Raspin, Kelsie, FitzGerald, Liesel M., Marthick, James R., Field, Matt A., Malley, Roslyn C., Banks, Annette, Donovan, Shaun, Thomson, Russell J., Foley, Georgea R., Stanford, Janet L., and Dickinson, Joanne L. (2021) ***A rare variant in EZH2 is associated with prostate cancer risk.*** **International Journal of Cancer, 149 (5) pp. 1089-1099.**

https://doi.org/10.1002/ijc.33584

Dickinson Joanne (Orcid ID: 0000-0003-4621-1703)

A rare variant in *EZH2* is associated with prostate cancer risk

Kelsie Raspin[1], Liesel M FitzGerald[1], James R Marthick[1], Matt A Field[2,3], Roslyn C Malley[4,5], Annette Banks[1], Shaun Donovan[4], Russell J Thomson[6], Georgea R Foley[1], Janet L Stanford[7], and Joanne L Dickinson[1,*]

[1] Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, 7000, Australia

[2] Australian Institute of Tropical Health and Medicine and Centre for Tropical Bioinformatics and Molecular Biology, James Cook University, Cairns, QLD, 4878, Australia

[3] John Curtin School of Medical Research, Australian National University, Canberra, ACT, 2601, Australia

[4] Hobart Pathology, Hobart, TAS, 7000, Australia

[5] Tasmanian School of Medicine, University of Tasmania, Hobart, TAS, 7000, Australia

[6] Centre for Research in Mathematics and Data Science, Western Sydney University, Sydney, NSW, 2150, Australia

[7] Fred Hutchinson Cancer Research Center, Seattle, WA, 98109, USA

* To whom correspondence should be addressed. Tel: +61 3 6226 7622; Email: jo.dickinson@utas.edu.au

Key words: prostate cancer, rare genetic variants, genetic susceptibility, transcriptome analysis.

Abbreviations: FFPE – formalin-fixed paraffin embedded; FHCRC – Fred Hutchinson Cancer Research Center; GS – Gleason score; HMTase – histone methyltransferase; MAF – minor allele frequency; M_{QLS} – Modified Quasi Likelihood Score; NGS – next-generation sequencing; OR – odds ratio; PCA – Principal components analysis; PcG – Polycomb group; PRC2 – Polycomb Repressive Complex 2; PrCa – prostate cancer; PROGRESS – Prostate Cancer Genetic Research Study; TCGA – The Cancer Genome Atlas; TCR – Tasmanian Cancer Registry; WES; whole-exome sequencing; WGS; whole-genome sequencing.

Novelty and Impact:

Few rare risk variants have been identified for prostate cancer despite evidence they contribute significantly to risk. Herein, we describe one of the few studies that have applied a whole-genome sequencing approach to a large multigenerational family with densely aggregated cases. A predicted splice variant in *EZH2* was identified and significantly associated with prostate cancer risk, with evidence of perturbation of *EZH2* function in the tumours of variant carriers.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ijc.33584

ABSTRACT

Prostate cancer (PrCa) is highly heritable, and although rare variants contribute significantly to PrCa risk, few have been identified to date. Herein, whole-genome sequencing was performed in a large PrCa family featuring multiple affected relatives spanning several generations. A rare, predicted splice site *EZH2* variant, rs78589034 (G>A), was identified as segregating with disease in all but two individuals in the family, one of whom was affected with lymphoma and bowel cancer and a female relative. This variant was significantly associated with disease risk in combined familial and sporadic PrCa data (n=1,551; OR=3.55, $p=1.20x10^{-05}$). Transcriptome analysis was performed on prostate tumour needle biopsies available for two rare variant carriers and two wild-type cases. Although no allele-dependent differences were detected in *EZH2* transcripts, a distinct differential gene expression signature was observed when comparing prostate tissue from the rare variant carriers with the wild-type samples. The gene expression signature comprised known downstream targets of EZH2 and included the top ranked genes, *DUSP1*, *FOS*, *JUNB* and *EGR1*, which were subsequently validated by qPCR. These data provide evidence that rs78589034 is associated with increased PrCa risk in Tasmanian men and further, that this variant may be associated with perturbed *EZH2* function in prostate tissue. Disrupted *EZH2* function is a driver of tumourigenesis in several cancers, including prostate, and is of significant interest as a therapeutic target.

INTRODUCTION

Prostate cancer (PrCa) is the most common, non-cutaneous malignancy in men and is a leading cause of male cancer-related deaths in the developed world[1]. The heritability of this cancer is estimated at 58%, greater than any other common cancer[2]. Numerous large-scale genome-wide association studies have identified more than 170 common genetic PrCa risk variants (reviewed in Benafif *et al.* (2018)[3]). Yet a recent meta-analysis suggests that these common variants still only explain a minor portion of risk (28.4%)[4].

Mancuso and colleagues report that approximately 42% of 'missing' PrCa heritability is likely to be explained by rare variants[5]. Rare variants typically play a more apparent role in disease causation, with stronger effect sizes[6] and are proving to be of significant clinical utility in guiding treatment decisions for several other cancers, such as breast cancer (reviewed in Berger *et al.* (2018)[7]). Despite these successes, few rare risk variants have been identified for PrCa. Targeted next-generation sequencing (NGS) in PrCa families exhibiting linkage to a region on chromosome 17q identified a rare mutation in *HOXB13.* The G84E (rs138213197) variant was found to be associated with a 3-16-fold increase in PrCa risk[8] and this association has subsequently been replicated in multiple studies[9]. Two whole-exome sequencing (WES) studies, conducted in a familial PrCa cohort from the Fred Hutchinson Cancer Research Center (FHCRC), identified two rare risk variants (minor allele frequency (MAF) <2%) in the *BTNL2* gene[10] and single variants in *TANGO2*,

OR5H14 and *CHAD*[11]. These variants were subsequently found to be significantly associated with PrCa risk in a FHCRC population-based, case-control dataset[10,11]. Other studies have taken a candidate gene approach; identifying rare mutations in DNA repair-associated genes *BRCA1/2*, *CHEK2*, *ATM*, *PALB2*, *BRIP2* and *NBN* that were associated with increased PrCa risk, and their clinical utility is now being realised[12]. However, these rare variants still only explain ~5-6% of PrCa heritability, thus there is a strong imperative to further elucidate rare variant contribution to PrCa risk and the role they play in tumourigenesis[13].

Although candidate gene approaches have had some success, there are calls for further agnostic approaches to rare variant discovery[14]. Herein, we apply whole-genome sequencing (WGS) to a large multigenerational Tasmanian family with a dense aggregation of PrCa. The *Tasmanian Familial Prostate Cancer Study* commenced in the late 1990s, utilising the Tasmanian Cancer Registry (TCR) and extensive Tasmanian genealogical records to identify large families with multiple cases of PrCa. Tasmania is an island state of Australia inhabited by the Palawa Indigenous people. It was colonised by Europeans as a penal settlement in the early 1800s. A mild-population bottleneck was experienced in the mid 1800s; at this time the population included an estimated 10,000 Northern European ancestry couples (at most) of childbearing age. It is estimated that these "founding families" were responsible for ~65% of Tasmanian residents in 1996[15]. The Tasmanian Familial Prostate Cancer Study commenced recruitment of families with multiple cases of PrCa in the late 1990s, prior to the introduction of prostate-specific antigen testing, thus families were largely ascertained on the basis of symptomatic PrCa. Here, WGS of individuals from a large Tasmanian PrCa family identified a rare *EZH2* variant. Subsequently, this variant was found to be associated with PrCa in a larger dataset comprising familial and sporadic cases and controls. Tumour-based analyses were then performed to examine the putative functional role this variant may play in PrCa development.

MATERIALS & METHODS

Study Resources

The Tasmanian Familial Prostate Cancer Study

This study comprises two Tasmanian PrCa datasets. The first is the *Tasmanian Familial Prostate Cancer Cohort*, which includes a rare collection of 73 PrCa families from the founder population of Tasmania and has been described previously[16]. DNA samples from blood and saliva were available for 293 affected men and 445 male and female relatives. The *Tasmanian Prostate Cancer Case-Control Study* is a Tasmanian population-based dataset, which includes blood and saliva samples from 472 cases and 341 controls[16]. Archived prostate tissue pathology blocks from *EZH2* rs78589034 carriers and a random selection of non-variant carriers were targeted for collection from local pathology laboratories. Sectioned formalin-fixed paraffin embedded (FFPE) pathology blocks

were histologically reviewed by pathologists (RM and SD) to provide a contemporary grading of the tumours (Supplementary Table S1) and to mark regions of malignant and benign cells.

The Tasmanian Clinical-Based Prostate Cancer Biopsy Cohort

This cohort is comprised of prostate tumour needle biopsies from 63 men participating in the *Tasmanian Tissue Resource* or *The Study of Prostate Cancer in Tasmanian Men* studies. These men underwent a radical prostatectomy and consented to their clinician taking subsequent needle biopsies from this tissue. Biopsies from both the right and left lobe of the prostate were taken. Table 1 provides clinical details of cases and the Gleason score (GS) reported from the radical prostatectomy pathology report.

The Fred Hutchinson Cancer Research Center Prostate Cancer Genetic Research Study

The FHCRC genetic resource includes families participating in the *Prostate Cancer Genetic Research Study* (*PROGRESS*), which comprises a collection of more than 300 families ascertained from across North America[11,17]. For the analysis included in this manuscript, WES data from 321 affected men diagnosed with early-onset and/or aggressive PrCa from 130 families was available. The number of individuals with WES data available in each family ranged from two to five. Eleven unaffected men were selected from the families for WES. The *PROGRESS* study was approved by the Institutional Review Board at the FHCRC.

Nucleic Acid Extractions

DNA from blood and saliva was extracted as previously described by FitzGerald and colleagues[16]. DNA and RNA were extracted from macro-dissected malignant and benign regions of FFPE tissue blocks from *EZH2* rs78589034 carriers and non-variant carriers (Supplementary Table S1)[16]. DNA was extracted from the left core of 63 fresh needle biopsies, using the AllPrep Micro Kit (Qiagen), as per the manufacturer's instructions, and quantitated using the Nanodrop® ND-1000 UV-vis spectrophotometer (Nanodrop® Technologies). RNA was extracted from 12 biopsies from seven men in the *Tasmanian Clinical-Based Prostate Cancer Biopsy Cohort,* using the RecoverAll Total Nucleic Acid Isolation Kit (ThermoFisher Scientific), according to the manufacturer's instructions. The SuperScript™ VILO™ cDNA Synthesis Kit (Invitrogen) was used for cDNA conversion, as per the manufacturer's instructions.

Next-Generation Sequencing

.

WGS data was generated for three PcTas12 individuals (Figure 1) at the Kinghorn Centre for Clinical Genomics, Australia, on the Illumina HiSeq X™ Ten platform using the TruSeq Nano library preparation. Quality control assessment of the NGS data is displayed in Supplementary Method 1 and Supplementary Table S2. Data was analysed using the Variant Analysis of Sequenced Pedigrees (VASP) analytical pipeline, developed specifically to detect disease-causing variants in sequenced pedigrees[18,19]. Variant reports for single nucleotide variants and insertions/deletions were generated and prioritised based on the following criteria: segregation with disease; MAF of <2% in the gnomAD non-Finnish European database; variant type (e.g. nonsynonymous or splice site) and a CADD[20] score of >10.

Sanger Sequencing

Sanger sequencing was performed on the ABI 3500 Genetic Analyser (Applied Biosystems) to validate the WGS results and determine variant segregation with disease in additional PcTas12 family members with germline samples (n=54). *EZH2* variant carrier status was also determined in deceased affected men with only a tumour specimen available (n=3). *EZH2* primers were designed using PrimerBLAST[21] (Supplementary Table S3) and PCR conditions are available on request.

Genotyping

A TaqMan SNP genotyping assay (Assay ID C_64633016_10, Life Technologies) was used to genotype the *EZH2* variant in the familial and case-control datasets on the LightCycler® 480 system (Roche). Heterozygous individuals were confirmed by Sanger sequencing.

Real-Time qPCR Analysis

SYBR green real-time quantitative PCR (RT-qPCR) assays were performed to determine gene expression of *EZH2* and two housekeeping genes, *β-Actin* and *GAPDH*. Amplification was performed on 50ng FFPE cDNA using *EZH2* primers designed in exons 8/9 of the gene, and published *β-Actin* and *GAPDH* primers[22] (IDT and Sigma Aldrich; Supplementary Table S4). In total, 18 malignant and 14 benign samples from seven *EZH2* variant carriers and 11 non-variant carriers were assessed. Standard curves were generated for each primer pair to determine PCR efficiency and normalise *EZH2* expression. PCR conditions are available on request.

Cell Culture

This method pertains to the results of the *in vitro* splicing assay shown in Supplementary Figure S2. PC3 (RRID:CVCL_0035) and 22Rv1 (RRID:CVCL_1045) cells were obtained from European

.

Collection of Authenticated Cell Cultures (ECACC, United Kingdom) and cultured in RPMI as described previously[23]. Cells were sub-cultured every 3-4 days and were maintained between 1 x 10^5 and 1 x 10^6 cells/ml. All were cultured in a humidified incubator at 37°C and 5% CO_2. Both cell lines have been authenticated using short tandem repeat profiling within the last three years. All experiments were performed with mycoplasma-free cells.

Transcriptome sequencing analysis

Transcriptome sequencing data were generated for four fresh prostate needle biopsies at the Australian Genome Research Facility, Australia, on the NovaSeq 6000 using the TruSeq stranded mRNA sample preparation. Samples were run in duplicate and 20 million, 100 base pair (bp) single end reads were generated for each. Read quality reports were generated for each library (using FastQC) and adapters trimmed from reads using TrimGalore. Quality control assessment of the transcriptome data is displayed in Supplementary Table S5. A principal component analysis (PCA) demonstrated good correlation (r^2=0.89) between the duplicates, and thus, the replicate with the most reads was chosen for subsequent analysis.

For specific gene transcript analysis, reads in fastQ format were aligned to the hg38 reference genome using a fast and sensitive splice alignment program, HiSat2 (version 2.1.0). Transcripts were assembled and quantified using StringTie (version 2.1.1) and the output processed by Ballgown. Transcript level expression reports, which included annotations such as RefSeq transcript identification numbers, total number of exons included in the transcript and transcript expression (RPKM), were generated for each sample.

For differential expression analysis, trimmed reads in fastQ format were aligned to the hg38 reference genome using TopHat2 (version 2.1.1), a gapped-read mapper for RNA-seq data. FeatureCounts (version 1.6.4) was used to count the expression of each gene in each sample and DESeq2 (version 2.11.40.6) was used to analyse differential expression of genes between the *EZH2* rs78589034 carriers and non-variant carriers. Statistically significant genes were determined by Bonferroni correction. All analyses and subsequent images were generated using Galaxy online, version 20.01. Four differentially expressed genes were chosen for RT-qPCR validation and the assay was performed as described above (Supplementary Table S4).

Statistical Analyses

Association analyses

EZH2 genotype data was analysed using Modified Quasi-Likelihood Score (M_{QLS})[24], a powerful association analysis method that deals with complex family structure and maximally uses available information whilst also appropriately accounting for relatedness of individuals. It also increases

.

power by inferring genotypes in related untyped individuals whilst permitting the inclusion of unrelated individuals and individuals of unknown phenotype[16]. Genotype data on sampled individuals is treated as random, while phenotype information is fixed, which additionally allows for valid association testing in the presence of possible phenotype misspecification, thereby maximising power but again appropriately managing type 1 error[25]. Evaluation of this method by Cummings *et al.* (2013)[26] has comprehensively examined possible inflation of type 1 error, by performing simulation analyses using a large Amish pedigree structure, revealing no inflation in type 1 error rates, whilst maintaining maximal use of power afforded by the large pedigree.

Linkage Analysis

Linkage analysis was performed for the main PcTas12 pedigree comprising 19 individuals; one female variant carrier and seven non-variant carriers were removed. Given PrCa is a complex trait, model-free linkage analysis was performed, consistent with previous linkage studies. Non-parametric linkage analysis was performed using Merlin 1.1.2 [27].

Gene expression

The unpaired Student's t-test was used to compare absolute gene expression of *DUSP1, FOS, JUNB* and *EGR1* between *EZH2* rs78589034 carriers and non-variant carriers. P-values that were <0.05 were considered to be significant.

RESULTS

EZH2 variant association with prostate cancer risk in the Tasmanian population

WGS was performed in three relatives from a large Tasmanian family exhibiting a dense aggregation of PrCa across multiple generations (PcTas12; Figure 1: with an extended PcTas12 pedigree showing all individuals also presented in Supplementary Figure S1). Interrogation of WGS data revealed four rare candidate PrCa risk variants (see Supplementary Table S6) present in an affected uncle/nephew pair but not in an older unaffected male relative, nor in any of the eight screened Tasmanian controls with WGS data. Additional genotyping in four PcTas12 cases and three relatives revealed that three of the four variants failed to segregate with disease in the larger family, and/or were found to be present in a screened Tasmanian control population (n=88) at a frequency inconsistent with a rare risk variant. The rare *EZH2* variant (rs78589034) was of particular interest given existing evidence in the literature that *EZH2* plays an established role in prostate tumourigenesis. This variant was also subsequently found to segregate with disease in the larger PcTas12 family, following genotyping of 57 additional family members, including three affected men who only had FFPE tumour DNA available. Sequencing of these samples revealed seven additional

.

variant carriers, including five affected males, one older unaffected male relative (diagnosed with bowel cancer and lymphoma), and a daughter of an affected carrier (Figure 1). Non-parametric linkage analysis was performed in a trimmed version of the main pedigree comprising 19 individuals (following exclusion of one female variant carrier and seven non-variant carriers). A single point maximum LOD score of 1.62 (p=0.003) was generated, suggestive of linkage. To examine whether this rare variant was associated with clinicopathological features we examined the clinical data available for six affected variant carriers and five affected non-variant carriers. Analyses revealed no significant difference in the age at diagnosis (p=0.94) or GS (p=0.54) between the two groups (Table 2).

To assess the contribution of rs78589034 to PrCa risk in the wider Tasmanian population, all additional individuals from the familial cohort (n=738) and the case-control study (n=813) were genotyped. A single affected variant carrier was identified in the familial cohort (an isolated case in a branch of PcTas9), and three cases and one control were identified as carriers in the case-control study (the total number of variant carriers in our datasets are displayed in Table 3). Association testing was performed by M_{QLS} analysis[24] which allows the combined familial and case-control genotyping data to be included, whilst appropriately accounting for relatedness. A significant association between PrCa risk and the *EZH2* rs78589034 variant in the Tasmanian population (OR=4.42; $p=1.99x10^{-07}$). As expected, removal of the three PcTas12 individuals included in the original variant discovery yielded a less significant result (OR=3.55; $p=1.20x10^{-05}$). The frequency of the variant in sporadic carriers was 0.64% compared with 0.29% in the unaffected controls (Table 3). The reported MAF in the gnomAD non-Finnish European population is 0.36%, and a global MAF of 0.06% is reported in ClinVar. Removal of the entire PcTas12 pedigree from the M_{QLS} analysis was not performed, as this analysis would have permitted the inclusion of only five variant carriers (one familial case, three sporadic cases and one unaffected control). This highlights the challenges encountered in the identification of rare high-risk variants in complex common disease. The rs78589034 variant lies 4bps from the AG dinucleotide predicted by the Human Splicing Finder as the acceptor splice site. The *in silico* predictive tool, FATHMM-MKL[28] predicts the rs78589034 variant to be deleterious and probably pathogenic (score of 0.69; a score of >0.5 is considered deleterious and a score of 0.7, pathogenic).

EZH2 variant association with prostate cancer risk in a North American population

Through our collaboration with the FHCRC (Seattle, US), WES data from 130 *PROGRESS* families were examined[11,17]. The number of individuals with WES data available in each family ranged from one to six. Of the 321 affected and 11 unaffected older relatives with WES data, six affected and one unaffected man carried the *EZH2* rs78589034 variant, representing seven different families. M_{QLS} analysis of the 332 *PROGRESS* individuals with WES data revealed a significant association between PrCa risk and the *EZH2* rs78589034 variant (p=0.0114). As the M_{QLS} association test

.

utilises all pedigree information available whilst also accounting for relationship between individuals, this result provides preliminary validation of the Tasmanian result. The observation that one *EZH2* variant carrier was detected per family is likely to be influenced by the WES selection strategy for the *PROGRESS* study[11,17]. Uncle-nephew and/or cousin pairs were prioritised for WES, and only a small number of individuals were sequenced in each family (minimum 2 and up to 5). Of the six affected variant carriers, four belonged to families where the most closely related genotyped case to the variant carrier were cousins and/or nephews. In two families where affected brothers were genotyped, only one brother was a carrier in each family. In the seventh family, the single unaffected carrier was the brother of a case, and all remaining cases genotyped were cousins (see Supplementary Figure S3 for all pedigree details).

EZH2 expression in variant and wild-type prostate tumour specimens

Given that alterations in *EZH2* gene expression are strongly associated with the development of a variety of cancers, including PrCa[29,30], *EZH2* expression was examined in prostate tumours from variant and wild-type carriers. FFPE pathology specimens were available for 18 cases, seven of which were identified as *EZH2* rs78589034 rare variant carriers (Supplementary Table S1). EZH2 gene and protein expression was not detectable in the FFPE prostate tumours by RT-PCR or immunohistochemistry, respectively (Supplementary Method 2; data not shown). However, two additional *EZH2* rare variant carriers (PT700 and PT1800) were identified in fresh prostate needle biopsies from radical prostatectomies sourced from the clinical-based cohort (Table 1). Total RNA was extracted from the biopsies of the two sporadic *EZH2* rare variant carriers and four wild-type non-carriers. Evaluation of *EZH2* mRNA levels revealed that all samples expressed low levels of *EZH2*, and expression was similar in rs78589034 rare variant and wild-type carriers (Figure 2). In the Human Protein Atlas, EZH2 expression is also low in prostate tissue, and the majority of prostate tumours display very low expression. Data derived from RNA sequencing analysis of the TCGA-PRAD tumour (n=494) and normal prostate samples (n=152) demonstrates that *EZH2* expression is generally low in the prostate but can vary widely in both tumour and normal prostate samples, with higher expression associated with shorter disease-free survival (Supplementary Figure S4)[31].

Examination of *EZH2* alternate transcripts in variant and wild-type prostate needle biopsies

Disruption of *EZH2* splicing has been observed in, and associated with a variety of cancers[29,32]. The rs78589034 variant is predicted to alter splicing (FATHMM-MKL) and lies 4bp from the AG acceptor site preceding the start of exon 16. A diagrammatic representation of *EZH2* transcripts in the normal prostate is presented in Supplementary Figure S5. Transcriptome sequencing of RNA from the fresh prostate needle biopsies of two *EZH2* variant carriers (PT700 and PT1800) and two non-variant carriers (PT300 and PT4400) was undertaken to determine whether alternate transcripts exist in the

.

variant carriers. Samples were sequenced in duplicate and PCA showed good correlation between the duplicates, and further that rs78589034 variant carrier status strongly influences the tumour transcriptome profile (Figure 3A). Consistent with the RT-qPCR data, low copy numbers of *EZH2* transcripts were detected in both variant and non-variant carriers (Table 4). Multiple different *EZH2* transcripts were detected in each of the four samples, suggesting that *EZH2* differential splicing events are frequent in prostate tissue. Seven different *EZH2* transcripts were identified across the four biopsies, however, no evidence for a correlation between the different transcripts and presence/absence of the variant allele was observed (Table 4). To further check data quality, two genes known to be highly expressed in the prostate, *AMACR* and *HOXB13*, were examined and showed expected levels of expression (Supplementary Table S7).

Given the rs78589034 variant allele is predicted to impact splicing, an *in vitro* splicing assay was undertaken (pSplice Express plasmid; a gift from Professor S Stamm[33]). The transcription of *EZH2* exons 16-19 was assessed using an *in vitro* splice assay employing the pSpliceExpress plasmid containing either the *EZH2* wild-type 'G' allele, or the rare variant 'A' allele, however no negative impact on splicing was observed in cells transfected with plasmids containing either allele (Supplementary Figure S2). However, the presence of the rare 'A' allele at this position is likely to strengthen the polypyrimidine tract (splicing factor recognition sequence), improving splice efficiency. Sequence changes at key sites alter binding affinity of selected splicing factors, thereby altering splicing factor efficiency[34], which may not be detectable using the pSplice Express assay.

Differential gene expression in *EZH2* variant and wild-type prostate needle biopsies

To examine whether there were transcriptome differences evident in the prostate tumours of *EZH2* rare variant carriers, analysis of transcriptome data from the fresh needle biopsies from two rs78589034 variant carriers and two wild-type non-carriers was undertaken. PCA of the transcriptome data showed clustering by genotype, providing evidence for a variant-dependent transcriptomic profile (Figure 3A). Subsequent DESeq2 analysis revealed a gene signature of 25 significantly differentially expressed genes between rs78589034 carriers (PT700 and PT1800) and non-variant carriers (PT300 and PT4400) (Bonferroni corrected p value, $<1.95 \times 10^{-06}$; representing the threshold met for a gene to be considered statistically different between the two groups). Of these 25 genes, 24 were downregulated in the two variant carriers, and one, *PCA3*, was upregulated (Figure 3B). There is also existing published evidence that EZH2 is associated with 14 of these 25 genes (Supplementary Table S8). Four were chosen for RT-qPCR validation; *DUSP1, FOS, JUNB* and *EGR1*. All four were found to be expressed at a lower level in the two *EZH2* variant carriers compared to the two non-variant carriers, validating the transcriptome data (Figure 3C). Whilst all gene expression results trended towards significance, expression of *DUSP1* and *EGR1* were significantly different between the two groups (p=0.004 and 0.012, respectively).

.

DISCUSSION

We observed a rare *EZH2* variant, rs78589034, segregating with PrCa in a large Tasmanian family and subsequently found it to be significantly associated with disease risk in the Tasmanian population (OR=3.55, p=1.2x10^{-05}). Analysis of independent WES data from 130 *PROGRESS* families[11,17] also revealed a statistically significant association of this variant with PrCa risk (p=0.0114), providing supporting evidence for our finding. Furthermore, previous familial PrCa genetic studies report linkage to chromosome 7q31-36[35], which spans the *EZH2* gene. Bachman *et al.* (2005) subsequently provided evidence that an *EZH2* haplotype was associated with risk of familial PrCa, which includes the common rs2072407 variant, located 15bp from the variant identified in our study[36].

The polycomb group (PcG) protein enhancer of zeste homolog 2, *EZH2,* is a histone methyltransferase (HMTase)[37] which plays a catalytic role in the Polycomb Repressive Complex 2 (PRC2)[38]. In cancer, including lymphomas, myelodysplastic and myeloproliferative disorders[39-41], germline and acquired *EZH2* mutations have been observed to have both activating and inactivating effects on tumour development. Interestingly, in our study, a PcTas12 male carrier of the *EZH2* rs78589034 variant was reported to be diagnosed with lymphoma. Although various common somatic *EZH2* variants have been associated with risk of cancer, mainly lymphoma[39-41], their functional impact on *EZH2* remains unexplored. Higher EZH2 expression has been strongly correlated with disease progression, including metastasis, and is associated with disease aggressiveness and a poor prognosis (reviewed in[42]). In prostate tumours, *EZH2* overexpression is mainly attributed to gene amplification, particularly in late-stage disease[42]. However, in our study, *EZH2* was expressed at a low level in fresh prostate biopsies from primary tumours, consistent with publicly available and published data; furthermore levels were comparable between rs78589034 variant and non-variant carriers.

EZH2 function is tightly regulated at the transcriptional, post-transcriptional and post-translational level and tissue-specific regulation is vital for normal tissue growth and differentiation[43]. Splicing dysregulation is one of the molecular hallmarks of cancer. The *EZH2* gene can give rise to over 30 different mRNA transcripts that can exist concurrently in tissues; the functional implications of the majority of these are as yet unknown[44]. Although, there are reports that alternative splicing of exon 14 in *EZH2* plays a major role in the tumorigenesis of renal cancer[32]. The *EZH2* rs78589034 variant lies 4bp downstream from the AG dinucleotide, an acceptor splice site preceding exon 16[45]. Exon 16 encodes residues included in the SET domain of the EZH2 protein, which is responsible for its HMTase activity.

The spliceosome comprises small nuclear ribonucleoprotein complexes including the U2 auxiliary factor (U2AF) proteins. These proteins bind and proof-read polypyrimidine tracts, which

precede AG acceptor splice sites. Studies of the physical interaction between U2AF proteins and their recognition sequences have shown that sequence changes at specific sites alter the strength of binding and impact downstream function[46]. Introduction of the rs78589034 variant introduces a run of uracil's close to the AG splice site potentially enhancing splice factor binding. Thus, it is possible that enhanced efficiency of *EZH2* splicing (to include exon 16) may increase the HMTase activity of EZH2, resulting in downstream effects on target genes. Our examination of the prostate biopsy transcriptomes from two rs78589034 variant carriers did not reveal any clear differences in *EZH2* splicing, however individual transcript counts were few and it is known that assembly of low-level transcripts is challenging[47].

Therefore, we sought to determine whether there were detectable allele dependent transcriptome differences in prostate tissue between variant and non-variant carriers. *EZH2* rs78589034 carrier status was demonstrated to impact the transcriptome signature as indicated by the PCA analysis. A 25 gene signature of significantly differentially expressed transcripts between *EZH2* rs78589034 variant carriers and non-variant carriers was identified. Top ranked genes in this signature are known downstream targets of *EZH2* and play a proven role in tumorigenesis (Supplementary Table S8). For example, EZH2 binds to the DUSP1 promoter inducing H3K27 trimethylation, which subsequently suppresses DUSP1 expression[48]. *EZH2* also directly regulates the key tumour suppressor, miRNA101, in a negative feedback loop[49] and decreasing levels of miRNA101 are associated with multiple tumours, including colon, liver, lung, ovarian and prostate[50]. Although miRNA101 was not detectable in our samples (the RNA extraction method was not specifically designed to capture miRNAs), this miRNA is known to regulate *DUSP1*[51], *FOS*[52] and *JUNB*[53], and expression of these genes was significantly downregulated in rs78589034 carriers. Finally, *EGR1*, like *EZH2*, is regulated by the PRC2 complex in PrCa and downregulation of *EGR1* is associated with tumour progression in PRC2-dependent tumours[54]. Although the evidence remains circumstantial, the observed changes in EZH2-regulated genes are consistent with a gain of function of EZH2 in variant carriers.

There remains the need to replicate the association of this rare, intronic *EZH2* variant in independent familial and sporadic datasets, in addition to individuals of other ethnicities. A further limitation of our study was that the availability of FFPE and fresh needle biopsies from rs78589034 variant carriers, limited our opportunity to fully elucidate the role of this variant on *EZH2* splicing. It is also recognised that the use of CADD to generate the initial list of prioritised variants, as opposed to a combination of predictive tools, would likely influence the variants for investigation. CADD was selected as it offers a less-biased larger-genomic training set than many other tools, which are trained on comparatively fewer genomic variants for which the pathogenic status is known[55]. Although CADD broadly performs well for both coding and non-coding variants, it is not specifically designed to predict functionality of splice-site variants. Therefore, an algorithm more suitable for intronic variants, FATHMM-MKL, was also used[28]. This tool also provided evidence that rs78589034

.

influences functionality of the splice site. Taken together, the evidence presented here raises the interesting possibility that this variant may be associated with chronically enhanced EZH2 activity, which may in turn contribute to prostate tumour development.

A variety of specific EZH2 inhibitors are in phase I/II clinical trials, in combination with standard therapies for metastatic castration-resistant PrCa. More recent interest has centred on their use for the treatment of earlier-stage androgen-responsive tumours (discussed in Mollica *et al.* (2019)[56]. The testing of EZH2-targeted therapies has also highlighted the value of molecular phenotyping patient populations, to better target therapies to those most likely to benefit. Further, there is also a strong interest in the characterisation of disrupted gene splicing in tumour development, particularly the impact of splicing on radiation and chemotherapy resistance[57].

In summary, this study has provided evidence of a significant association between the *EZH2* rs78589034 variant and PrCa risk in two distinct Caucasian populations. Here, we have also provided circumstantial evidence that this variant may be associated with a gain of function of EZH2, however this remains to be established. Importantly, this study highlights the complexities encountered in unravelling the genetic determinants of PrCa. In evaluating candidate genetic risk variants and their likelihood as being causative, one must also consider that prostate tumours are often slow to progress and may have developed over a long period of time. Despite the challenges encountered, the pursuit to improve our understanding of genetic factors in disease risk should remain a priority if we are to deliver better screening, diagnostic and therapeutic options for PrCa patients.

ACKNOWLEDGEMENTS

We are greatly indebted to the participants of our prostate cancer studies, both those participating in the Tasmanian Prostate Cancer Studies, and those participating in the Fred Hutchinson Cancer Research Centre led, PROGRESS study. In addition, we would like to thank Tasmanian Cancer Registry staff, Tasmanian urologists, pathologists and the wider Tasmanian clinical and research community for their ongoing support. In particular, Hobart clinicians, Mr Ian Middleton, Dr Marketa Skala, Dr Frank Redwig and Mr Anthony Eaton for supporting the *Tasmanian Clinical-Based Prostate Cancer Biopsy Cohort.*

CONFLICT OF INTEREST

The authors declare no competing financial interests.

DATA AVAILABILITY STATEMENT

.

RNA-seq data generated in this study are available in ArrayExpress under accession number E-MTAB-10271. Additional data including genotyping data are either available in the electronic Supplementary Material document, or from the corresponding author on reasonable request.

ETHICS STATEMENT

Ethics approval for all Tasmanian cohort studies was obtained from the Human Research Ethics Committee Tasmania, Australia (H0017040) and written informed consent was obtained from all participating individuals. For deceased familial cases, a waiver of consent was obtained to collect prostate tumour tissue specimens**.** Written informed consent was obtained from all participants of the *PROGRESS* study was approved by the Institutional Review Board at the FHCRC.

FUNDING

JLD is supported by a Select Foundation Cancer Research Fellowship and was previously supported by an Australian Research Council Future Fellowship. KR has previously been supported by an Australian Government Research Training Program Scholarship and a Cancer Council Tasmania Evelyn Pedersen Elite Research Scholarship. KR is now supported by funding from the Royal Hobart Hospital Research Foundation. LMF is supported by a Cancer Council Tasmania/College of Health and Medicine Senior Research Fellowship. MAF is supported by NHMRC APP5121190. The *Tasmanian Familial Prostate Cancer Study* has been supported by the Royal Hobart Hospital Research Foundation, Cancer Council Tasmania, Cancer Australia, The Mazda Foundation, Perpetual Trustees, Max Bruce Trust, The Estate of Dr RA Parker, and the Tasmanian Community Fund.

AUTHORS' CONTRIBUTIONS

KR and JLD wrote the main manuscript and KR prepared the tables and figures. KR and JRM performed the laboratory work. KR, MAF, RJT and GF performed the analyses. AB traced the participants and RM and SD sourced and reviewed the pathology material. JLD and LMF directed the study, and all authors reviewed the manuscript.

.

REFERENCES

1 Attard G, Parker C, Eeles RA, Schroder F, Tomlins SA, Tannock I, Drake CG, de Bono JS. Prostate cancer. Lancet 2016; 387: 70-82.

2 Hjelmborg JB, Scheike T, Holst K, Skytthe A, Penney KL, Graff RE, Pukkala E, Christensen K, Adami HO, Holm NV, Nuttall E, Hansen S, Hartman M, Czene K, Harris JR, Kaprio J, Mucci LA. The heritability of prostate cancer in the Nordic Twin Study of Cancer. Cancer Epidemiol Biomarkers Prev 2014; 23: 2303-2310.

3 Benafif S, Kote-Jarai Z, Eeles RA, Consortium P. A Review of Prostate Cancer Genome-Wide Association Studies (GWAS). Cancer Epidemiol Biomarkers Prev 2018; 27: 845-857.

4 Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, Dadaev T, Leongamornlert D, Anokian E, Cieza-Borrella C, Goh C, Brook MN, Sheng X, Fachal L, Dennis J, Tyrer J, Muir K, Lophatananon A, Stevens VL, Gapstur SM, Carter BD, Tangen CM, Goodman PJ, Thompson IM, Jr., Batra J, Chambers S, Moya L, Clements J, Horvath L, Tilley W, Risbridger GP, Gronberg H, Aly M, Nordstrom T, Pharoah P, Pashayan N, Schleutker J, Tammela TLJ, Sipeky C, Auvinen A, Albanes D, Weinstein S, Wolk A, Hakansson N, West CML, Dunning AM, Burnet N, Mucci LA, Giovannucci E, Andriole GL, Cussenot O, Cancel-Tassin G, Koutros S, Beane Freeman LE, Sorensen KD, Orntoft TF, Borre M, Maehle L, Grindedal EM, Neal DE, Donovan JL, Hamdy FC, Martin RM, Travis RC, Key TJ, Hamilton RJ, Fleshner NE, Finelli A, Ingles SA, Stern MC, Rosenstein BS, Kerns SL, Ostrer H, Lu YJ, Zhang HW, Feng N, Mao X, Guo X, Wang G, Sun Z, Giles GG, Southey MC, MacInnis RJ, FitzGerald LM, Kibel AS, Drake BF, Vega A, Gomez-Caamano A, Szulkin R, Eklund M, Kogevinas M, Llorca J, Castano-Vinyals G, Penney KL, Stampfer M, Park JY, Sellers TA, Lin HY, Stanford JL, Cybulski C, Wokolorczyk D, Lubinski J, Ostrander EA, Geybels MS, Nordestgaard BG, Nielsen SF, Weischer M, Bisbjerg R, Roder MA, Iversen P, Brenner H, Cuk K, Holleczek B, Maier C, Luedeke M, Schnoeller T, Kim J, Logothetis CJ, John EM, Teixeira MR, Paulo P, Cardoso M, Neuhausen SL, Steele L, Ding YC, De Ruyck K, De Meerleer G, Ost P, Razack A, Lim J, Teo SH, Lin DW, Newcomb LF, Lessel D, Gamulin M, Kulis T, Kaneva R, Usmani N, Singhal S, Slavov C, Mitev V, Parliament M, Claessens F, Joniau S, Van den Broeck T, Larkin S, Townsend PA, Aukim-Hastie C, Dominguez MG, Castelao JE, Martinez ME, Roobol MJ, Jenster G, van Schaik RHN, Menegaux F, Truong T, Koudou YA, Profile S, Xu J, Khaw KT, Cannon-Albright L, Pandha H, Michael A, Thibodeau SN, McDonnell SK, Schaid DJ, Lindstrom S, Turman C, Ma J, Hunter DJ, Riboli E, Siddiq A, Canzian F, Kolonel LN, Le Marchand L, Hoover RN, Machiela MJ, Cui Z, Kraft P, Australian Prostate Cancer B, Study I, Canary PI, Breast, Prostate Cancer Cohort C, Consortium P, Cancer of the Prostate in S, Prostate Cancer Genome-wide Association Study of Uncommon Susceptibility L, Genetic A, Mechanisms in Oncology /Elucidating Loci Involved in Prostate Cancer Susceptibility C, Amos CI, Conti DV, Easton DF, Wiklund F, Chanock SJ, Henderson BE, Kote-Jarai Z, Haiman CA, Eeles RA. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet 2018; 50: 928-936.

5 Mancuso N, Rohland N, Rand KA, Tandon A, Allen A, Quinque D, Mallick S, Li H, Stram A, Sheng X, Kote-Jarai Z, Easton DF, Eeles RA, consortium P, Le Marchand L, Lubwama A, Stram D, Watya S, Conti DV, Henderson B, Haiman CA, Pasaniuc B, Reich D. The contribution of rare variation to prostate cancer heritability. Nat Genet 2016; 48: 30-35.

6 Rosenthal SL, Bamne MN, Wang X, Berman S, Snitz BE, Klunk WE, Sweet RA, Demirci FY, Lopez OL, Kamboh MI. More evidence for association of a rare TREM2 mutation (R47H) with Alzheimer's disease risk. Neurobiol Aging 2015; 36: 2443 e2421-2446.

7 Berger MF, Mardis ER. The emerging clinical relevance of genomics in cancer medicine. Nat Rev Clin Oncol 2018; 15: 353-365.

8 Ewing CM, Ray AM, Lange EM, Zuhlke KA, Robbins CM, Tembe WD, Wiley KE, Isaacs SD, Johng D, Wang Y, Bizon C, Yan G, Gielzak M, Partin AW, Shanmugam V, Izatt T, Sinari S, Craig DW, Zheng SL, Walsh PC, Montie JE, Xu J, Carpten JD, Isaacs WB, Cooney KA. Germline mutations in HOXB13 and prostate-cancer risk. N Engl J Med 2012; 366: 141-149.

9 Breyer JP, Avritt TG, McReynolds KM, Dupont WD, Smith JR. Confirmation of the HOXB13 G84E germline mutation in familial prostate cancer. Cancer Epidemiol Biomarkers Prev 2012; 21: 1348-1353.

.

10 Fitzgerald LM, Kumar A, Boyle EA, Zhang Y, McIntosh LM, Kolb S, Stott-Miller M, Smith T, Karyadi DM, Ostrander EA, Hsu L, Shendure J, Stanford JL. Germline missense variants in the BTNL2 gene are associated with prostate cancer susceptibility. Cancer Epidemiol Biomarkers Prev 2013; 22: 1520-1528.

11 Karyadi DM, Geybels MS, Karlins E, Decker B, McIntosh L, Hutchinson A, Kolb S, McDonnell SK, Hicks B, Middha S, FitzGerald LM, DeRycke MS, Yeager M, Schaid DJ, Chanock SJ, Thibodeau SN, Berndt SI, Stanford JL, Ostrander EA. 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 2017; 8: 1495-1507.

12 Vysotskaia V, Kaseniit KE, Bucheit L, Ready K, Price K, Johansen Taber K. Clinical utility of hereditary cancer panel testing: Impact of PALB2, ATM, CHEK2, NBN, BRIP1, RAD51C, and RAD51D results on patient management and adherence to provider recommendations. Cancer 2020; 126: 549-558.

13 Paulo P, Maia S, Pinto C, Pinto P, Monteiro A, Peixoto A, Teixeira MR. Targeted next generation sequencing identifies functionally deleterious germline mutations in novel genes in early-onset/familial prostate cancer. PLoS Genet 2018; 14: e1007355.

14 Isaacs WB, Xu J. Current progress and questions in germline genetics of prostate cancer. Asian J Urol 2019; 6: 3-9.

15 Rubio JP, Bahlo M, Butzkueven H, van Der Mei IA, Sale MM, Dickinson JL, Groom P, Johnson LJ, Simmons RD, Tait B, Varney M, Taylor B, Dwyer T, Williamson R, Gough NM, Kilpatrick TJ, Speed TP, Foote SJ. Genetic dissection of the human leukocyte antigen region by use of haplotypes of Tasmanians with multiple sclerosis. Am J Hum Genet 2002; 70: 1125-1137.

16 FitzGerald LM, Raspin K, Marthick JR, Field MA, Malley RC, Thomson RJ, Blackburn NB, Banks A, Charlesworth JC, Donovan S, Dickinson JL. Impact of the G84E variant on HOXB13 gene and protein expression in formalin-fixed, paraffin-embedded prostate tumours. Sci Rep 2017; 7: 17778.

17 Schaid DJ, McDonnell SK, FitzGerald LM, DeRycke L, Fogarty Z, Giles GG, MacInnis RJ, Southey MC, Nguyen-Dumont T, Cancel-Tassin G, Cussenot O, Whittemore AS, Sieh W, Ioannidis NM, Hsieh CL, Stanford JL, Schleutker J, Cropp CD, Carpten J, Hoegel J, Eeles R, Kote-Jarai Z, Ackerman MJ, Klein CJ, Mandal D, Cooney KA, Bailey-Wilson JE, Helfand B, Catalona WJ, Wiklund F, Riska S, Bahetti S, Larson MC, Cannon Albright L, Teerlink C, Xu J, Isaacs W, Ostrander EA, Thibodeau SN. Two-stage Study of Familial Prostate Cancer by Whole-exome Sequencing and Custom Capture Identifies 10 Novel Genes Associated with the Risk of Prostate Cancer. Eur Urol 2021; 79: 353-361.

18 Field MA, Cho V, Cook MC, Enders A, Vinuesa CG, Whittle B, Andrews TD, Goodnow CC. Reducing the search space for causal genetic variants with VASP. Bioinformatics 2015; 31: 2377-2379.

19 Field MA, Cho V, Andrews TD, Goodnow CC. Reliably Detecting Clinically Important Variants Requires Both Combined Variant Calls and Optimized Filtering Strategies. PLoS One 2015; 10: e0143199.

20 Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014; 46: 310-315.

21 Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 2012; 13: 134.

22 Mori R, Wang Q, Danenberg KD, Pinski JK, Danenberg PV. Both beta-actin and GAPDH are useful reference genes for normalization of quantitative RT-PCR in human FFPE tissue samples of prostate cancer. Prostate 2008; 68: 1555-1560.

23 Oakford PC, James SR, Qadi A, West AC, Ray SN, Bert AG, Cockerill PN, Holloway AF. Transcriptional and epigenetic regulation of the GM-CSF promoter by RUNX1. Leuk Res 2010; 34: 1203-1213.

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

25 Hong X, Hao K, Ladd-Acosta C, Hansen KD, Tsai HJ, Liu X, Xu X, Thornton TA, Caruso D, Keet CA, Sun Y, Wang G, Luo W, Kumar R, Fuleihan R, Singh AM, Kim JS, Story RE, Gupta RS, Gao P, Chen Z, Walker SO, Bartell TR, Beaty TH, Fallin MD, Schleimer R, Holt PG, Nadeau KC, Wood RA, Pongracic JA, Weeks DE, Wang X. Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children. Nat Commun 2015; 6: 6304.

.

26 Cummings AC, Torstenson E, Davis MF, D'Aoust LN, Scott WK, Pericak-Vance MA, Bush WS, Haines JL. Evaluating power and type 1 error in large pedigree analyses of binary traits. PLoS One 2013; 8: e62615.
27 Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 2002; 30: 97-101.
28 Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day IN, Gaunt TR, Campbell C. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. Bioinformatics 2015; 31: 1536-1543.
29 Chase A, Cross NC. Aberrations of EZH2 in cancer. Clin Cancer Res 2011; 17: 2613-2618.
30 Chen Z, Yang P, Li W, He F, Wei J, Zhang T, Zhong J, Chen H, Cao J. Expression of EZH2 is associated with poor outcome in colorectal cancer. Oncol Lett 2018; 15: 2953-2961.
31 Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017; 45: W98-W102.
32 Chen K, Xiao H, Zeng J, Yu G, Zhou H, Huang C, Yao W, Xiao W, Hu J, Guan W, Wu L, Huang J, Huang Q, Xu H, Ye Z. Alternative Splicing of EZH2 pre-mRNA by SF3B3 Contributes to the Tumorigenic Potential of Renal Cancer. Clin Cancer Res 2017; 23: 3428-3441.
33 Kishore S, Khanna A, Stamm S. Rapid generation of splicing reporters with pSpliceExpress. Gene 2008; 427: 104-110.
34 Chang JW, Yeh HS, Park M, Erber L, Sun J, Cheng S, Bui AM, Fahmi NA, Nasti R, Kuang R, Chen Y, Zhang W, Yong J. mTOR-regulated U2af1 tandem exon splicing specifies transcriptome features for translational control. Nucleic Acids Res 2019; 47: 10373-10387.
35 Janer M, Friedrichsen DM, Stanford JL, Badzioch MD, Kolb S, Deutsch K, Peters MA, Goode EL, Welti R, DeFrance HB, Iwasaki L, Li S, Hood L, Ostrander EA, Jarvik GP. Genomic scan of 254 hereditary prostate cancer families. Prostate 2003; 57: 309-319.
36 Bachmann N, Hoegel J, Haeusler J, Kuefer R, Herkommer K, Paiss T, Vogel W, Maier C. Mutation screen and association study of EZH2 as a susceptibility gene for aggressive prostate cancer. Prostate 2005; 65: 252-259.
37 Vire E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, Morey L, Van Eynde A, Bernard D, Vanderwinden JM, Bollen M, Esteller M, Di Croce L, de Launoit Y, Fuks F. The Polycomb group protein EZH2 directly controls DNA methylation. Nature 2006; 439: 871-874.
38 Karanikolas BD, Figueiredo ML, Wu L. Polycomb group protein enhancer of zeste 2 is an oncogene that promotes the neoplastic transformation of a benign prostatic epithelial cell line. Mol Cancer Res 2009; 7: 1456-1465.
39 Bodor C, O'Riain C, Wrench D, Matthews J, Iyengar S, Tayyib H, Calaminici M, Clear A, Iqbal S, Quentmeier H, Drexler HG, Montoto S, Lister AT, Gribben JG, Matolcsy A, Fitzgibbon J. EZH2 Y641 mutations in follicular lymphoma. Leukemia 2011; 25: 726-729.
40 Majer CR, Jin L, Scott MP, Knutson SK, Kuntz KW, Keilhack H, Smith JJ, Moyer MP, Richon VM, Copeland RA, Wigle TJ. A687V EZH2 is a gain-of-function mutation found in lymphoma patients. FEBS Lett 2012; 586: 3448-3451.
41 Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R, Paul JE, Boyle M, Woolcock BW, Kuchenbauer F, Yap D, Humphries RK, Griffith OL, Shah S, Zhu H, Kimbara M, Shashkin P, Charlot JF, Tcherpakov M, Corbett R, Tam A, Varhol R, Smailus D, Moksa M, Zhao Y, Delaney A, Qian H, Birol I, Schein J, Moore R, Holt R, Horsman DE, Connors JM, Jones S, Aparicio S, Hirst M, Gascoyne RD, Marra MA. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet 2010; 42: 181-185.
42 Sauvageau M, Sauvageau G. Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. Cell Stem Cell 2010; 7: 299-313.
43 Chou RH, Yu YL, Hung MC. The roles of EZH2 in cell lineage commitment. Am J Transl Res 2011; 3: 243-250.
44 Grzenda A, Lomberk G, Svingen P, Mathison A, Calvo E, Iovanna J, Xiong Y, Faubion W, Urrutia R. Functional characterization of EZH2beta reveals the increased complexity of EZH2 isoforms involved in the regulation of mammalian gene expression. Epigenetics Chromatin 2013; 6: 3.

.

45 Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009; 37: e67.
46 Shiozawa Y, Malcovati L, Galli A, Sato-Otsubo A, Kataoka K, Sato Y, Watatani Y, Suzuki H, Yoshizato T, Yoshida K, Sanada M, Makishima H, Shiraishi Y, Chiba K, Hellstrom-Lindberg E, Miyano S, Ogawa S, Cazzola M. Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. Nat Commun 2018; 9: 3649.
47 Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 2016; 11: 1650-1667.
48 Chen Z, Chen Q, Cheng Z, Gu J, Feng W, Lei T, Huang J, Pu J, Chen X, Wang Z. Long non-coding RNA CASC9 promotes gefitinib resistance in NSCLC by epigenetic repression of DUSP1. Cell Death Dis 2020; 11: 858.
49 Huang D, Wang X, Zhuang C, Shi W, Liu M, Tu Q, Zhang D, Hu L. Reciprocal negative feedback loop between EZH2 and miR-101-1 contributes to miR-101 deregulation in hepatocellular carcinoma. Oncol Rep 2016; 35: 1083-1090.
50 Frediani JN, Fabbri M. Essential role of miRNAs in orchestrating the biology of the tumor microenvironment. Mol Cancer 2016; 15: 42.
51 Ye Y, Bao C, Fan W. Overexpression of miR-101 May Target DUSP1 to Promote the Cartilage Degradation in Rheumatoid Arthritis. J Comput Biol 2019; 26: 1067-1079.
52 Llorens F, Banez-Coronel M, Pantano L, del Rio JA, Ferrer I, Estivill X, Marti E. A highly expressed miR-101 isomiR is a functional silencing small RNA. BMC Genomics 2013; 14: 104.
53 Wang L, Zhang X, Jia LT, Hu SJ, Zhao J, Yang JD, Wen WH, Wang Z, Wang T, Zhao J, Wang RA, Meng YL, Nie YZ, Dou KF, Chen SY, Yao LB, Fan DM, Zhang R, Yang AG. c-Myc-mediated epigenetic silencing of MicroRNA-101 contributes to dysregulation of multiple pathways in hepatocellular carcinoma. Hepatology 2014; 59: 1850-1863.
54 Rizq O, Mimura N, Oshima M, Saraya A, Koide S, Kato Y, Aoyama K, Nakajima-Takagi Y, Wang C, Chiba T, Ma A, Jin J, Iseki T, Nakaseko C, Iwama A. Dual Inhibition of EZH2 and EZH1 Sensitizes PRC2-Dependent Tumors to Proteasome Inhibition. Clin Cancer Res 2017; 23: 4817-4830.
55 Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 2019; 47: D886-D894.
56 Mollica V, Di Nunno V, Cimadamore A, Lopez-Beltran A, Cheng L, Santoni M, Scarpelli M, Montironi R, Massari F. Molecular Mechanisms Related to Hormone Inhibition Resistance in Prostate Cancer. Cells 2019; 8.
57 Di C, Syafrizayanti, Zhang Q, Chen Y, Wang Y, Zhang X, Liu Y, Sun C, Zhang H, Hoheisel JD. Function, clinical application, and strategies of Pre-mRNA splicing in cancer. Cell Death Differ 2019; 26: 1181-1194.

.

FIGURE LEGENDS

Figure 1. Segregation of the *EZH2* variant in the Tasmanian prostate cancer pedigree, PcTas12. PcTas12 pedigree depicting the number and relationships of PrCa cases (shaded squares), as well as *EZH2* rs78589034 carrier status. Squares indicate males and circles females, with a slash indicating the subject is deceased. The disease status of earlier generations is generally unknown, unless this information was obtained from clinical records. The rs78589034 variant (+) was originally identified in individuals, PcTas12-1 and PcTas12-132, by WGS (indicted by an arrow). An additional seven carriers were identified by subsequent genotyping, including one older unaffected male relative diagnosed with bowel cancer and lymphoma (PcTas12-73). Individual identification numbers, age at diagnosis of cases and availability of FFPE tissue are shown under each symbol. We note that PcTas12-33 also has a strong family history of PrCa, with potential bilineal status for his progeny. However, all 11 individuals in his extended pedigree with DNA available were *EZH2* rs78589034 non-variant carriers (including PcTas12-02, 05 and 254; please refer to Supplementary Figure S1).

.

Figure 2. *EZH2* gene expression analysis in fresh prostate needle biopsies. The *EZH2* rs78589034 carriers (PT700 and PT1800) had levels of expression similar to the non-variant carriers (n=4; Note: 0.1 and 0.2 represent two different biopsies from a single radical prostatectomy). *EZH2* expression was normalised to the expression of housekeeping genes, *β-Actin* and *GAPDH*; here, the normalised expression is plotted as absolute gene expression. Error bars represent the difference between the technical triplicates performed for each sample.

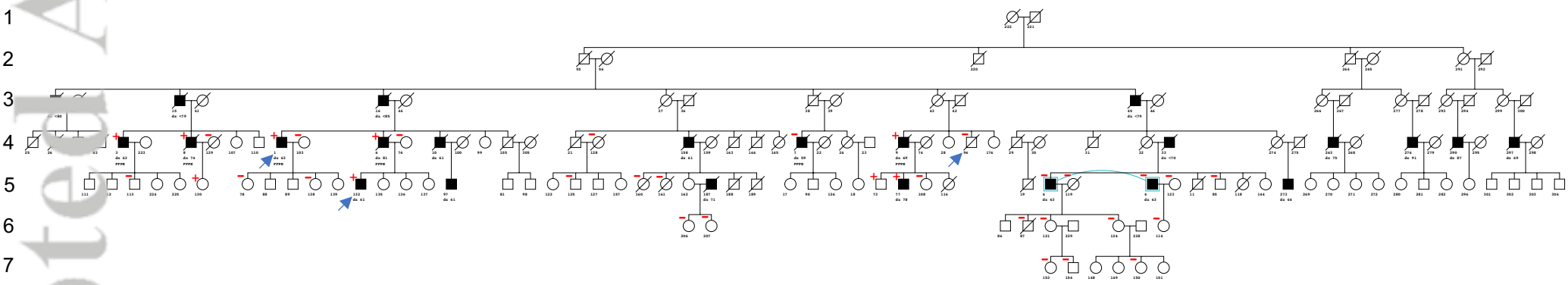
.

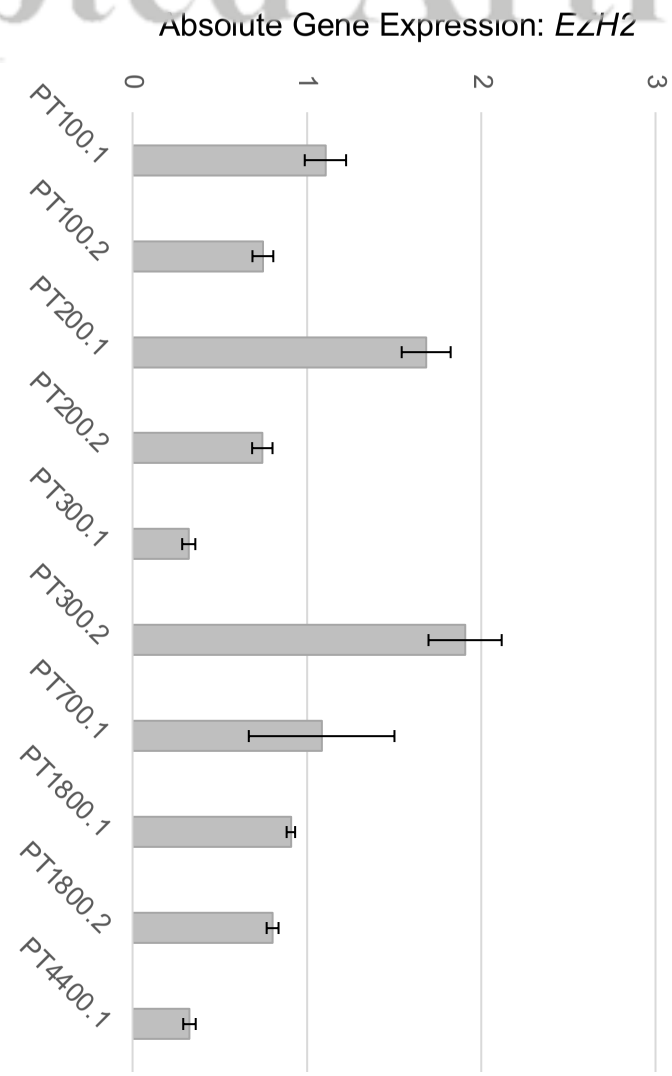
Figure 3. Significantly differentially expressed genes between *EZH2* rs78589034 carriers and non-variant carriers.

(**A**) Principal component analysis of transcriptome data generated from fresh prostate needle biopsies. Samples were run in duplicate across separate lanes, and each pair of samples showed good correlation (plot generated using Galaxy online, version 20.01).

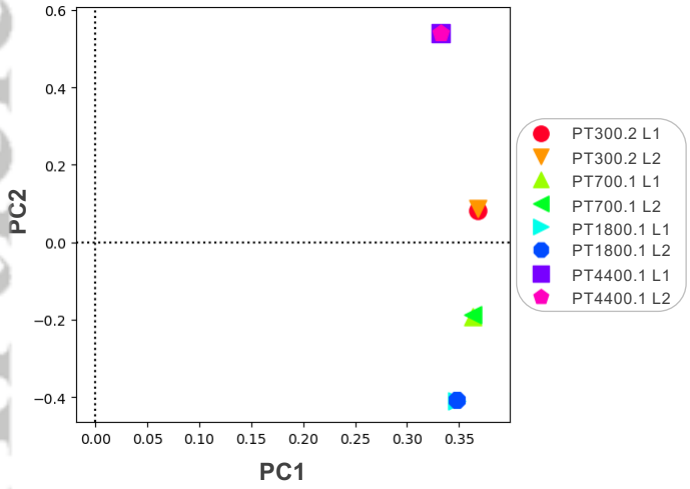
(**B**) Significantly differentially expressed genes between *EZH2* variant carriers (n=2) and non-variant carriers (n=2), observed in transcriptome data from fresh prostate needle biopsies. *DUSP1, FOS, JUNB* and *EGR1* expression were normalised to the expression of housekeeping genes, *β-Actin* and *GAPDH*; here, the normalised expression is plotted as absolute gene expression. Error bars represent the difference between the technical triplicates performed for each sample.

(**C**) RT-qPCR gene expression data validated the decreased expression of *DUSP1, FOS, JUNB* and *EGR1* in *EZH2* variant carriers (PT700.1 and PT1800.1) *versus* non-variant carriers (PT300.2 and PT4400.1). *statistically significant ($p < 0.01$). #Expression is presented as 1.0×10^2.

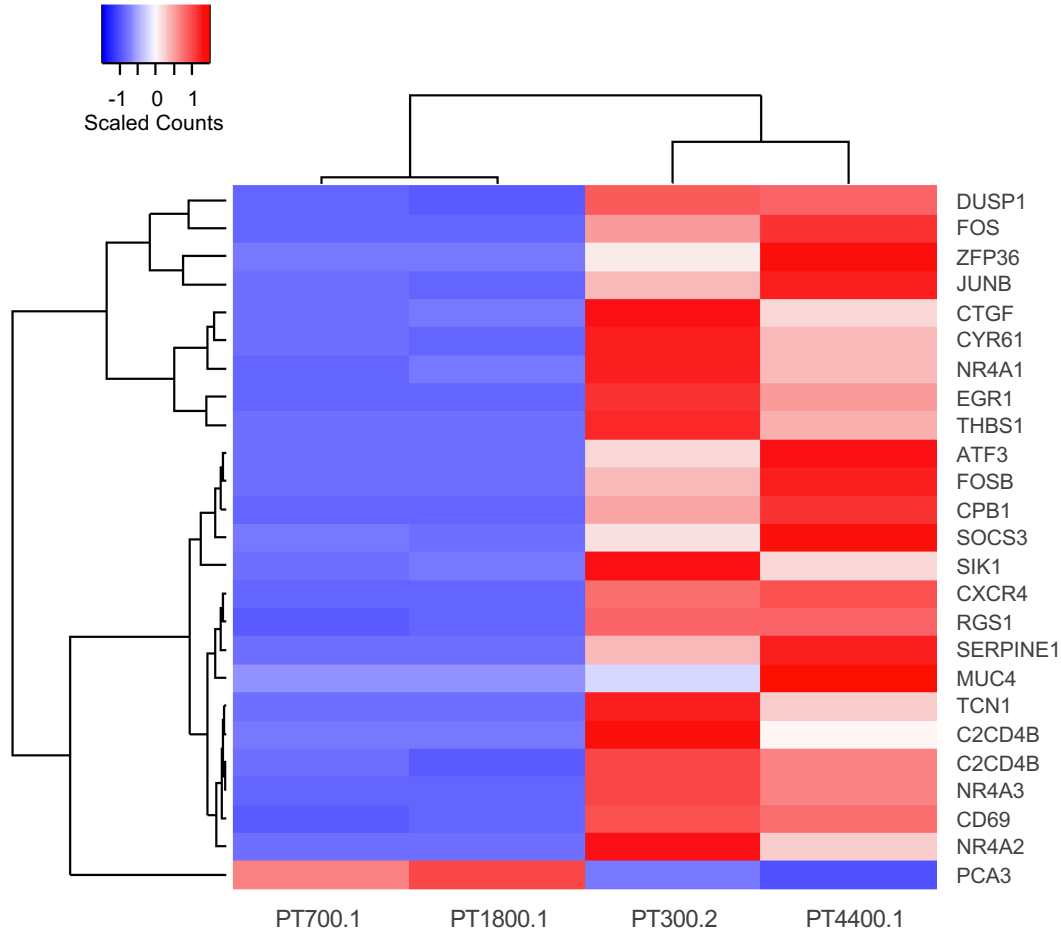




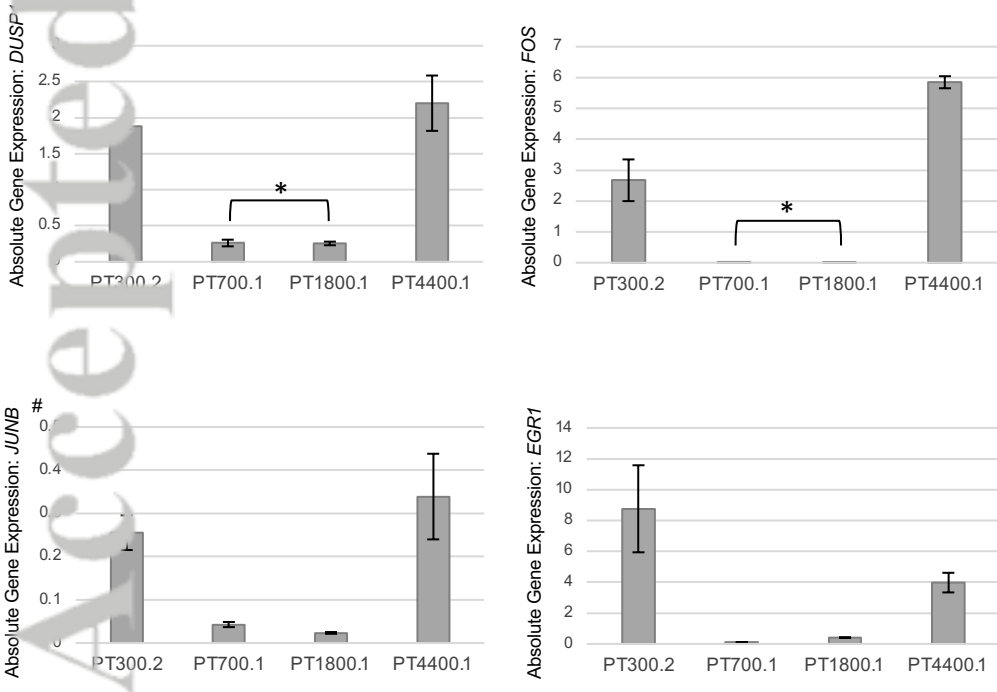
(A)



(B)



(C)



TABLES

Table 1. Clinicopathological characteristics of fresh prostate needle biopsies obtained for *EZH2* rs78589034 carriers and non-variant carriers.

Identification	Age at Diagnosis	Tissue Source	Germline Genotype	Tumour Genotype	Tumour Grade[1]	Contemporary Gleason Score[2]
PT100	70	Fresh needle biopsy	N/A	GG	PD	9 (4+5)
PT200	73	Fresh needle biopsy	N/A	GG	PD	9 (4+5)
PT300	61	Fresh needle biopsy	N/A	GG	MD	7 (4+3)
PT4400	64	Fresh needle biopsy	N/A	GG	PD	10 (5+5)
PT700	75	Fresh needle biopsy	N/A	GA	MD	7 (4+3)
PT1800	59	Fresh needle biopsy	N/A	GA	WD	6 (3+3)

N/A: sample not available; [1]Tumour grade obtained from pathology report; [2]GS was obtained from the original prostatectomy pathology report; WD: well differentiated; MD: moderately differentiated; PD: poorly differentiated.

Table 2. Clinicopathological characteristics of prostate cancer cases from the PcTas12 family, including *EZH2* rs78589034 carriers and non-variant carriers.

Identification	Age at Diagnosis	Germline Genotype	Tumour Genotype	Tumour Grade[1]	Contemporary Gleason Score[2]
PcTas12-02	80	GG	N/A	MD	6 (3+3)
PcTas12-04	63	GG	N/A	MD	6 (3+3)
PcTas12-05	64	GG	N/A	WD	-
PcTas12-07	59	N/A	GG	PD	9 (4+5)
PcTas12-254	75	GG	N/A	WD	6 (3+3)
PcTas12-01	63	GA	GA	MD	6 (3+3)
PcTas12-03	62	N/A	GA	WD	4 (2+2)
PcTas12-06	80	N/A	GA	PD	7 (3+4)
PcTas12-08	73	N/A	GA	-	6 (3+3)
PcTas12-09	68	N/A	GA	-	6 (3+3)
PcTas12-132	61	GA	GA	-	8 (4+4)

N/A: sample not available; [1]Tumour grade obtained from pathology report; [2]GS obtained from pathology report; WD: well differentiated; MD: moderately differentiated; PD: poorly differentiated; -: information not present in original pathology report. Please note: PcTas12-02, 05 and 254 are from PcTas12-33's extended family.

22 **Table 3. Association analysis of rs78589034 with prostate cancer risk in the combined Tasmanian familial and case-control datasets.**

23

Familial Case Carriers	Familial Unaffected Carriers	Sporadic Case Carriers	Control Carriers	gnomAD non-Finnish European MAF	Odds Ratio	p-value
8/293 (2.73%)	2/434 (0.46%)	3/470 (0.64%)	1/339 (0.29%)	0.36%	3.55	$1.2x10^{-05}$
A genotype for rs78589034 was unable to be determined for 11 unaffected family members, two sporadic cases and two controls.						

Table 4. *EZH2* transcripts present in the fresh prostate needle biopsies.

Biopsy ID	*EZH2* Genotype	Chromosome: base pair	Total number of exons included in transcript	RPKM[1]
PT300.2	GG	7:148,807,384-7:148,847,290	19	3.11
		7:148,816,551 – 7:148,819,019	3[2]	0.85
PT700.1	GA	7:148,807,398 - 7:148,814,290	6[3]	1.35
		7:148,807,398 - 7:148,826,512	13	1.57
		7:148,807,398 - 7:148,826,512	12	1.30
		7:148,827,189 - 7:148,829,772	3[2]	1.15
PT1800.1	GA	7:148,807,451 – 7:148,832,744	17	2.19
PT4400.1	GG	7: 148,807,399 – 7:148,809,616	3[2]	3.48
		7:148,810,382 – 7:148,816,872	6[3]	1.55
		7:148,817,111 – 7:148,818,029	2	0.44

[1]Reads Per Kilobase of transcript per Million mapped reads; [2,3]Represents the same *EZH2* transcript.