3')
F16	AGCCTGGCAT	R20	GTCCACCAGT
F17	AGTTCGGACT	R21	ATCCGCCAGT
F18	AGTTCTTGAC	R23	ACTTGCCGAT
F19	ACGGTCCATG	R24	GCTTACCGAT
F20	ACTTGTTCAG	R27	GACCTAACTG
F21	ACTTGCCGAT	R28	ATGGCAACTG
F22	ACGGTGGATC	R33	GCTTACCATG
F23	ATCCGCCTAG	R43	CGAATGGTCA
F47	TCAAGCCAGT	R44	ATCCGTTGCA
F48	TCAAGAATGC		
F49	CTGGACCTGA		
F50	CGTTACCGTA		
F51	TGCCATTGCA		
F52	TCGGATTCGA		
F53	AGCCTGGCTA		
F54	AGCCTCCTGA		
F55	ACTTGTTCGA		
F56	ATCCGCCGTA		
F57	ATCCGTTCGA		
F58	ATGGCGGTCA		
F59	CAGGTGGCTA		
F60	CTAAGTTGCA		
F61	CGTTAGGTCA		
F62	TCGGACCGTA		

APPENDIX 14

Allele-specific next-generation sequencing; Galaxy FastQC report for PC12-08.

Figure **A** depicts 'Per Base Sequence Quality', with high quality scores (for each base position) shown in the green area and lower quality in red. As the 'position in read' increases the quality of the reads decrease significantly, which may be due to poor quality FFPE RNA used in this experiment. Figure **B** shows 'Per Sequence Quality Scores'.
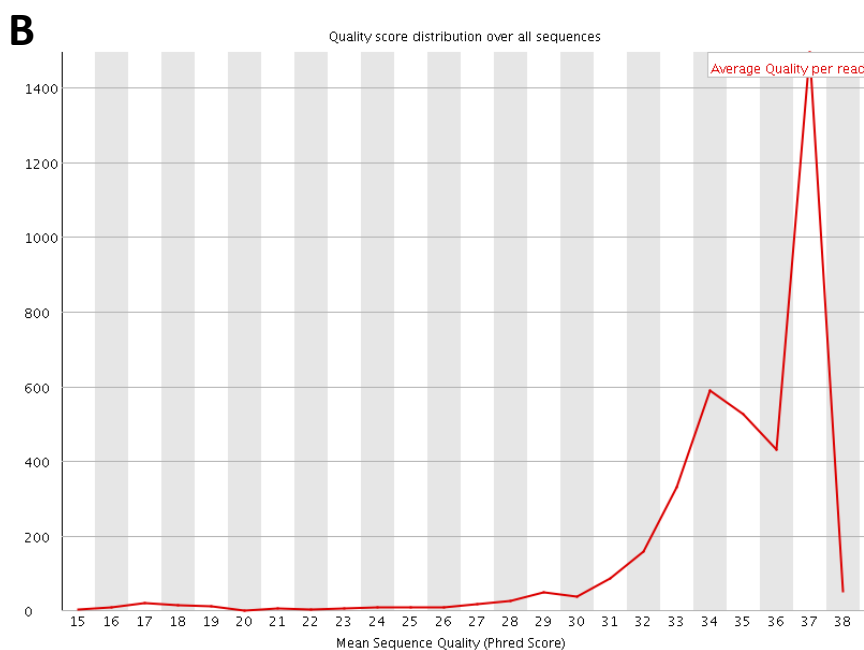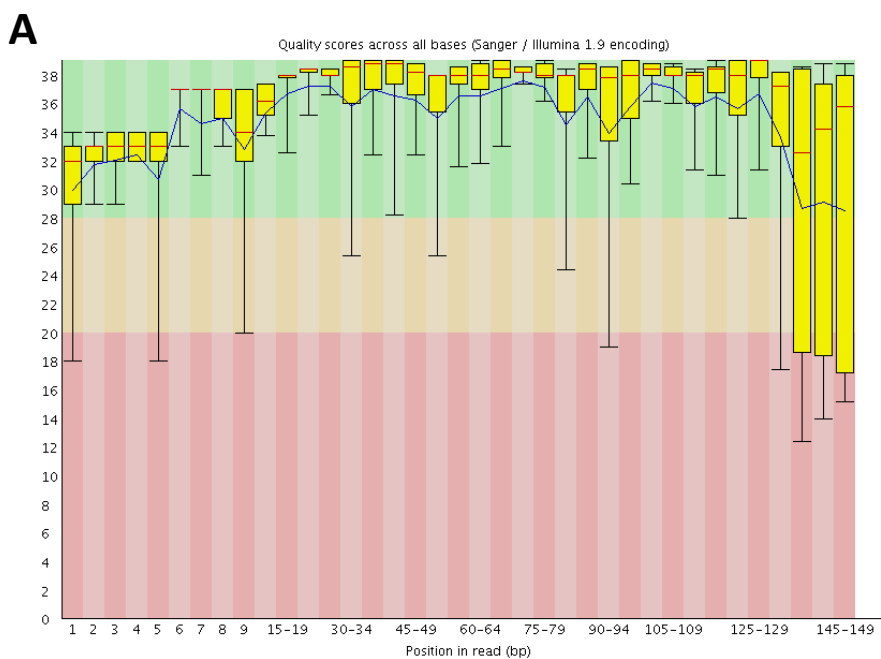
APPENDIX 15

Rare variants in known cancer-associated/DNA repair genes, following examination of WGS data from five Tasmanian prostate cancer families.

Gene	rs number	Chromosome: base pair	ExAC[1] MAF (%)	Family Identification	Segregation in WGS individuals (affected carriers/unaffected carriers)	CADD[2] Score	Allele Change; Amino Acid Change	Number of Control Carriers	ClinVar Search[3]
AR	Novel	X:66,766,342	N/A	PcTas22 Main	3 out of 5/ 0 out of 1	6.738	G > T; G452C	0 out of 8	N/A
ATM	Novel	11:108,235,819	N/A	PcTas3	1 out of 5	20.5	A > G; Y2954C	0 out of 8	N/A
ATM	rs1800056	11:108,137,753	0.90	PcTas4	3 out of 4/ 0 out of 1	8.76	T > C; F858L	0 out of 8	Hereditary cancer: Benign
BRCA1	rs4986852	17:41,244,179	1.32	PcTas4	1 out of 4/ 1 out of 1	14.94	C > T; S1040N	0 out of 8	Hereditary breast and ovarian cancer: Benign
BRCA1	rs28897673	17:41,256,016	<0.01	PcTas22 Sub	1 out of 4/ 1 out of 2	24.7	T > C; Y105C	0 out of 8	Hereditary breast and ovarian cancer: Benign

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; MAF: minor allele frequency; WGS: Whole-genome sequenced; [2]CADD: Combined Annotation Dependent Depletion [164]; Control: Control from the *Tasmanian Prostate Cancer Case-Control Study*, eight were WGS; [3]Associated condition: Interpretation of variant [163].

Gene	rs number	Chromosome: base pair	ExAC[1] MAF (%)	Family Identification	Segregation in WGS individuals (affected carriers/unaffected carriers)	CADD Score	Allele Change; Amino Acid Change	Number of Control Carriers	ClinVar Search[2]
BRCA2	rs56403624	13:32,906,750	0.02	PcTas3	1 out of 5	16.12	A > G; E462G	0 out of 8	Hereditary breast and ovarian cancer: Benign
BRCA2	rs28897727	13:32,912,500	0.68	PcTas4	1 out of 4/ 0 out of 1	16.04	G > T; D1420Y	0 out of 8	Hereditary breast and ovarian cancer: Benign
BRCA2	rs4987117	13:32,913,986	1.79	PcTas12	1 out of 2/ 0 out of 1	7.558	C > T; T1915M	1 out of 8	Hereditary breast and ovarian cancer: Benign
NBN	Novel	8:90,976,638	N/A	PcTas3	1 out of 5	25	C > T; G332R	0 out of 8	N/A
NBN	Novel	8:90,958,439	<0.01	PcTas22 Sub	2 out of 4/ 0 out of 2	0.073	A > G; S667P	0 out of 8	N/A
NKX3-1	rs199879315	8:23,540,125	0.39	PcTas22 Sub	2 out of 4/ 0 out of 2	10.18	C > G; G10R	0 out of 8	Not reported
OR5H14	rs112084609	3:97,868,154	0.73	PcTas22 Main	2 out of 5/ 0 out of 1	12.72	A > G; M59V	0 out of 8	Not reported

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; MAF: minor allele frequency; WGS: Whole-genome sequenced; [2]CADD: Combined Annotation Dependent Depletion [164]; Control: Control from the *Tasmanian Prostate Cancer Case-Control Study*, eight were WGS; [3]Associated condition: Interpretation of variant [163].

Gene	rs number	Chromosome: base pair	ExAC[1] MAF (%)	Family Identification	Segregation in WGS individuals (affected carriers/unaffected carriers)	CADD Score	Allele Change; Amino Acid Change	Number of Control Carriers	ClinVar Search[2]
PALB2	rs45494092	16:23,646,607	1.43	PcTas3, 4, 22 Sub, 72	1 out of 5; 2 out of 4/ 1 out of 1; 2 out of 1/ 1 out of 2; 0 out of 4/ 1 out of 4	8.68	A > G; L337S	0 out of 8	Hereditary breast cancer: Benign
RAD51C	Novel	17:56,787,286	N/A	PcTas22 Main	1 out of 5/ 0 out of 1	34	C > T; R258C	0 out of 8	N/A
RNASEL	rs56250729	1:182,555,403	0.77	PcTas3	2 out of 5	13.87	T > G; I97L	0 out of 8	Not reported
SLX4	rs759305861	16:3,633,131	0.02	PcTas4	2 out of 4/ 1 out of 1	5.784	G > C; P1624A	0 out of 8	Not reported
SLX4	rs148542931	16:3,638,822	<0.01	PcTas12	1 out of 2/ 0 out of 1	14.12	C > G; E1532Q	0 out of 8	Not reported
TANGO2	Novel	22:20,050,921	0.01	PcTas3	1 out of 5	15.48	C > A; S222R	0 out of 8	N/A

[1]ExAC, non-Finnish European, non-The Cancer Genome Atlas database; MAF: minor allele frequency; WGS: Whole-genome sequenced; [2]CADD: Combined Annotation Dependent Depletion [164]; Control: Control from the *Tasmanian Prostate Cancer Case-Control Study*, eight were WGS; [3]Associated condition: Interpretation of variant [163].

APPENDIX 16

HOXB13 **gene and protein expression analysis, G84E allele transcription and methylation analysis in FFPE prostate tissue samples (raw data).**

	Sample ID	Tissue Cell Type	Absolute *HOXB13* Gene Expression	IHC Score[1]	Protein Expression Final Score[2]	Methylation Assays[3]
***HOXB13* G84E non-carrier**	PC4-03	Malignant	1.36	2 (69%)	1.38	1, 2, 3
		Benign	0.02	3 (83%)	2.49	
	PC11-11	Malignant	0.59	3 (94%)	2.82	1, 2, 3
		Benign	0.17	2 (88%)	1.76	
	PC11-12	Malignant	0.50	3 (98%)	2.94	
		Benign		2 (59%)	1.18	
		Benign	0.44	3 (89%)	2.67	3
		Benign	0.23	3 (79%)	2.37	2, 3
	PC12-01	Malignant	2.72	1 (77%)	0.77	
		Benign	0.38	1 (34%)	0.34	
	PC12-06	Malignant	0.69	2 (81%)	0.62	3
		Benign		2 (72%)	1.44	
	PC12-09	Malignant	0.43	2 (92%)	1.84	1, 3
		Benign	0.37	3 (93%)	2.79	
	PC22-06	Benign	0.42	3 (91%)	2.73	
	PC47-02	Benign	0.30	1 (55%)	0.55	
	PC60-01	Malignant	0.58	2 (80%)	1.6	
		Benign	0.02	2 (59%)	1.18	
	PC63-24	Malignant		2 (100%)	2	
	PC72-04	Malignant	1.34	2 (88%)	1.76	3
		Benign	0.19	1 (74%)	0.74	
***HOXB13* G84E carrier**	PC12-03	Malignant	0.25	3 (90%)	2.7	1, 2, 3
		Benign	0.05	3 (80%)	2.4	3
	PC12-07	Malignant	0.40	2 (91%)	1.82	1, 2, 3
	PC12-08	Malignant	0.38	1 (69%)	0.69	
		Benign		2 (77%)	1.54	3
	PC22-203	Malignant		2 (69%)	1.38	1, 2, 3*
	PC22-203	Benign		2 (65%)	1.3	

Blank cell= sample was not analysed; [1]Staining intensity: 1=weak, 2=moderate, 3=strong (% of HOXB13 positive cells); [2]Final score is calculated by multiplying staining intensity (1, 2 or 3) by % of HOXB13 positive cells; [3]Primer pair used to assess CpG island methylation (as per Figure 5.7 and Appendix 12); *RNA/DNA extracted from a mixed cell population.

	Sample ID	Tissue Cell Type	Absolute *HOXB13* Gene Expression	IHC Score[1]	Protein Expression Final Score[2]	Methylation Assays[3]
HOXB13 G84E carrier	PC22-576	Malignant	0.97	3 (93%)	2.79	3
		Benign	0.21	1 (72%)	0.72	3
	PC22-637	Malignant		3 (80%)	2.4	
		Benign		3 (78%)	2.34	
	PC72-06	Malignant	0.78	3 (91%)	2.73	3
		Benign	0.37	3 (81%)	2.43	3
	PC72-154	Malignant		2 (81%)	1.62	
		Benign		3 (80%)	2.4	
	PC3250-01	Malignant	1.06	1 (100%)	1	3
		Benign	1.08	1 (100%)	1	

Blank cell= sample was not analysed; [1]Staining intensity: 1=weak, 2=moderate, 3=strong (% of HOXB13 positive cells); [2]Final score is calculated by multiplying staining intensity (1, 2 or 3) by % of HOXB13 positive cells; [3]Primer pair used to assess CpG island methylation (as per Figure 5.7 and Appendix 12).

APPENDIX 17

Array-Based Comparative Genomic Hybridisation fine-mapped regions.

Those listed are regions of loss and gain previously identified in tumours from PcTas9 cases (unpublished). These regions were targeted for fine-mapping on the array, as described in Chapter 6 of this thesis.

	Chromosome Band	Start (bp)	Stop (bp)	Region Size
Loss	1p22-1p31.1	80,299,794	82,299,794	2,000,000
	1q23.3-1q25.2	170,547,117	178,547,117	8,000,000
	6p25.1-6p25.3	1	2,000,000	2,000,000
	6q22-6q22.1	121,938,845	124,438,845	2,500,000
	7p21-7p21.3	7,300,002	21,089,819	13,789,818
	10q26.2	126,918,814	128,918,814	13,789,818
	17p13-17p13.3	1	6,500,000	6,500,500
	19p13.3	1	6,900,000	6,900,000
Gain	1p36.21-1p36.22	12,336,786	15,336,786	3,000,000
	6p22.1-6p22.3 6p24-6p25	6,516,515	15,616,515	9,100,000
	6q25.3-6q27	160,828,366	162,828,366	2,000,000
	17p13-17p13.3	1	6,500,000	6,500,000
	20p12-20p12.2	7,555,344	11,055,344	3,500,000

APPENDIX 18

Array-Based Comparative Genomic Hybridisation quality report.

This report was prepared by PathWest Pathology Laboratory (Drs Thomas and Robinson) and details the quality assessment of the 12 PcTas9 FFPE tissue samples assayed on the aCGH.

Analysis of FFPE Prostate Cancer Tissue Samples by Array CGH: A Pilot Feasibility Study.

Preliminary report prepared by Dr Carla Thomas and Dr Cleo Robinson

Molecular Anatomical Pathology, PathWest, QEII Medical Centre, WA

December 2017

Important Note

Please note that this is a research project only and the results have not been reviewed or validated by a pathologist and are not for clinical use.

Overview

Genomic DNA that had been extracted from 12 prostate cancer samples was sent to PathWest for analysis on a customised SurePrint G3 Human 8 x 60K Microarray (Agilent Technologies) that was designed by Dr Liesel Fitzgerald. Array CGH procedure was carried out by the Molecular Anatomical Pathology laboratory, PathWest, QEII Medical Centre, Nedlands, Perth WA according to the manufacturer's protocol and a workflow that has been validated for melanoma FFPE tissue samples.

Materials and Methods

DNA re-purification and quantification

The DNA received from Tasmania University was re-eluted with ddH_2O on Qiagen DNeasy Blood and Tissue Kit columns, as recommended by Agilent, prior to use with the ULS labelling system (Agilent, Australia). The purified DNA was quantified by spectrophotometry using a Nanodrop ND-2000 (Nanodrop, USA). Ratios of absorbance A260/280 were used to assess DNA purity, and all samples had ratios >1.80 and were thus regarded as sufficiently pure and suitable for ULS labelling.

DNA labelling, Array Hybridisation and Scanning

An optimised version of the manufacturer's protocol for ULS labelling of FFPE DNA (Agilent, Australia) was used as described in Ardakani et al., (2017). This comprised a step prior to labelling, whereby reference DNA and FFPE DNA was heat fragmented at 95°C for 10 and 1 min respectively. 250 ng of tumour and Promega reference DNA was then chemically labelled by incubating with 0.25 µl of ULS-Cy3 or Cy5, respectively, in a thermal cycler with a heated lid for 30 minutes. Unreacted dye was removed using KREApure filters (Agilent, Australia). Hybridisation was carried out for 40 hours and wash steps were carried out according to the manufacturer's protocol. The washed slide was scanned on the Agilent SureScan microarray scanner using protocol AgilentG3_CGH and then analysed using Agilent Cytogenomics software (Agilent, Australia).

Parameters used for Data Analysis

The data was imported into the Cytogenomics software via the Tiff image that is generated by the scanner. The software identifies copy number variations using the ADSM-2 algorithm according to the following parameters:

- normalising of the fluorescent intensity of both dyes at each probe and calculation of their ratio, expressed on a logarithmic scale (probe log2 ratio). Thus, where there

is no CNV the ratio should =0, however, the algorithm sets a threshold value of 0.25, so that a log2 ratio below this is considered no CNV and a significant CNV must be more than this threshold.

- NB. Based on data from our validation study we set a further log2 ratio threshold of >0.3 (gains/amplifications) or >-0.3 (losses/deletions) and genes were only called aberrant if they exceeded these limits.
- The optimal sensitivity threshold for ADM-2 (Aberation Detection Method 2) algorithm was defined as 6.1 based on our previous experience and validation study (Ardakani et al 2017) and this is in line with another similar published study (Wang et al 2013).

The software also computes a set of Quality Control (QC) metrics including the average green and red signal intensity at all probes and using non-hybridising control probes, determines the background signal (noise) and signal-to-noise ratio.

The software colour codes the QC metrics as follows: green: excellent; blue: good; pink: evaluate, however, please note, this is based on DNA from fresh samples. When using FFPE samples the threshold QC values need to be considered differently, as described below and which we have based on the results from our validation study:

- Derivative Log Ratio (DLR) Spread. This is the most important of the QC metrics. It calculates the probe to probe log ratio noise of an array. A poor DLR spread, defined by Agilent as >0.3, means that it is more difficult to accurately call amplification or deletions. A DLR spread of <0.1 is considered excellent and will allow small aberrations to be called, however, DLR spread this low is difficult to achieve with FFPE samples and typically they lie between 0.3 and 0.6. When validating the technique for FFPE samples we determined an appropriate DLR spread threshold of ≤0.6 was acceptable if the sex chromosome patterns were as expected.

- Average red (Cy3) and green (Cy5) signal intensity for all probes. For a non-FFPE sample array, a minimum value of 350 counts (Cy3) and 250 counts (Cy5) is expected for the red and green channels respectively. This is indicative of DNA quality and labelling efficiency (which are usually picked up at the time of labelling, as mentioned above). However, for FFPE samples the minimum levels are usually lower and values around 100 (Cy3) and 50 (Cy5) counts, for red and green signals respectively, are considered acceptable. Please note that the reference sample (green, Cy5) is consistently higher in these data because it is not FFPE derived and reflects better quality DNA.

- Background signal (noise) should be between 5 and 10. A value less than 5 is excellent. A high value is usually indicative of technical issues with the washing steps.

- Signal-to-noise ratio. This indicates how clearly the spots can be detected above background level. It is also dependent on washing steps. An excellent value would be >100, but values between 30 and 100 are good for fresh samples, however for FFPE the values are generally between 20 and 50.

- Opposite sex reference DNA was hybridised against each test DNA as an internal control to confirm the quality and validity of results for each experiment, ie. the X/Y patterns were as expected.

Results

Refer to the attached 'Prostate Cancer aCGH pilot study_Results' spreadsheet for results tables.

DNA Labelling

The degree of labelling (DoL) was determined according to the manufacturer's recommendations using the Nanodrop ND 2000. Labelling is considered optimal for DoL values between 1.75 - 3.5% for Cy3 and 0.75 - 2.5% for Cy5. Labelling passed these criteria for all samples.

Data Quality Analysis

The Derivative Log Ratio (DLR) Spread is considered the most important QC parametric and this was within the acceptable values for FFPE samples of ≤0.6 for 10/12 samples. The DLR range was 0.44 to 0.63 and 2 samples had a DLR spread >0.6 as follows:

> PC9-211 (0.63 in array #1 and in the repeat array #2, the DLR spread was 0.58)
> PC9-158 (0.61 and 0.62 in arrays 1 and 2 respectively).

Although the Green (Cy5) signal intensity and background noise were acceptable (range 108-145 and 3.6-5.5, respectively), the red signal intensity was very low ranging from 3 to 67 across the two array experiments, indicating poor quality samples and the red background noise was high, ranging from 3.8 to 21.4. Similarly Green signal to noise ratios were good, but red signal to noise ratios were poor.

The sex chromosome patterns were as expected for 9/12 cases. Cases PC9-211, 13 and 659 all had gain of Y but didn't show a loss of X chromosome. Of these 211 and 13 were repeated and the same result was found.

Chromosomal Aberrations

Chromosomal aberrations detected in this pilot study are listed in the excel spreadsheet and should be interpreted with the above mentioned quality metrics in mind. We have listed any regions with a log ratio of >0.3 (gains) or >-0.3 (losses), regardless of number of probes (see comment below). Dr Daniele Belluoccio, Agilent Technologist, has access to the raw data Tiff files for further consultation and access to the Cytogenomics software.

The genes within the region of chromosomal loss/gain are only listed in the spreadsheet if there are 5 genes or less. For aberrations involving > 5 genes, the genes are listed by sample and chromosomal location in the attached document, or can be examined in the cytogenomics software.

Comments

Genes were only called aberrant if they had a log ratio of >0.3 (gains) or >-0.3 (losses). This may not be the correct threshold for Prostate Cancer samples and a validation study would be needed to determine this and the threshold for other QC metrics for full and robust data interpretation. There were some aberrant regions with a high number of probes with a log2 ratio of 0.25-0.3 or -0.25—0.3 that have not been included in the spreadsheet.

For example:

PC9-158: 8q gain (q11.21-24.3), at log2 ratio 0.254, 750 probes.

PC9-13: 5q gain (q11.2-q33.1), log2 ratio 0.257, 723 probes

PC9-532: 17p gain (p13.2), log2 ratio 0.295, 50 probes and 6p loss (p24.3), log2 ratio -0.277, 106 probes

PC9-645: 17p loss (p13.2) -0.291, 423 probes

References

- MesbahArdakani, N, Thomas, C, Robinson, C, Mina, K, Harvey, NT, Amanuel, B, Wood BA. Detection of melanocytic lesions utilising array based comparative genomic hybridisation (2017) *Pathology*, 49(3): 285-291

- Wang L, Rao M, Fang Y, et al. A genome-wide high-resolution array-CGH analysis of cutaneous melanoma and comparison of array-CGH to FISH in diagnostic evaluation (2013). *The Journal of Molecular Diagnostics*, 15: 581-91.

284

APPENDIX 19

Cytogenomics 5.0.2.5 (Agilent) plots of the PcTas9 tumour samples that failed quality control.

PC9-13 (array 1 & 2), PC9-211 (array 1 & 2) and PC9-659 (array 2) failed QC as there was no loss of the X chromosome (in comparison to a female control). PC9-13 (array 1), PC9-158 (array 1 & 2), PC9-211 (array 1) failed QC as the DLR (derivative log spread) spread was >0.60.

PC9-211 Array 1
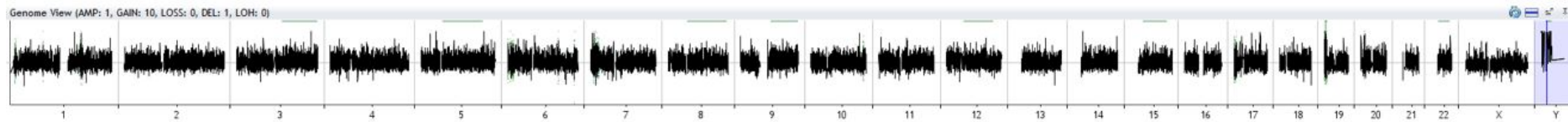
Genome View (AMP: 11, GAIN: 17, LOSS: 4, DEL: 2, LOH: 0)

PC9-211 Array 2

Genome View (AMP: 5, GAIN: 11, LOSS: 10, DEL: 6, LOH: 0)

PC9-659 Array 2

Genome View (AMP: 4, GAIN: 12, LOSS: 2, DEL: 1, LOH: 0)

286
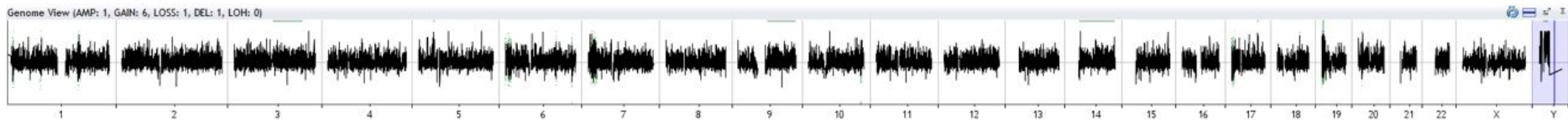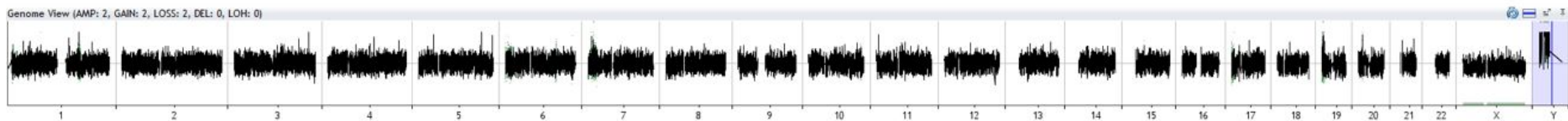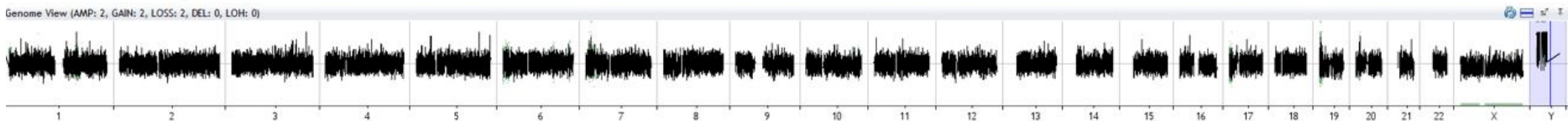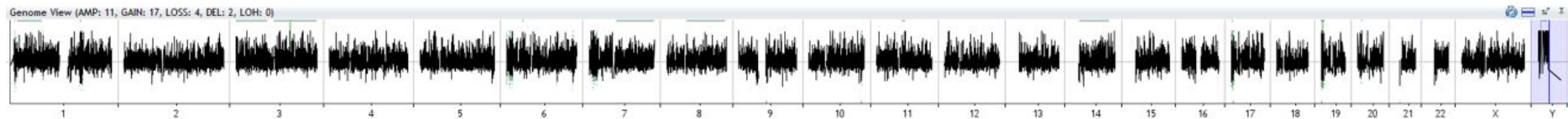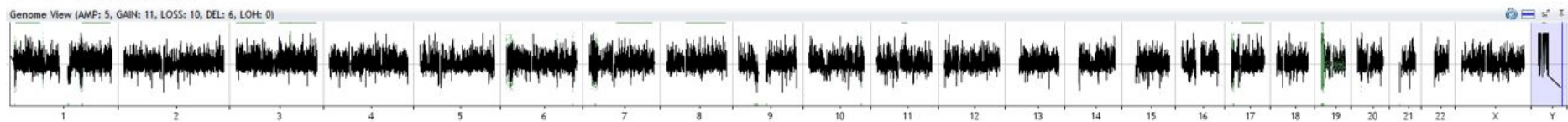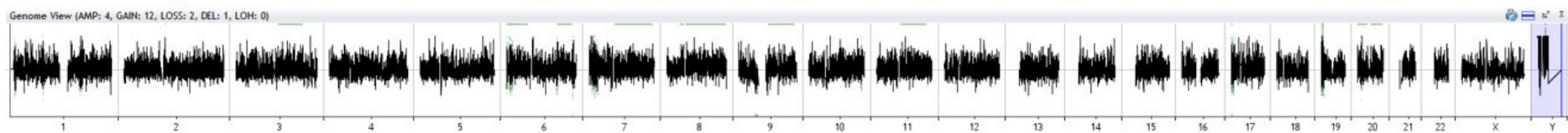
APPENDIX 20

Cytogenomics 5.0.2.5 (Agilent) plots of the PcTas9 tumour samples that passed quality control on array 1.

PC9-477
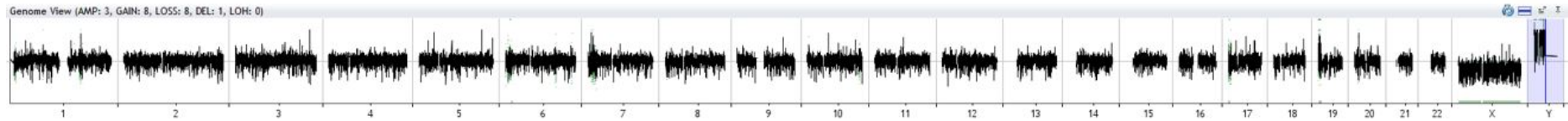
PC9-532
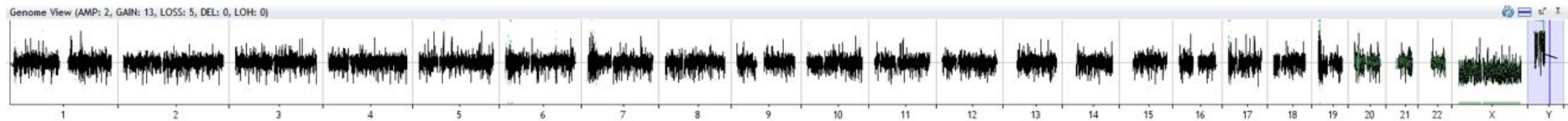
PC9-588
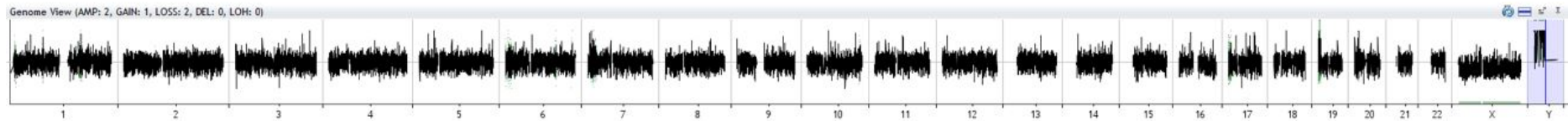
PC9-620
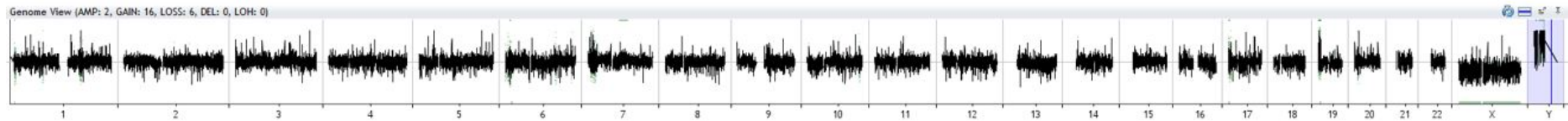
PC9-645
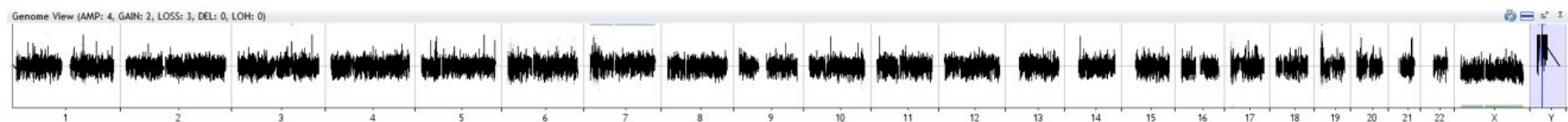
APPENDIX 21

Cytogenomics 5.0.2.5 (Agilent) plots of the PcTas9 tumour samples that passed quality control on array 2.

288
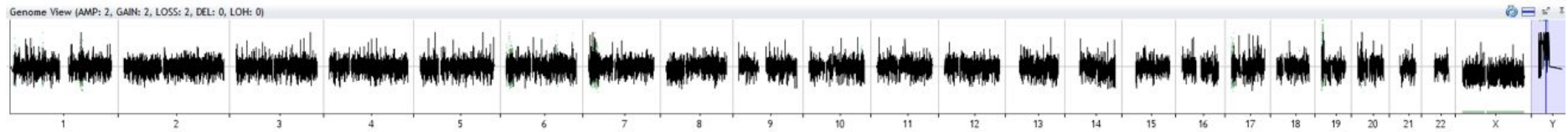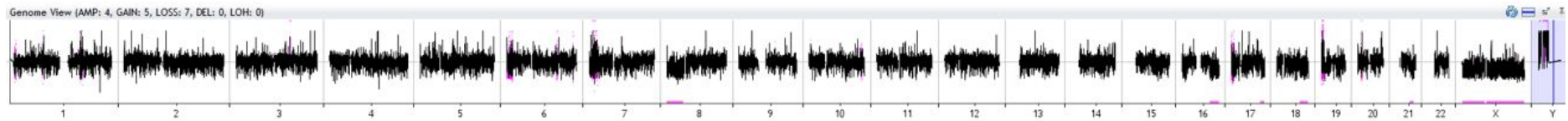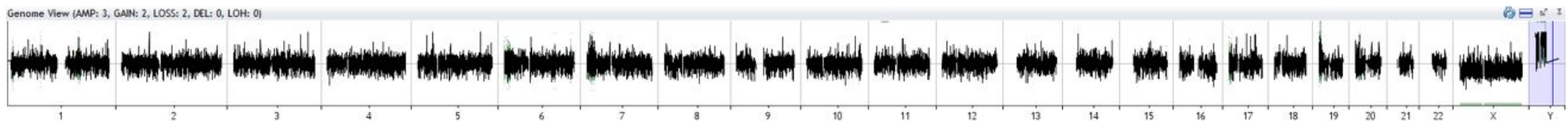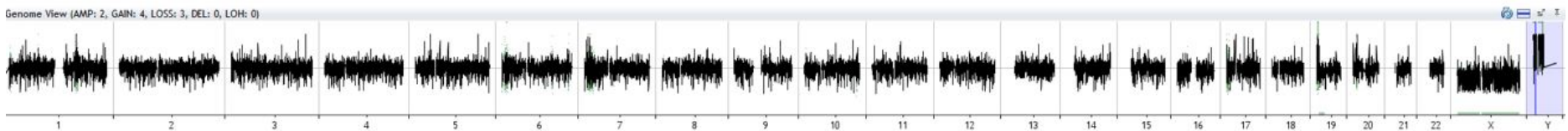
APPENDIX 22

Chromosomal losses and gains for each PcTas9 tumour sample assayed by aCGH.

Sample Identification	Cytoband	Number of Probes	Gain/Loss	p-value	Genes underlying the region
PC9-12	3q11.1-q26.32	634	0.160	1.89E-12	*PROS1, ARL13B, ARL6*
PC9-12	17p13.2	1,745	-0.195	2.77E-50	*ANKFY1, UBE2G1, SPNS3*
PC9-12	19p13.3	7	3.589	1.18E-72	*EEF2, MIR1268A, SNORD37*
PC9-20	1p34.3-p13.2	1684	0.153	6.34E-31	*RRAGC, MYCBP, GJA9*
PC9-20	3q12.3-q29	723	0.224	1.03E-30	*RPL24, CEP97, NFKBIZ*
PC9-20	6p23-p22.3	134	-0.215	2.09E-13	*JARID2, JARID2-AS1*
PC9-20	7p21.2	66	0.570	1.90E-13	*ETV1*
PC9-20	8p23.3-p11.21	311	-0.418	8.51E-35	*CLN8, ARHGEF10, MCPH1*
PC9-20	16q22.1-q24.3	155	-0.379	1.73E-16	*COG8, HAS3, CHTF8*
PC9-20	17p13.3	558	0.195	4.39E-21	*DPH1, RTN4RL1, OVCA2*
PC9-20	17q25.1-q25.3	72	-0.439	7.69E-12	*DNAI2, RPL38, TTYH2*
PC9-20	18q21.32-q23	174	-0.347	2.61E-14	*RAX, LMAN1, CCBE1*
PC9-20	19p13.3	7	3.593	1.38E-84	*EEF2, MIR1268A, SNORD37*
PC9-20	19p13.3	11	1.043	1.40E-13	*ZBTB7A*
PC9-20	20p12.3-p11.21	1690	0.156	2.03E-33	*PLCB1, PLCB4, SNAP25*
PC9-20	20p12.2	20	0.840	6.45E-11	*JAG1, MIR6870*
PC9-20	21q22.11-q22.3	77	-0.399	8.89E-10	*KCNE2, KCNE1, ITSN1*
PC9-477	1p36.21	206	-0.195	1.02E-11	*LRRC38, PDPN, LINC01784*
PC9-477	6p24.3-p24.2	451	-0.175	4.57E-19	*TFAP2A, GCNT2, MAK*
PC9-477	6p24.2	25	0.687	5.17E-17	*NEDD9*
PC9-477	6p23-p22.3	155	-0.279	4.23E-17	
PC9-477	7p21.1	298	0.162	2.10E-11	*HDAC9*
PC9-477	17p13.3	48	-0.428	1.63E-12	*RPA1, SMYD4*
PC9-477	17p13.3	53	0.362	9.58E-11	*DPH1, OVCA2, MIR132*
PC9-477	17p13.3	28	0.524	1.04E-11	*MNT*
PC9-477	17p13.3	105	-0.398	8.65E-12	*PAFAH1B1*

Sample Identification	Cytoband	Number of Probes	Gain/Loss	p-value	Genes underlying the region
PC9-477	19p13.3	139	0.268	1.21E-14	*ABCA7, GPX4, GRIN3B*
PC9-477	19p13.3	107	-0.266	2.27E-11	*ZNF77, ZNF554, ZNF555*
PC9-477	19p13.3	7	2.550	2.46E-58	*EEF2, MIR1268A, SNORD37*
PC9-477	19p13.3	38	0.468	2.20E-12	*PTPRS*
PC9-477	19p13.3	5	-1.295	2.95E-11	
PC9-477	19p13.3	70	-0.348	1.15E-12	*TNFSF9, CD70*
PC9-477	19p13.3	73	-0.312	7.90E-11	*SH2D3A, VAV1*
PC9-477	19p13.3-p11	113	-0.241	3.53E-10	*INSR, ARHGEF18, MCOLN1*
PC9-532	1p36.21	219	-0.211	2.76E-12	*LRRC38, PDPN, PRDM2*
PC9-532	3q13.11-q25.32	410	0.152	3.54E-12	*ALCAM, CBLB, DUBR*
PC9-532	5q11.2-q12.1	13	1.138	7.50E-21	*PDE4D*
PC9-532	6p24.3	105	-0.280	7.22E-10	*DSP, CAGE1, RIOK1*
PC9-532	6p24.2	30	0.830	5.49E-25	*NEDD9*
PC9-532	6p23	56	0.391	3.33E-11	*SIRT5, NOL7, RANBP9*
PC9-532	6p23-p22.3	176	-0.263	6.16E-15	*JARID2, JARID2-AS1*
PC9-532	7p21.3	87	0.336	1.50E-12	*ICA1, LOC100505938*
PC9-532	17p13.3	15	0.704	3.39E-10	*MYO1C*
PC9-532	17p13.3	468	0.175	1.27E-17	*DPH1, RTN4RL1, OVCA2*
PC9-532	17p13.3	9	1.080	3.87E-10	*SMG6*
PC9-532	17p13.2	126	0.151	2.19E-10	*CAMKK1, P2RX1, ATP2A3*
PC9-532	17p13.2	47	0.308	2.54E-10	*SPNS2, MYBBP1A, GGT6*
PC9-532	19p13.3	119	0.333	9.28E-17	*ABCA7, GPX4, GRIN3B*
PC9-532	19p13.3	14	0.852	2.58E-13	*MKNK2*
PC9-532	19p13.3	7	3.663	6.91E-96	*EEF2, MIR1268A, SNORD37*
PC9-532	19p13.3	10	0.850	5.71E-10	*ZBTB7A*
PC9-532	19p13.3	88	-0.303	1.18E-10	*TICAM1, FEM1A, PLIN3*
PC9-532	19p13.3	15	0.925	2.50E-16	*PTPRS*
PC9-532	19p13.3	162	-0.238	6.45E-12	*PTPRS, ZNRF4, TINCR*
PC9-588[1]	8q12.1-q24.3	634	0.154	1.64E-10	*CA8, CHC7, ASPH*
PC9-588[1]	10q25.2-q26.2	915	0.171	5.56E-18	*TCF7L2, HABP2, ADRB1*
PC9-588[1]	20p12.2-p11.1	378	0.201	8.20E-11	*JAG1, NDUFAF5, FLRT3*
[1]Assayed on array 1; [2]Assayed on array 2.					

290

Sample Identification	Cytoband	Number of Probes	Gain/Loss	p-value	Genes underlying the region
PC9-588[2]	6q22.31-q26	2602	0.155	9.79E-43	*TRDN, LAMA2, ARG1*
PC9-588[2]	7p22.1-p15.3	7097	0.159	1.28E-114	*ACTB, RNF216, PMS2*
PC9-588[2]	10q25.3-q26.2	1030	0.187	2.24E-28	*PNLIP, VAX1, KCNK18*
PC9-588[2]	11p15.1-p13	134	0.297	2.61E-10	*KCNJ11, ABCC8, USH1C*
PC9-588[2]	20p12.3-p12.2	1588	0.164	7.44E-31	*PLCB1, PLCB4, JAG1*
PC9-620	1p36.22	42	0.435	2.96E-13	*DHRS3, MIR6730*
PC9-620	1p36.21	33	-0.561	7.18E-16	*HNRNPCL3, HNRNPCL4, HNRNPCL1*
PC9-620	2p13.1-p11.1	124	-0.242	4.53E-11	*DCTN1, MOGS, HTRA2*
PC9-620	6p25.3	244	0.182	1.36E-12	*EXOC2, HUS1B*
PC9-620	6p24.3-p24.2	255	-0.169	1.65E-11	*GCNT2, MAK, GCM2*
PC9-620	6p24.2	24	0.892	2.77E-29	*NEDD9*
PC9-620	6p23-p22.3	77	-0.305	1.40E-11	
PC9-620	6q12-q21	329	-0.176	6.95E-15	*EYS, LMBRD1, COL9A1*
PC9-620	7p22.3-p11.2	7407	0.168	1.11E-281	*FAM20C, DNAAF5, MAD1L1*
PC9-620	7q21.11-q22.1	168	0.258	1.70E-17	*MAGI2, CD36, HGF*
PC9-620	10p15.1-p11.21	241	-0.230	1.34E-18	*AKR1C2, AKR1C4, IL2RA*
PC9-620	13q14.12-q34	520	-0.162	1.51E-19	*HTR2A, SUCLA2, NUDT15*
PC9-620	16q22.2-q24.1	113	-0.233	3.91E-10	*TAT, DHODH, HP*
PC9-620	17p13.3	1300	0.169	6.20E-54	*BHLHA9, INPP5K, DPH1*
PC9-620	17p13.3	38	-0.254	1.64E-11	*ABR, BHLHA9*
PC9-620	17p13.3	13	0.939	6.45E-13	*MY01C*
PC9-620	17p13.3	9	1.264	2.20E-17	*SMG6*
PC9-620	17p13.3	20	0.849	3.66E-15	*MNT*
PC9-620	17p13.2	244	-0.169	2.25E-11	*KIF1C, SLC52A1, INCA1*
PC9-620	19p13.3	216	0.228	5.26E-18	*ABCA7, GPX4, STK11*
PC9-620	19p13.3	11	0.801	5.76E-12	*MKNK2*
PC9-620	19p13.3	5	1.762	3.68E-24	*AES*
PC9-620	19p13.3	7	3.278	1.32E-87	*EEF2, MIR1268A, SNORD37*
PC9-620	19p13.3	10	1.036	2.82E-17	*ZBTB7A*
PC9-620	19p13.3	50	0.463	3.65E-17	*PTPRS*
PC9-620	19p13.3	28	-0.491	3.14E-11	*CATSPERD*
[1]Assayed on array 1; [2]Assayed on array 2.					

Sample Identification	Cytoband	Number of Probes	Gain/Loss	p-value	Genes underlying the region
PC9-620	19p13.3	62	0.392	1.95E-15	*RFX2*
PC9-620	19p13.3	9	0.800	4.28E-10	*MLLT1*
PC9-620	19p13.3	10	0.884	5.32E-13	*C3*
PC9-627	6p24.2	26	0.659	1.69E-14	*NEDD9*
PC9-627	6p23-p22.3	232	-0.187	1.45E-10	*JARID2, JARID2-AS1*
PC9-627	10q26.2	374	0.159	4.48E-12	*FANK1, ADAM12, C10orf90*
PC9-627	17p13.3	172	0.209	3.03E-10	*SRR, TSR1, MNT*
PC9-627	19p13.3	10	0.885	8.78E-11	*ZBTB7A*
PC9-627	19p13.2-p12	89	-0.298	2.68E-10	*CD320, RPS28, ANGPTL4*
PC9-645	3q13.31-q26.2	422	0.23	2.35E-15	*ZBTB20, ARHGAP31, POGLUT1*
PC9-645	4q12-q35.2	1016	0.179	7.43E-21	*SGCB, CHIC2, PDGFRA*
PC9-645	6q12-q26	3401	0.162	2.13E-52	*EYS, LMBRD1, COL9A1*
PC9-645	7p22.3-p11.2	7393	0.309	4.9E-324	*FAM20C, DNAAF5, MAD1L1*
PC9-645	7q11.21-q36.3	705	0.271	1.89E-34	*GUSB, ASL, KCTD7*
PC9-645	10q25.1-q26.2	1180	0.162	2.27E-21	*ADD3, MXI1, SMC3*
PC9-645	11q12.1-q24.1	487	0.167	3.26E-10	*CTNND1, FAM111B, FAM111A*
PC9-645	17p13.3-p13.2	3118	0.164	2.12E-60	*VPS53, BHLHA9, INPP5K*
PC9-645	17p13.2	423	-0.291	7.52E-25	*ZMYND15, CHRNE, GP1BA*
PC9-645	19p13.3	7	3.351	1.55E-60	*EEF2, MIR1268A, SNORD37*
PC9-645	20p12.3-11.1	1895	0.184	1.88E-39	*PLCB1, PLCB4, SNAP25*

APPENDIX 23

Consistently observed regions of loss and gain identified in three or more PcTas9 tumour samples.

Loss or Gain	Chromosomal Region	Frequency of CNV in PcTas9 tumours	Tumours with CNV	Known association with cancer	Interesting genes underlying the region of alteration
Loss	1p36.21	38% (3/8)	PC9-477, 532, 620	Ovarian cancer; CGH analysis of 28 ovarian tumours found that the 1p36 region was lost in 40% of tumours [295]. A study of pheochromocytomas and abdominal paragangliomas also found this region to be frequently deleted [407].	Genes underlying this region of loss, include genes in the *PRAME* and *HNRNPCL* gene families. *PRAME* family members are expressed in many cancer types, but also function in reproductive tissues during development [116]. *HNRNPCL* genes encode for RNA binding proteins, which influence pre-mRNA splicing processes and alterations could lead to alternative transcripts [116].
Loss	19p13.3	38% (3/8)	PC9-477, 532, 620	Prostate cancer; The 19p region of deletion has previously been identified in tumours from familial PCa cases by Rokman and colleagues (2001), however has not been identified in sporadic tumours [296].	Present in this region are a number of interesting genes which play a role in the antigen presentation process, the generation of cytotoxic T cells, and the activation and development of T and B cells [116]. Particularly interesting is the *TINCR* lncRNA (LIC00036), which has been suggested to have altered expression in multiple human cancers [319,320].
Gain	6p23-p22.3	63% (5/8)	PC9-20, 477, 532, 620, 627	Bladder cancer [297] and retinoblastoma tumours [298]	This region of gain encompasses the *JARID2* gene, which is a putative transcription factor that plays a role in DNA binding, nuclear localisation, transcriptional repression and recruitment of the Polycomb-repressive complex 2 [322-324]. Whilst the gene has never been found to be associated with

					PCa, *JARID2* has consistently been identified to play a role in the initiation, proliferation and maintenance of tumour cells in other cancers.
Gain	6p24.2	50% (4/8)	PC9-477, 532, 620, 627	No known association	*NEDD9* is frequently overexpressed in diverse cancer types and has been linked to tumorigenesis of many different malignancies, including PCa and is reasonably expressed in the normal prostate [116]. Interestingly, the region of amplification of *NEDD9* encompasses only the small transcript (NM_006403) and upon further investigation using the GTEx Portal, this is the most highly expressed transcript in the prostate [137].
Gain	17p13.3	63% (5/8)	PC9-20, 477, 532, 620, 627	Prostate cancer; Gain of 17p was reported by Rokman *et al.* (2001) in their study of familial PCa. This region of gain was not identified in any sporadic tumours, suggesting an association with familial prostate tumourigenesis [296].	A number of interesting genes are present in this region that play a role in transcriptional repression, initiation of transcription, the replication and maintenance of chromosome ends, and cell growth and differentiation. The *DPH1* gene was amplified in three out of the five tumours. *DPH1* is an enzyme involved in the biosynthesis of diphthamide, a modified histidine found only in *EEF2* [116].
Gain	17p13.3	25% (2/8)	PC9-477, 620	Prostate cancer; Gain of 17p was reported by Rokman *et al.* (2001) in their study of familial PCa. This region of gain was not identified in any sporadic tumours, suggesting an association with familial prostate tumourigenesis [296].	*MNT*, a member of the Myc/Max/Mad network of transcription factors that co-interact to regulate gene-specific transcription [116]. As *MYC* plays a role in cell cycle progression, apoptosis and cellular transformation, this interaction could be a key driver in prostate carcinogenesis [116]. In fact, it has now emerged that the MNT protein has

					the most substantial impact on MYC activities (reviewed in [408]).
Gain	17p13.3	25% (2/8)	PC9-532, 620	Prostate cancer; Gain of 17p was reported by Rokman *et al.* (2001) in their study of familial PCa. This region of gain was not identified in any sporadic tumours, suggesting an association with familial prostate tumourigenesis [296].	*SMG6* is a gene which encodes a component of the telomerase ribonucleoprotein complex, which is responsible for the replication and maintenance of chromosome ends [116]. Whilst this gene has never been implicated in PCa [409], it has recently been identified as a 5' fusion partner of *ALK* in cases of non-small-cell lung cancer [410].
Gain	19p13.3	38% (3/8)	PC9-477, 532, 620	Prostate cancer; The 19p region of amplification has previously been identified in tumours from familial PCa cases by Rokman and colleagues (2001), however has not been identified in sporadic tumours [296].	*PTPRS*, like other *PTP* family members, is a signaling molecule that regulates a variety of cellular processes including, cell growth, differentiation, the mitotic cycle and oncogenic transformation [116].
Gain	19p13.3	50% (4/8)	PC9-20, 532, 620, 627	Prostate cancer; The 19p region of amplification has previously been identified in tumours from familial PCa cases by Rokman and colleagues (2001), however has not been identified in sporadic tumours [296].	ZBTB7A is a zinc finger protein that is moderately expressed in the prostate. Given that ZBTB7A upregulation in gastric cancer cells promotes apoptosis and represses cell migration [318], the amplification identified in these four PCa samples may promote carcinogenesis by downregulation of the gene.
Gain	19p13.3	100% (8/8)	PC9-12, 20, 477, 532, 588, 620, 627, 645	Prostate cancer; The 19p region of amplification has previously been identified in tumours from familial PCa cases by Rokman and colleagues	All eight tumours had amplification of the *EEF2* gene. *EEF2* is an essential factor for protein synthesis as it promotes the GTP-dependent translocation of the nascent protein chain from the A to the P-site of the ribosome [116].

				(2001), however has not been identified in sporadic tumours [296].	It is overexpressed in a diverse range of cancer types, including PCa, and interestingly, has been suggested as a potential biomarker of PCa [299].

APPENDIX 24

EEF2 gene and protein expression analysis in FFPE prostate tissue samples (raw data).

Sample Identification	Tissue Cell Type	Absolute *EEF2* Gene Expression					EFF2 Protein Expression	
		5'UTR/ Exon 2	Exon 2/3	Exon 4/5	Exon 9/10	Exon 14/15	IHC Score[1]	Final Score[2]
DVA 67	Malignant						1 (70%)	0.70
	Benign						1 (20%)	0.20
DVA 157	Malignant	120.12	104.18	5.04	23.91	41.16	N/A	N/A
	Benign						2 (100%)	2.00
DVA 167	Malignant	17.32	32.38	3.30	26.12	4.22	2 (80%)	1.60
DVA 216	Malignant	177.66	146.06	14.45	71.32	49.96	1 (20%)	0.20
	Benign	1.78	47.64	0.45	3.95	13.79	1 (50%)	0.50
DVA 220	Malignant	176.21	107.26	7.00	39.63	142.54	0 (0%)	0
	Benign						1 (30%)	0.30
DVA 302	Malignant						3 (90%)	2.70
	Benign						2 (50%)	1.00
DVA 303	Malignant						3 (80%)	2.40
	Benign						1 (50%)	0.50
DVA 402	Malignant	158.41	167.64	0.88	37.67	70.09	2 (100%)	2.00
	Benign	168.12	43.45	3.40	16.38	22.07	2 (50%)	1.00
DVA 416	Malignant	2.07	25.98	5.00	84.45	21.99	2 (100%)	2.00
	Benign	29.08	62.83	2.57	12.09	17.26	2 (80%)	1.60
DVA 422	Malignant						2 (80%)	1.60
	Benign						1 (10%)	0.10
DVA 1002	Malignant	20.32	86.21	4.89	20.79	30.86	2 (90%)	1.80
	Benign						2 (50%)	1.00
DVA 1006	Malignant						1 (70%)	0.70
	Benign						2 (90%)	1.80
DVA 1036	Malignant						1 (50%)	0.50
	Benign						2 (80%)	1.60
DVA 1050	Malignant						2 (30%)	0.60
	Benign						2 (50%)	1.00
DVA 1086	Malignant						3 (100%)	3.00
	Benign						3 (100%)	3.00

Blank cell= sample was not analysed; [1]Staining intensity: 0=none, 1=weak, 2=moderate, 3=strong (% of EEF2 positive cells); [2]Final score is calculated by multiplying staining intensity (0, 1, 2 or 3) by % of EEF2 positive cells; N/A: No tissue.

Sample Identification	Tissue Cell Type	Absolute *EEF2* Gene Expression					EFF2 Protein Expression	
		5'UTR/ Exon 2	Exon 2/3	Exon 4/5	Exon 9/10	Exon 14/15	IHC Score[1]	Final Score[2]
PC3-08	Malignant						2 (100%)	2.00
	Benign						3 (100%)	3.00
PC3-31	Malignant						1 (80%)	0.80
	Benign						1 (80%)	0.80
PC4-03	Malignant	47.16	171.9	9.49	190.3	61.05	2 (70%)	1.40
	Benign	5.88	67.59	1.56	12.56	248.79	1 (50%)	0.50
PC9-04	Malignant						2 (80%)	1.60
	Benign						2 (80%)	1.60
PC9-05	Benign						2 (80%)	1.60
PC9-06	Malignant	225.41	123.58	12.37	57.57	33.20	2 (80%)	1.60
	Benign						3 (100%)	3.00
PC9-07	Malignant	27.81	152.67	13.87	66.94	21.37	1 (50%)	0.50
PC9-12	Malignant	985.42	149.0	75.64	142.7	99.97	2 (100%)	2.00
	Benign	314.16	71.72	9.05	38.02	57.34	2 (80%)	1.60
PC9-13	Malignant	141.86	91.75	21.64	71.03	28.56	2 (70%)	1.40
PC9-14	Malignant	134.43	200.8	21.81	433.0	55.66	2 (80%)	1.60
	Benign						1 (80%)	0.80
PC9-15	Malignant	96.21	123.6	16.47	56.91	40.83	3 (80%)	2.40
	Benign						3 (80%)	2.40
PC9-20	Malignant	837.59	113.7	41.97	74.16	55.60	1 (70%)	0.70
PC9-140	Benign						2 (100%)	2.00
PC9-158	Malignant	3740.95	69.12	22.42	49.99	36.94	1 (60%)	0.60
	Benign	33.25	21.37	27.82	62.13	13.62	1 (80%)	0.80
PC9-211	Malignant	81.70	27.91	1.57	11.14	9.98	1 (70%)	0.70
	Benign						1 (40%)	0.40
PC9-338	Malignant						2 (100%)	2.00
	Benign						2 (100%)	2.00
PC9-474	Malignant						2 (90%)	1.80
	Benign						1 (50%)	0.50
PC9-477	Malignant	140.60	45.18	9.94	43.90	75.36	N/A	N/A
	Benign	105.02	33.87	13.30	29.96	25.48	1 (50%)	0.50
PC9-532	Malignant	1813.13	74.64	18.95	48.58	59.52	1 (80%)	0.80
	Benign	86.61	53.84	12.15	32.49	33.20	0 (0%)	0

Blank cell= sample was not analysed; [1]Staining intensity: 0=none, 1=weak, 2=moderate, 3=strong (% of EEF2 positive cells); [2]Final score is calculated by multiplying staining intensity (0, 1, 2 or 3) by % of EEF2 positive cells; N/A: No tissue.

		Absolute *EEF2* Gene Expression					EFF2 Protein Expression	
Sample Identification	Tissue Cell Type	5'UTR/ Exon 2	Exon 2/3	Exon 4/5	Exon 9/10	Exon 14/15	IHC Score[1]	Final Score[2]
PC9-545	Malignant	151.38	111.5	11.12	36.02	65.20	1 (80%)	0.80
PC9-561	Malignant						2 (100%)	2.00
	Benign						2 (100%)	2.00
PC9-588	Malignant	45.01	44.13	9.52	44.53	23.38	2 (100%)	2.00
	Benign	153.76	49.99	5.98	23.05	18.04	2 (80%)	1.60
PC9-603	Malignant	151.59	74.62	13.40	50.30	24.20	N/A	N/A
	Benign						2 (30%)	0.60
PC9-620	Malignant	190.74	42.67	9.14	39.97	23.36	2 (80%)	1.60
	Benign	63.17	62.21	5.37	32.41	22.73	2 (80%)	1.60
PC9-627	Malignant	649.75	50.71	20.53	45.99	40.31	2 (70%)	1.40
	Benign						0 (0%)	0
PC9-645	Malignant	769.13	96.96	25.97	72.63	65.58	1 (10%)	0.10
	Benign	49.75	19.62	2.17	12.44	7.79	1 (70%)	0.70
PC9-659	Malignant	61.31	44.21	6.02	21.41	20.46	2 (80%)	1.60
	Benign						1 (20%)	0.20
PC9-951	Malignant						1 (100%)	1.00
	Benign						1 (100%)	1.00
PC11-11	Malignant	82.14	65.00	25.63	50.50	60.08	3 (80%)	2.40
	Benign	164.61	58.28	11.92	21.21	61.78	3 (80%)	2.40
PC11-12	Malignant	9.78	58.90	4.47	20.54	8.64	N/A	N/A
PC11-13	Benign						2 (80%)	1.60
PC11-16	Benign						3 (90%)	2.70
PC11-19	Malignant						0 (0%)	0
	Benign						1 (20%)	0.20
PC12-01	Malignant	37.17	78.73	6.97	68.61	48.21	0 (0%)	0
	Benign	27.19	16.93	5.65	23.23	5.29	1 (10%)	0.10
PC12-03	Malignant	13.19	76.79	2.20	18.50	35.56	0 (0%)	0
	Benign	17.32	25.87	2.41	13.20	27.07	1 (5%)	0.05
PC12-06	Malignant	293.02	63.63	3.24	22.52	30.55	2 (60%)	1.20
	Benign						1 (40%)	0.40
PC12-07	Malignant	70.24	32.91	2.56	17.47	23.78	1 (60%)	0.60
PC12-08	Malignant	1.10	4.59	0.64	0.43	3.01	3 (60%)	1.80
	Benign						3 (90%)	2.70

Blank cell= sample was not analysed; [1]Staining intensity: 0=none, 1=weak, 2=moderate, 3=strong (% of EEF2 positive cells); [2]Final score is calculated by multiplying staining intensity (0, 1, 2 or 3) by % of EEF2 positive cells; N/A: No tissue.

Sample Identification	Tissue Cell Type	Absolute *EEF2* Gene Expression					EFF2 Protein Expression	
		5'UTR/ Exon 2	Exon 2/3	Exon 4/5	Exon 9/10	Exon 14/15	IHC Score[1]	Final Score[2]
PC12-09	Malignant	32.67	47.83	2.39	21.24	19.288	0 (0%)	0
	Benign	17.17	46.82	3.18	24.52	23.75	1 (30%)	0.30
PC19-02	Malignant	0.79	50.50	0.54	38.04	11.59	2 (60%)	1.20
	Benign						2 (90%)	1.80
PC22-06	Benign						2 (80%)	1.60
PC22-17	Malignant	3.66	118.39	22.19	73.73	31.71	1 (80%)	0.80
	Benign						1 (5%)	0.05
PC22-576	Malignant	452.95	120.51	18.14	69.37	80.49	2 (80%)	1.60
	Benign	27.97	42.13	1.53	11.86	17.10	1 (50%)	0.50
PC23-02	Malignant						1 (50%)	0.50
	Benign						1 (10%)	0.10
PC27-01	Malignant						1 (30%)	0.30
PC31-01	Malignant	11.58	69.97	0.61	14.62	67.11	1 (100%)	1.00
	Benign						1 (50%)	0.50
PC47-02	Benign						2 (80%)	1.60
PC60-01	Malignant						2 (80%)	1.60
	Benign						2 (80%)	1.60
PC72-04	Malignant	27.20	88.43	5.39	42.10	31.53	1 (70%)	0.70
	Benign	0.98	42.45	1.31	17.29	43.37	0 (0%)	0
PC72-06	Malignant	53.08	35.37	1.15	9.24	3.27	1 (70%)	0.70
	Benign	13.80	61.55	1.08	49.22	6.02	1 (50%)	0.50
PC213-991	Malignant						2 (100%)	2.00
	Benign						1 (10%)	0.10
PC3250-01	Malignant	224.99	112.8	42.17	92.27	98.37	1 (50%)	0.50
	Benign	36.37	131.1	42.25	74.96	18.11	N/A	N/A

Blank cell= sample was not analysed; [1]Staining intensity: 0=none, 1=weak, 2=moderate, 3=strong (% of EEF2 positive cells); [2]Final score is calculated by multiplying staining intensity (0, 1, 2 or 3) by % of EEF2 positive cells; N/A: No tissue.

APPENDIX 25

DAPK3 gene expression analysis in FFPE prostate tissue samples (raw data).

Sample Identification	Tissue Cell Type	Absolute *DAPK3* Gene Expression		
		Exon 3/4	Exon 4/5	Exon 7/8
DVA 157	Malignant	5.18	5.06	12.40
DVA 220	Malignant	7.93	5.80	10.37
DVA 416	Benign	14.44	83.77	
PC4-03	Malignant	0.76	6.01	6.90
	Benign	3.27	0.39	8.97
PC9-12	Malignant	0.35	1.07	5.57
	Benign	1.23	4.43	7.90
PC9-13	Malignant	1.93	4.04	8.34
PC9-14	Malignant	1.09	4.04	9.78
PC9-15	Malignant	0.45	3.52	6.21
PC9-158	Malignant	0.47	3.05	5.54
	Benign	0.98	1.53	5.13
PC9-20	Malignant	0.38	2.59	7.12
PC9-211	Malignant	2.37	3.85	5.35
PC9-477	Malignant	0.60	2.14	9.08
	Benign	1.38	3.51	14.23
PC9-532	Malignant	2.12	2.58	7.15
PC9-532	Benign	0.83	2.340	11.87
PC9-545	Malignant	6.56	12.48	29.91
PC9-588	Malignant	0.61	1.32	7.92
	Benign	1.13	3.29	5.59
PC9-06	Malignant	1.58	4.72	13.54
PC9-603	Malignant	0.85	2.61	7.92
PC9-620	Malignant	0.39	3.02	12.44
	Benign	1.01	3.22	6.16
PC9-627	Malignant	0.28	1.49	3.89
PC9-645	Malignant	0.61	2.96	5.34
	Benign	0.21	2.98	4.96
PC9-659	Malignant	1.28	2.92	8.54
PC11-11	Malignant	1.36	1.59	5.57
	Benign	0.47	1.20	4.12
PC12-01	Malignant		3.17	6.59
Blank cell= sample was not analysed.				

Sample Identification	Tissue Cell Type	Absolute *DAPK3* Gene Expression		
		Exon 3/4	Exon 4/5	Exon 7/8
PC12-01	Benign	2.44	2.40	6.05
PC12-03	Benign	0.89	2.74	4.88
PC12-06	Malignant	9.04	3.86	7.92
PC12-07	Malignant	2.78	2.18	9.06
PC12-09	Malignant	0.87	1.13	8.60
	Benign		3.52	6.10
PC22-576	Malignant	23.25	5.68	4.64
	Benign	3.02	6.16	11.77
PC72-04	Malignant	2.64	1.82	2.96
PC72-06	Malignant	0.55	3.41	5.78
	Benign	1.59	3.02	6.23
PC3250-01	Malignant	0.79	2.08	5.25
	Benign	0.32	8.63	5.99

APPENDIX 26

TaqMan® *TMPRSS2:ERG* **expression assay identification numbers (Applied Biosystems).**

Fusion/Gene	Assay Identification	Assay Location	Amplicon Length
T1E2	Hs04396946_ft	60	105
T1E4	Hs03063375_ft	49	106
T1E4	Custom	N/A	112
β-Actin	Hs01060665_g1	208	63
T1E2: TMPRSS2 (Exon 1): *ERG* (Exon 2); *T1E4: TMPRSS2* (Exon 1): *ERG* (Exon 4)			

APPENDIX 27

Primers designed for Sanger sequencing validation of *TMPRSS2:ERG* **fusion positive tumours.**

Fusion	*TMPRSS2* Forward Primer (5'-3')	*ERG* Reverse Primer (5'-3')	Product size (bp)	Optimal annealing temperature (°C)
T1E2	CGCGAGCTAAGCAGGAGGCG [411]	TAAGCCAGCCCATCTACCAG	211	64
T1E4	GGAGGCGGAGGCGGAGGG [411]	TTTTGATGGTGACCCTGGCT	236	60
T1E2: TMPRSS2 (Exon 1): *ERG* (Exon 2); *T1E4: TMPRSS2* (Exon 1): *ERG* (Exon 4)				

APPENDIX 28

Clinicopathological characteristics of *TMPRSS2:ERG* fusion negative prostate tumours.

Sample Identification	Age at Diagnosis	Tumour Grade[1]	Contemporary Gleason Score[2]	Age at Death[3]	Cause of Death[3]
DVA 67	61	-	6 (2+4)	74	Non-Cancer
DVA 157	66	-	7 (3+4)		
DVA 167	53	PD	9 (5+4)	60	PCa
DVA 216	64	-	5 (3+2)	68	Other
DVA 220	63	MD	6 (3+3)		
DVA 402	52	MD	6 (3+3)		
DVA 416	62	MD	6 (3+3)		
DVA 1002	61	WD	6 (3+3)		
PC2-46	52	M/PD	7 (4+3)		
PC2-47	51	-	6 (3+3)		
PC3-08	69	MD	6 (3+3)	85	Non-Cancer
PC3-31	54	-	5 (3+2)		
PC4-03	80	M/PD	7 (4+3)	84	Non-Cancer
PC9-06	79	-	-	88	PCa
PC9-07	71	PD	10 (5+5)	73	PCa
PC9-13	83	-	-	87	Non-Cancer
PC9-15	64	MD	5 (2+3)	75	PCa
PC9-20	76	PD	-	83	PCa
PC9-158	63	-	6 (3+3)		
PC9-603	73	MD	6 (3+3)	86	Non-Cancer
PC9-659	65	-	9 (4+5)		
PC11-11	85	-	7 (3+4)	87	Non-Cancer
PC11-12	58	-	9 (4+5)	60	Other
PC12-03	62	WD	4 (2+2)		
PC12-07	59	PD	9 (4+5)	71	PCa
PC12-08	73	-	6 (3+3)	75	Other
PC12-09	68	-	6 (3+3)	82	Non-Cancer
PC12-132	61	-	8 (4+4)		

[1]Tumour grade obtained from pathology report; [2]Contemporary Gleason Score from pathology report (if known) or FFPE tissue block chosen for microdissection of nucleic acids; WD: well differentiated; MD: moderately differentiated; PD: poorly differentiated; -: information not present in original pathology report; [3]Age at death and cause of death information was obtained from the Tasmanian Cancer Registry (as at April 2019); PCa: Prostate Cancer; Other: Other cancer.

Sample Identification	Age at Diagnosis	Tumour Grade[1]	Contemporary Gleason Score[2]	Age at Death[3]	Cause of Death[3]
PC19-02	50	-	-		
PC22-17	56	MD	6 (3+3)		
PC22-576	69	M/PD	7 (3+4)		
PC23-02	78	MD	7 (3+4)	86	Non-Cancer
PC31-01	61	PD	5 (3+2)		
PC60-01	58	WD	6 (3+3)	70	Other
PC72-06	62	-	8 (4+4)	72	PCa
PC213-991	68	-	9 (4+5)		
PC3250-01	51	PD	9 (4+5)		

[1]Tumour grade obtained from pathology report; [2]Contemporary Gleason Score from pathology report (if known) or FFPE tissue block chosen for microdissection of nucleic acids; WD: well differentiated; MD: moderately differentiated; PD: poorly differentiated; -: information not present in original pathology report; [3]Age at death and cause of death obtained from the Tasmanian Cancer Registry (as at April 2019); PCa: Prostate Cancer; Other: Other cancer.

APPENDIX 29

ETV1 gene and protein expression analysis in FFPE prostate tissue samples (raw data).

Sample Identification	Tissue Cell Type	Absolute *ETV1* Gene Expression			ETV1 Protein Expression	
		Exon 8/10	Exon 16/17	Exon 21/22	IHC Score[1]	Final Score[2]
DVA 67	Malignant				0 (0%)	0
	Benign				2 (10%)	0.2
DVA 157	Malignant		5.04		N/A	N/A
	Benign				0 (0%)	0
DVA 167	Malignant				0 (0%)	0
DVA 216	Malignant				1 (40%)	0.4
	Benign				0 (0%)	0
DVA 220	Malignant	0.87	3.41	2.21	0 (0%)	0
	Benign				0 (0%)	0
DVA 302	Malignant				0 (0%)	0
	Benign				0 (0%)	0
DVA 303	Malignant				1 (20%)	0.2
	Benign				0 (0%)	0
DVA 402	Malignant				0 (0%)	0
	Benign		6.07		0 (0%)	0
DVA 416	Malignant				0 (0%)	0
	Benign		2.58		0 (0%)	0
DVA 422	Malignant				0 (0%)	0
	Benign				0 (0%)	0
DVA 1002	Malignant				0 (0%)	0
	Benign				0 (0%)	0
DVA 1006	Malignant				0 (0%)	0
	Benign				0 (0%)	0
DVA 1036	Malignant				0 (0%)	0
	Benign				0 (0%)	0
DVA 1050	Malignant				0 (0%)	0
	Benign				0 (0%)	0
DVA 1086	Malignant				1 (70%)	0.7
	Benign				0 (0%)	0

Blank cell= sample was not analysed; [1]Staining intensity: 0=none, 1=weak, 2=moderate, 3=strong (% of ETV1 positive cells); [2]Final score is calculated by multiplying staining intensity (0, 1, 2 or 3) by % of ETV1 positive cells; N/A: No tissue.

Sample Identification	Tissue Cell Type	Absolute *ETV1* Gene Expression			ETV1 Protein Expression	
		Exon 8/10	Exon 16/17	Exon 21/22	IHC Score[1]	Final Score[2]
PC3-31	Malignant				0 (0%)	0
PC4-03	Malignant		7.62		0 (0%)	0
	Benign		8.95	2.55	0 (0%)	0
PC9-04	Malignant				0 (0%)	0
	Benign				0 (0%)	0
PC9-05	Benign				0 (0%)	0
PC9-06	Malignant		2.05		0 (0%)	0
	Benign				0 (0%)	0
PC9-07	Malignant		3.03	0.68	0 (0%)	0
PC9-12	Malignant	0.07	0.49	0.19	0 (0%)	0
	Benign	0.72	1.43	0.85	1 (50%)	0.5
PC9-13	Malignant	0.09	2.63	1.49	0 (0%)	0
PC9-14	Malignant	0.59	3.73	0.31	0 (0%)	0
	Benign				0 (0%)	0
PC9-15	Malignant				0 (0%)	0
	Benign				0 (0%)	0
PC9-20	Malignant	0.13	4.57	2.57	1 (50%)	0.5
PC9-140	Benign				0 (0%)	0
PC9-158	Malignant	0.28	24.46	18.21	0 (0%)	0
	Benign	0.38	2.29	1.13	2 (50%)	1
PC9-211	Malignant	0.87	7.42	0.86	0 (0%)	0
	Benign				0 (0%)	0
PC9-338	Malignant				0 (0%)	0
	Benign				0 (0%)	0
PC9-474	Malignant				0 (0%)	0
	Benign				0 (0%)	0
PC9-477	Malignant	0.08	0.44	0.28	N/A	N/A
	Benign	0.06	0.57	0.19	0 (0%)	0
PC9-532	Malignant	0.79	1.00	1.55	0 (0%)	0
	Benign	0.36	0.72	0.38	1 (10%)	0.1
PC9-545	Malignant	0.53	4.94	2.01	0 (0%)	0
PC9-561	Malignant				0 (0%)	0
	Benign				0 (0%)	0

Blank cell= sample was not analysed; [1]Staining intensity: 0=none, 1=weak, 2=moderate (% of ETV1 positive cells); [2]Final score is calculated by multiplying staining intensity (0, 1 or 2) by % of ETV1 positive cells; N/A: No tissue.

		Absolute *ETV1* Gene Expression			ETV1 Protein Expression	
Sample Identification	**Tissue Cell Type**	**Exon 8/10**	**Exon 16/17**	**Exon 21/22**	**IHC Score[1]**	**Final Score[2]**
PC9-588	Malignant	0.32	1.08		1 (5%)	0.05
	Benign	0.37	1.39		1 (10%)	0.1
PC9-603	Malignant	0.14	4.16	1.22	N/A	N/A
	Benign				0 (0%)	0
PC9-620	Malignant	0.13	1.52	0.66	0 (0%)	0
	Benign	0.15	2.47	1.37	1 (10%)	0.1
PC9-627	Malignant	0.11	1.21	0.85	0 (0%)	0
	Benign				0 (0%)	0
PC9-645	Malignant	0.21	0.56	0.69	0 (0%)	0
	Benign	0.22	0.96	0.15	0 (0%)	0
PC9-659	Malignant	0.19	1.77	0.37	0 (0%)	0
	Benign				0 (0%)	0
PC9-951	Malignant				0 (0%)	0
	Benign				0 (0%)	0
PC11-11	Malignant	0.32	0.85	1.59	0 (0%)	0
PC11-11	Benign	1.14	0.45	0.26	0 (0%)	0
PC11-13	Benign				1 (10%)	0.1
PC11-16	Benign				1 (5%)	0.05
PC11-19	Malignant				0 (0%)	0
	Benign				0 (0%)	0
PC12-01	Malignant	0.17	2.50		0 (0%)	0
	Benign		1.74	0.62	0 (0%)	0
PC12-03	Malignant				0 (0%)	0
	Benign	0.22			0 (0%)	0
PC12-06	Malignant	0.76	2.11	4.17	0 (0%)	0
	Benign				0 (0%)	0
PC12-07	Malignant	0.51	1.35	0.73	0 (0%)	0
PC12-08	Malignant				0 (0%)	0
	Benign				0 (0%)	0
PC12-09	Malignant	0.58	1.48	0.42	0 (0%)	0
	Benign	0.71	2.49	0.89	0 (0%)	0
PC19-02	Malignant				1 (20%)	0.2
	Benign				1 (20%)	0.2

Blank cell= sample was not analysed; [1]Staining intensity: 0=none, 1=weak, 2=moderate (% of ETV1 positive cells); [2]Final score is calculated by multiplying staining intensity (0, 1 or 2) by % of ETV1 positive cells; N/A: No tissue.

		Absolute *ETV1* Gene Expression			ETV1 Protein Expression	
Sample Identification	Tissue Cell Type	Exon 8/10	Exon 16/17	Exon 21/22	IHC Score[1]	Final Score[2]
PC22-06	Benign				0 (0%)	0
PC22-17	Malignant				0 (0%)	0
	Benign				0 (0%)	0
PC22-576	Malignant				0 (0%)	0
	Benign			1.66	0 (0%)	0
PC23-02	Malignant				0 (0%)	0
	Benign				0 (0%)	0
PC27-01	Malignant				1 (70%)	0.7
PC31-01	Malignant				0 (0%)	0
	Benign				0 (0%)	0
PC47-02	Benign				0 (0%)	0
PC60-01	Malignant				0 (0%)	0
	Benign				0 (0%)	0
PC72-04	Malignant	0.61	2.62		0 (0%)	0
	Benign				0 (0%)	0
PC72-06	Malignant	0.43	1.91	0.71	0 (0%)	0
	Benign	7.24	2.38		0 (0%)	0
PC213-991	Malignant				0 (0%)	0
	Benign				1 (10%)	0.1
PC3250-01	Malignant	0.41	1.17	0.54	0 (0%)	0
	Benign	0.81	3.44		N/A	N/A

Blank cell= sample was not analysed; [1]Staining intensity: 0=none, 1=weak, 2=moderate (% of ETV1 positive cells); [2]Final score is calculated by multiplying staining intensity (0, 1 or 2) by % of ETV1 positive cells; N/A: No tissue.

APPENDIX 30

This condensed PcTas9 pedigree indicates those tumours assessed for *EEF2* 5'UTR/exon 2 expression and the two *TMPRSS2:ERG* fusion events; *EEF2* 5'UTR/exon 2 overexpressing and/or *TMPRSS2:ERG* fusion positive tumours are shown in yellow and tumours with no overexpression of *EEF2* and *TMPRSS2:ERG* fusion negative, in grey.