

Pathway analysis of genes identified through Post-GWAS to understand prostate cancer aetiology

Samaneh Farashi^{1,2}, Thomas Kryza^{1,2,3}, Jyotsna Batra^{1,2*}

¹Cancer Program, School of Biomedical Sciences, Institute of Health and Biomedical Innovation, Australian Prostate Cancer Research Centre – Queensland, Queensland University of Technology, Brisbane, Queensland, 4102, Australia.

²Translational Research Institute, 37 Kent Street, Woolloongabba, Queensland, 4102, Australia.

³Mater Research Institute, University of Queensland, Translational Research Institute, 37 Kent Street, Woolloongabba, Queensland, 4102, Australia.

*Email: jyotsna.batra@qut.edu.au

Abstract

Understanding the role of risk regions identified by genome-wide association studies (GWAS) have made considerable progress lately referred to the post-GWAS era. Annotation of the genes to the GWAS and fine-mapped functional variants, and understanding their biological pathway/gene networks enrichments is expected to give rich dividend by elucidating the mechanisms underlying prostate cancer. To this aim, we compiled and analysed currently available post-GWAS data on genes identified through GWAS and validated through experimental studies in prostate cancer to investigate molecular biological pathways enriched for assigned functional genes. The results highlight some well-known cancer signalling pathways, antigen presentation process and enrichment in cell growth and development gene networks suggesting prostate cancer may result from the accumulation of the effects of functional variants through multiple gene sets and pathways. The upstream analysis identifies critical transcription factors, which supplements the results regarding the regulatory role of the post-GWAS genes. We also identified the common genes between post-GWAS and three well-annotated prostate cancer Oncomine data in patient samples in order to uncover possible main genes in prostate cancer development/progression. Post-GWAS generated knowledge of gene networks and pathways, if analysed further and targeted appropriately, will have an important impact on clinical management of the disease.

Keywords: Prostate cancer, post-GWAS, functional variants, pathway analysis, upstream analysis, OncoPrint

Introduction:

Prostate cancer (PrCa) is the second leading cause of cancer death in the Western world [1]. Genetic and non-genetic (environmental) factors are known to promote PrCa with the prominent effect of genetics as demonstrated by 58% heritability that has been discovered by the large scale studies such as twin and familial segregation studies evaluating the role of genetic influence on PrCa development [2]. In particular, during the last decade, genome-wide association studies (GWAS) have successfully identified >160 loci associated with the risk of PrCa [3].

In general, GWAS identified loci represent a lead SNP (or so-called tag SNP, since they “tag” the disease locus), which is the primarily designed probe for that SNP in an array. A tag SNP represents multiple SNPs within a linkage disequilibrium (LD) block (around 10-100 kb) [4], consequently, identification of the exact functional variant is additionally complicated. Of note, the distant functional impact of some variants that are located away from their target genes on a different chromosome (*trans*-variants) could be misleading by local impacts of *cis*-variants on the same chromosome. For example, a genome-wide analysis of the risk SNPs demonstrated regulatory variant on chromosome 19 that impose a functional impact on potential target genes on chromosomes 10 (*MSMB*) and X (*NUDT11*) [5]. Additionally, the subtle direct/indirect effect of functional variants needs to be kept into consideration so that the most appropriate choice of an experiment is made. In particular, the majority of SNPs are located within non-coding regions mostly involved in regulating the gene expression by either regulating the chromatin remodelling, consequently changing chromatin interactions, and/or altering the binding of certain transcription factors (TFs). This is the main reason that many post-GWAS studies include the expression quantitative trait loci (eQTLs) as functional variants that represent the regulatory role of germline loci associated with gene expression levels. Validating the mechanism of action of these loci is challenging and difficult to guarantee using the present techniques. Other approaches such as allele-specific recruitment of TFs has been investigated by genotype-specific protein interactions such as Chromatin-immunoprecipitation sequencing (ChIP-seq) and DNase I hypersensitive sites sequencing

(DNase-seq) assays [6,7]. The electrophoretic mobility shift assay (EMSA) is another experimental approach to investigate the differential binding of TFs to a certain genomic region. The growing accessibility of Chromosome Conformation Capture (3C) method and its derivative including circular chromosome conformation capture (4C), chromosome conformation capture carbon copy (5C), Hi-C, Capture-C, high-throughput chromosome conformation capture (HiC-seq) and Capture-Hi-C and tethered conformation capture, Repli-seq or ChIA-PET techniques [8] have greatly help to identify long-distance DNA interactions of regulatory elements harbouring functional variants [9]. Genome editing methods such as Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)/Cas9 have been developed to examine the causality of potential variants in response elements within intergenic regions. While all the above-mentioned techniques can help post-GWAS analysis refining the role of PrCa-GWAS loci in prostate tumorigenesis, still, majority of signals are unknown and many other genes, in particular, non-coding RNAs (ncRNAs), in the vicinity of risk loci or distant yet to be discovered [10]. This brings up the urgent need for other approaches to implementing the GWAS and post-GWAS analysis in the improvement of clinical management of PrCa.

Pathway-based analysis of GWAS assigned genes (PWAS) has been an emerging tool used to define a group of genes that are involved in the same biological and/or molecular process in prostate tumorigenesis [11,12]. Notably, mapping GWAS assigned genes into gene networks [13] and molecular pathways [14] has the potential to reveal GWAS identified and/or novel genes within these pathways for the successful drug targeting [3]. Post-GWAS inference on the basis of related genes to GWAS signals has shown that this approach can be accounted for indicating biologically interpretable links to human diseases/traits including PrCa [14]. This has undeniable benefits to reveal putative targets to utilise post-GWAS hits aiming to drug repurposing for the identified dysregulated pathways that those genes act through [2,15]. For example, functional variants acting through the GWAS identified oncogene MYC [16] or androgen receptor (AR) [17] may drive PrCa, or through modulating known PrCa biomarkers such as MSMB or KLK3 [3], and/or other gene networks/pathways involved in cancer cell proliferation, invasion or metastasis. Therefore, pathways and regulatory networks produced by post-GWAS can pinpoint molecular pathways/gene networks that have functional roles in the development and progression of PrCa. In concordance with this

hypothesis, focusing on gene networks identified from the post-GWAS in PrCa can significantly increase the chances for clinical success and productivity for this highly polygenic cancer [18]. In other words, in contrast to GWAS alone, integrating post-GWAS data could add value for effective treatments for both known and newly identified pathways. For this purpose, we undertook a pathway post-GWAS analysis (PPWAS) including all the genes assigned to functional variants discovered so far (Figure 1, Supplementary Data S1) [10,19]. We used the Ingenuity Pathway Analysis (IPA) algorithm [20] and several publicly available databases of gene annotations such as Gene Ontology (GO) [21], REACTOME [12], Gene Set Enrichment Analysis (GSEA) [13] that also includes the Kyoto Encyclopedia of Genes and Genomes (KEGG) dataset [15]. The findings pinpoint several well-known pathways involved in PrCa as well as enrichment in cancer-related gene networks and upstream regulators. We further used Oncomine dataset to investigate the expression status of post-GWAS assigned genes in prostate primary and metastatic clinical samples [22]. The findings may provide us with clues to understand the possible mechanism differentiating molecular processes of prostate tumour development that implicate several key genes that are also affected by functional variants.

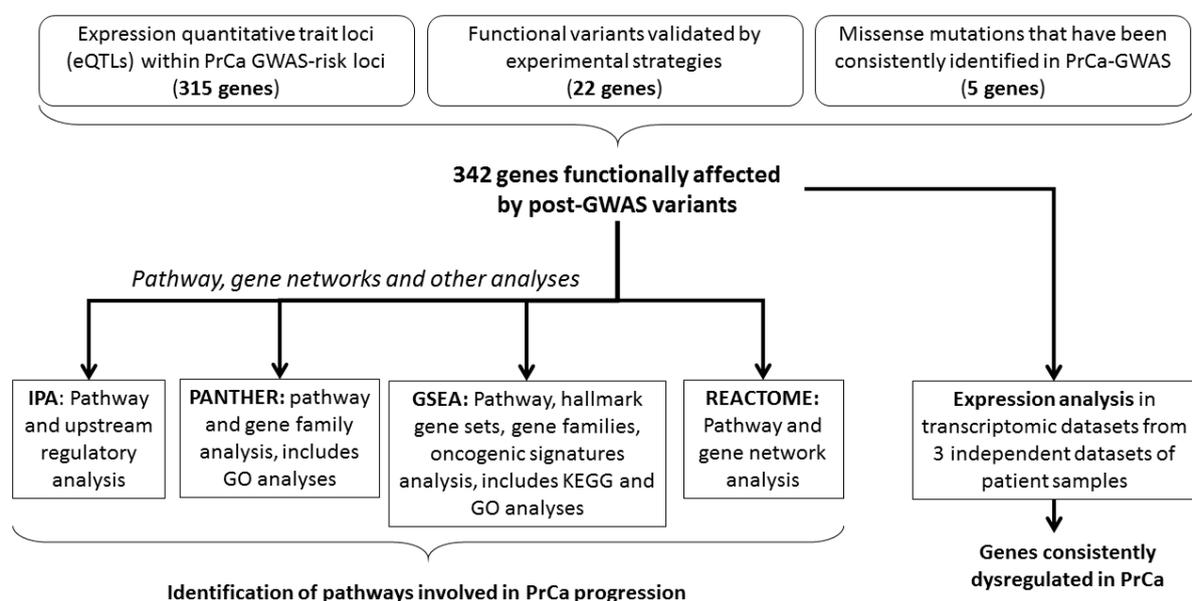


Figure 1. The study design. This flow chart depicts the flow of the analyses performed in this study. All the published available post-GWAS data was integrated in order to compile the assigned genes, 342 genes in total resulted from post-GWAS studies until 2019. Assigned

(target) genes represent functional variants within the target genes that are expression quantitative trait loci (eQTLs) or functional variants contributing to the molecular mechanisms by which contributing to the PrCa tumorigenesis evaluated by experiments or *in-silico* studies. Further investigations were performed to pathway analysis of contributed genes using four different algorithms. In addition, the investigation of the expression status of assigned genes in other studies reported in Oncomine dataset demonstrated a list of 8 overlapped dysregulated genes in those studies and post-GWAS genes.

Taken together, the results presented here suggest that PPWAS may help to prioritise the critical networks of genes involved in relevant pathways for further translational studies.

Material and methods:

Prostate cancer risk associated, functional SNPs and genes:

Apart from the lead GWAS-loci, we included individual SNPs in a set of variants within an LD block that have been identified by extended analysis (i.e. imputation and fine-mapping methods [23]) and further functional annotations. Additionally, we considered post-GWAS data carried out by dense genotyping and re-sequencing to identify independent PCA associated variants [24,25]. The eQTLs resulted from fine-mapping or transcriptome-wide association studies (TWAS) consist of the majority of the functional variants in PrCa post-GWAS studied here. TWAS approach investigates the association of transcriptome-wide gene expression with PrCa-risk to discover independent genes from a previously reported risk variant adding to the list of susceptibility genes. The current approach to further interpret the possible function of SNPs is to assign them to the nearby genes. Even though, nowadays we know that this approach is overlooking the active nature of the genome and has limitations (not in the scale of this study to discuss) but it has been the main criteria to assign the identified SNPs to the genes. The functional coding variants (synonymous or non-synonymous change) affect the structure, biochemical properties (e.g. charge) or the stability of the produced protein of a given assigned gene, thus subsequently change its molecular function [26]. All of the genes with coding variants identified through the GWAS/post-GWAS approach have been included in our pathway analysis. The prominent consequence of non-coding functional variants included in this study [27,28] is mainly due to: i) change in TF binding site

pattern [28,29], ii) change in DNA methylation marks [30] and iii) Chromatin architecture alteration [31], or a combination of these mechanisms.

Pathway Analysis

The PPWAS approach was applied to a list of 342 assigned genes to GWAS identified PrCa loci following a workflow presented in Figure 1. We used several publicly available and commercial tools to explore the relevant pathway/gene network enrichments of these 342 genes (Supplementary data S1) for at P-value <0.05 throughout all analyses described below.

Ingenuity pathway analysis (IPA): The IPA tool is based on the statistical significance of the relationship pattern of the molecules matched with the prior published biological knowledge [32]. IPA allows relating the most recent literature findings to a certain hypothesis that is the relation of the input genes in certain gene networks/pathways. It integrates detailed information about those genes and isoforms and creates interactive and customized pathways, accordingly. In addition, we investigated likely upstream regulators that are connected, directly or indirectly, to those genes. IPA scores those regulators based on their statistical significance by measuring the overlap of observed and predicted regulated gene sets as previously described [20] (Supplementary data S2). The pathway analysis was performed using IPA mapped 314 post-GWAS genes to the known IPA pathways (Supplementary data S2). 28 genes were not included in this analysis with the majority of long ncRNAs as unmapped genes. The upstream analysis feature was utilised provides insights into how TFs and chemicals affect biological processes related to the genes examined.

The PANTHER (Protein ANalysis THrough Evolutionary Relationships) [33] was used to investigate the function of genes (corresponding proteins) in homo sapiens, their ontology, involved pathways. The statistical enrichment test is based on differential gene expression levels or P-values from large-scale genomics experiments [33]. 53 genes were not mapped in this analysis, which is mainly long non-coding RNAs (Supplementary data S1). PANTHER-based pathway analysis was performed considering 55 gene identifiers (gene ids, gene symbols) that compare the functions of gene products from organisms across all kingdoms of life [34]. In PANTHER, pathway roles are assigned to the related individual gene ontologies (GO) using a variety of GO codes (Supplementary data S3) [21,35]. In addition, the PANTHER gene family analysis was conducted for 63 gene identifiers that were mapped for this analysis.

Gene Set Enrichment Analysis (GSEA) was used to determine whether a priori defined set of genes and our gene list of interest shows statistically significant, concordant alterations in gene expression associated with a disease that manifests at the level of biological pathways or co-regulated gene sets [36]. The functional enrichment analysis method by GSEA to identify classes of genes or proteins that are over-represented in a large set of genes or proteins, to derive reproducible associations with a biological state of interest (e.g. disease phenotypes) leading to identify pathways enriched in ranked gene lists.

This tool converted 342 submitted identifiers into 306 Entrez genes (Supplementary data S1) that are recognized for further annotation analysis of their significance in a pathway/network based on the Molecular Signatures Database (MSigDB) [37]. The genes underwent through the following analyses available in GSEA: “hallmark gene sets”, “canonical pathways”, “motif gene sets”, which includes microRNA (miRNA) targets and transcription factor targets. Moreover, a “computational gene set” analysis that considers cancer gene neighbourhoods and cancer modules were performed. We investigated “oncogenic signatures” resulted from the gene list as well as immunologic signatures. Furthermore, GSEA evaluates the overlap of the provided gene set based on KEGG and GO analyses with the estimation of the statistical significance of $FDR < 0.05$ (Supplementary data S4).

REACTOME is also a knowledge-based source of biomolecular pathways that contains data from other resources e.g. NCBI, Ensembl, UniProt, KEGG, ChEBI, PubMed and GO [12]. This analysis was performed using REACTOME version 67 including 282 identifiers that have been found or mapped to an equivalent element in REACTOME (Note: there are more than one identifiers for some of the genes, see Supplementary data S1). 118 identifiers were not found mapped to any entity in REACTOME. A binomial test was used to calculate the probability shown for each result, and the P-values are corrected for the multiple testing (Benjamini–Hochberg procedure) that arises from evaluating the submitted list of identifiers against every pathway. The P-values determine whether certain REACTOME pathways are over-represented (enriched) in the submitted data (Supplementary data S5).

Expression analysis of Post-GWAS identified genes in clinical samples

We utilised previously published Oncomine cancer microarray database (<http://www.oncomine.org>) to identify the expression of the shortlisted genes in patient samples.

OncoPrint datasets are derived from differential expression analyses that compared defined samples in groups of cancerous, normal and metastatic tissues or cell lines [22]. We utilised three studies with the highest number of samples for primary prostate carcinoma, normal and metastatic tumour site. The up- or down-expressed genes in each study were filtered for significantly differentially expressed in primary tumour versus normal/metastatic samples (P-value <0.05) by more than 1.5 fold change. The comparison of post-GWAS genes dysregulated genes was performed, separately, for primary tumour versus normal and metastatic samples investigating for overlapped genes based on Entrez gene identifiers. OncoPrint originated studies utilised in the study are: 1) Taylor dataset (OncoPrint ID: n9205) including 150 prostate carcinoma tissue specimens (131 specimens from primary and 19 metastatic tumours) and 29 paired normal adjacent prostate tissue specimens [38], 2) Yu (OncoPrint ID: n5345) that used 23 normal prostate and 64 prostate carcinoma samples [22] and 3) Grasso dataset (OncoPrint ID: n6252) that describes 59 localized prostate carcinoma and 28 benign prostate tissue specimens [39].

Results:

Analysis of compiled 342 genes assigned to PrCa-functional germline variants in this study highlights the value of integrating post-GWAS and provides testable hypotheses for future functional validation of PrCa aetiology. The genes that have been identified through experimental validation in addition to the variants that have a functional impact on the regulation of gene expression provides a powerful means to characterise the molecular mechanisms responsible for PrCa pathogenesis. Out of 342 genes, 191 genes have been identified as GWAS-eQTL pairs. We considered the contributions of these genes to investigate the gene expression-related functional form of variants (i.e. eQTLs) in possible pathway enrichments (Supplementary data S1). Out of 191 genes, 70 genes identified by TWAS which were included to this study in order to identify the possible same biological pathways enriched for the risk associated and dysregulated genes that may work in concert and lead to the development of PrCa.

The implicated canonical pathways and gene networks among the post-GWAS assigned genes were analysed including and excluding major histocompatibility complex (HLA) genes, respectively. The later analysis was performed to avoid the possible effect of the relatively

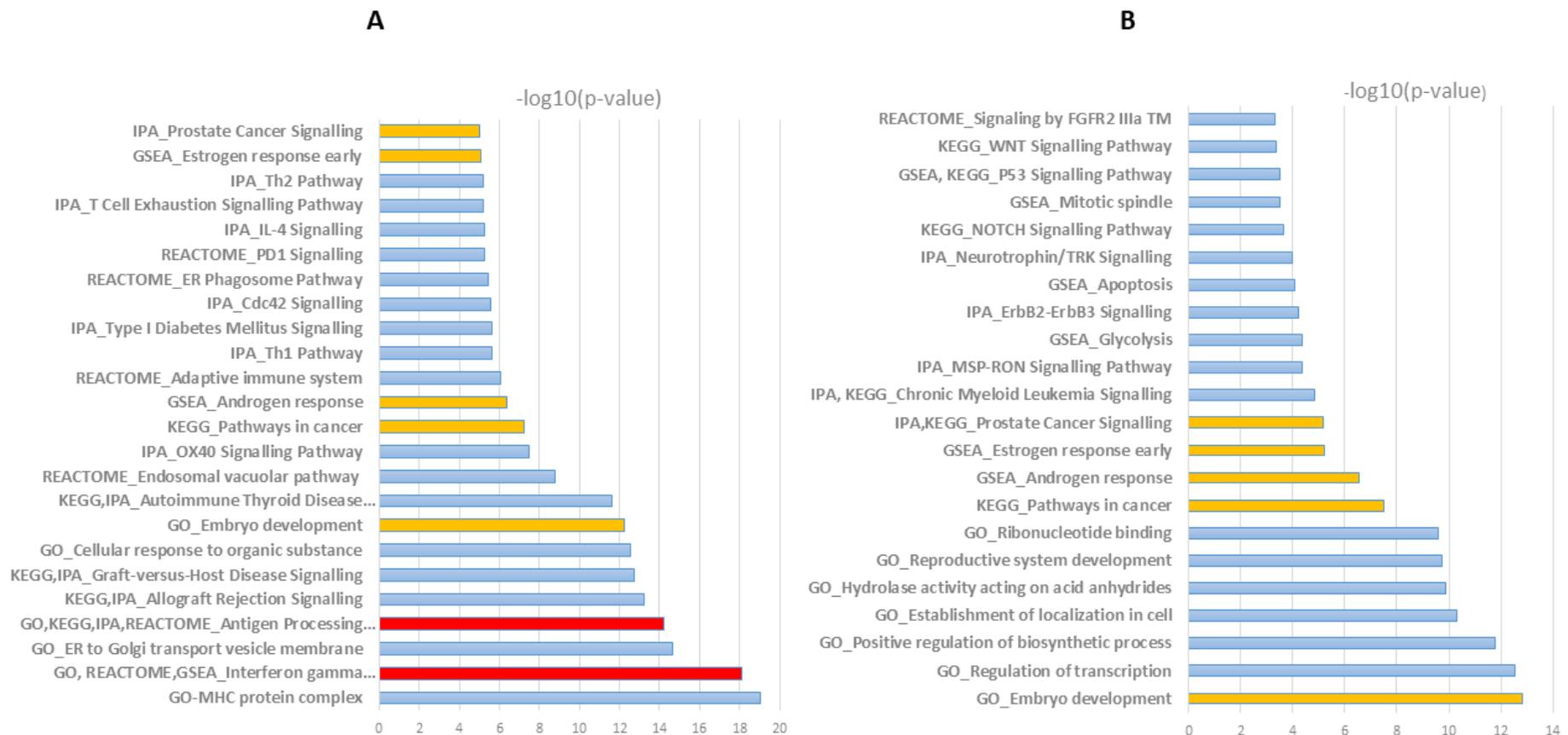
high number of HLA genes (consisting 5% of the total number of genes) on the results of this study. In total, more than 200 canonical pathways were significantly (P -value ≤ 0.05) enriched in assigned genes using four different tools in both analyses (including and excluding HLA genes). PANTHER pathways were similar in both analyses representing pathways such as p53 and Angiogenesis as the top-ranked pathways with the highest number of genes involved (Supplementary Data S3). One possible reason could be that the number of genes identified by PANTHER might have not been enough to generate significant P -values.

Pathway enriched including HLA genes. The most significant pathways identified by four tools were pathways related to immune signalling such as interferon-gamma mediated signalling (INFG) and antigen processing and presentation pathways (Figure 2A, Table 1). Additionally, this analysis revealed significant enrichments in components of the presentation of endogenous peptide antigen via Estrogen Receptor (ER) pathway and Allograft rejection (Figure 2A). ERs mediate physiological effects of estrogens play important roles via MHC molecules in growth, development, reproduction, and maintenance of a diverse range of mammalian tissues [22]. Including HLA genes, a significant enrichment for cancer-related gene networks was demonstrated (Supplementary data S2-6). In particular, cell growth and development processes such as cell cycle checkpoints and cell adhesion are identified as important gene sets (Table 1). The most significant gene sets are involved in extracellular matrix organization such as MHC protein complex (P -value: $9.72E-20$, Figure 2A). In addition, several observed categories involved in apoptosis, cell development, and cell-death were identified at high significance (Figure 2A). GSEA analysis revealed Androgen response as the most significant pathway involved (Supplementary data S4).

Table 1. The significant canonical pathways, gene sets and molecular functions that the genes are involved. In these analyses, we used the gene list described in Supplementary data S1.

Tool	Hallmark gene sets ©	The most significant gene network(s)	Canonical Pathway	(main) gene families	Function (Biological Process)	Cancer Module/ Oncogenic Signature	Top upstream regulators
IPA	Lipid metabolism, molecular transport, small molecule biochemistry	Tumorigenesis of tissue	Antigen presentation pathway Immune signalling pathways	Transcription factors (TFs)	Cancer, endocrine system disorders, organismal Injury and abnormalities	Cell death and survival Cellular Development	STAT5A NLRC5 TDP2
GSEA	Early response to estrogen, Response to interferon gamma	interferon-gamma response	Antigen presentation Pathway immune signalling pathways	TFs	Immune response	Cancer module 293 (see Supplementary data S4). All molecules involved in this module are HLAs except for B2M.	-
PANTHER (GO)	metallopeptidase activity	-Pachytene checkpoint protein 2 homolog -Prostate-specific antigen	p53 pathway Angiogenesis	MHC class I protein complex	cell cycle checkpoint;meiotic telophase ;negative regulation of meiotic nuclear division;reciprocal meiotic recombination;signal transduction	-	MAPK8-10 Protein kinase A JNK1/2 Junc PKCs
KEGG	Pathways in cancer	Allograft rejection	Allograft rejection	TFs	-	-	-
REACTOME	Genes involved in Adaptive Immune System	Endosomal/Vacuolar pathway	Endosomal/Vacuolar pathway	TFs	-	-	-

©The gene networks with the highest number of genes involved in.



0

1 **Figure 2. Overlapped significant canonical pathways and gene sets resulted from different tools used in this study.** The diagrams illustrate
 2 overlapped pathways/gene networks enriched in biological processes represented using four tools that have been shown as a prefix for each **(A)**

3 including HLA genes and **(B)** excluding HLA genes. The $-\log_{10}$ (P-value) has been used to create the graphs. The number of significant pathways
4 was higher including HLA genes, therefore we set the significance $-\log_{10}$ (P-value) cut-off at ≥ 5 for (A) and >3 for (B). A complete list of these
5 pathways and gene networks has been represented in Supplementary data S2-5. Orange bars are the common biological processes/canonical
6 pathways in both analysis, including and excluding HLA genes, and red bars represent significant biological processes/canonical pathways
7 demonstrated by more than two tools. The $-\log_{10}$ (P-value) presented in this graph is the highest (therefore, the lowest P-value and the most
8 significant) for the pathways or gene networks that are significant for more than one tool. The $-\log$ P-values are as followings: in **(A)** GO=18.08,
9 REACTOME=15.95, GSEA=5.05 for INFG pathway. GO=12.96, KEGG=12.04, IPA=11.8, REACTOME=15.95 for Antigen processing and presentation
10 pathways. KEGG=13.23 and IPA=7.7 for Allograft Rejection Signalling. KEGG=12.74 and IPA=7.9 for Graft versus Host Disease Signalling.
11 KEGG=11.65 and IPA=7.82 for Autoimmune Thyroid Disease Signalling. In **(B)** KEGG=4.64 and IPA=5 for Prostate Cancer Signalling. KEGG=3.98
12 and IPA=4.68 for Chronic Myeloid Leukemia Signalling. GSEA=3.00 and KEGG=3.32 for P53 Signalling Pathway. Note: PANTHER pathways are not
13 presented in these graphs due to the lack of P-values in output data. However, pathways such as p53 and Angiogenesis were the top rank
14 pathways with the highest number of genes involved Supplementary Data S3).

15 **Pathway enrichment of non-HLA genes.** Excluding HLA genes represents enrichments in additional cancer-related pathways such as p53
16 pathway, fibroblast growth factor receptor 2 (FGFR2) signalling related pathway and the well-established PrCa signalling pathway, AR (Figure
17 2B). Several molecular mechanisms enriched in cell growth such as cytokine activity and mitosis processes were observed (Table 2). These highly
18 significant pathways involved in androgen production, regulation, and receptor signalling, known to be involved in PrCa [40,41]. In addition, IPA
19 analysis demonstrated PrCa signalling pathway as the most significant canonical pathway ($7.73e-11$) with involving molecules such as *PIK3C2B*,
20 *KLK3*, *RALB*, *NKX3-1*, *FGFR2*, *CREB3L4*, *CDKN1B*, *MAP2K1* and *ATM* (Supplementary data S2).

- 21 **Table 2. The significant canonical pathways, gene sets and molecular functions that the genes are involved.** In these analyses, we used the
- 22 gene list described in Supplementary data S1 excluding HLA genes (HLA genes are listed in Supplementary Data S1).

Tool	The biological process with the highest number of genes involved in	Gene sets (the most significant)	Canonical Pathway	Gene families	Function (Biological Process)/ cellular component	Oncogenic Signature	Cancer Module	Hallmark gene sets*
IPA	Cell Morphology, Cell-To-Cell Signalling and Interaction, Cardiovascular System Development and Function	Cell Morphology, Cell-To-Cell signalling and Interaction, Cardiovascular System Development and Function	Prostate cancer signalling	transcription factors	Tumorigenesis of tissue	-	-	Cell Morphology, Cell-To-Cell signalling and Interaction, Cardiovascular System Development and Function,
GSEA	positive regulation of immune response	-	AR pathway	transcription factors	mitosis	Genes up-regulated during late stages of differentiation of embryoid bodies from V6.5 embryonic stem cells.	Genes in the cancer module 3 : http://robotics.stanford.edu/~erans/cancer/modules/module_3	Androgen response
PANTHER (GO)	Regulation of transcription from RNA polymerase II promoter	Regulation of transcription from RNA polymerase II promoter	p53 pathway	TGF-beta family	cytokine activity cytokine receptor binding signalling receptor activity response to hypoxia	-	-	-
KEGG	Pathways in cancer	Pathways in cancer	Pathways in cancer	TF	-	-	-	-
REACTOME	Cell Cycle	Extracellular matrix organization	Signalling by FGFR2 IIIa TM	TF	-	-	-	-

- 23 *The majority of gene networks are involved in these diseases.

Gene network and upstream regulatory analysis. Gene network analysis in IPA identified gene networks involved in different molecular and cellular functions including growth, metabolism and cancer-related processes (Supplementary data S2.2). Lipid metabolism, molecular transport and small molecule biochemistry was the top-networks considering all the genes while cardiovascular system development and function, cell morphology, and cell-to-cell signalling and interaction were the top-gene networks for non-HLA gene list. The interactions of the molecules involved in these top-gene networks have been illustrated in Figure 3A and Figure 3A. In addition, upstream regulation analysis in IPA revealed STAT5A as the most significant TF (in both analyses including/excluding HLA genes P-value= 7.7E-08 and 4.1E-08, respectively) (Figure 3C). The AR was also identified as an upstream regulator that can modulate the expression of TFs, biomarkers and vital tumour promoters in PrCa development such as *KLK2*, *KLK3*, *MYC*, *MSMB* and *TMPRSS2* [42]. This is in line with the fact that AR can activate other signalling cascades like the MAPK, Akt, JAK-STAT3 pathways [43,44] and stimulate growth processes in cells. GSEA regulatory analysis described as “motif gene sets” that is a comparative analysis of the human genome [13], also demonstrated TFs as the main gene family (Supplementary data S4). PANTHER algorithm analysis revealed MAPK8-10 as significant upstream regulators (Supplementary data S3). MAP kinases act as an integration point for multiple biochemical signals and are involved in a wide variety of cellular processes such as proliferation, differentiation, transcription regulation and development. For example, the main apoptotic cell-death mediated by JNK, and ERK pathways are regulated by a variety of MAPKs cascades in PrCa cells [37].

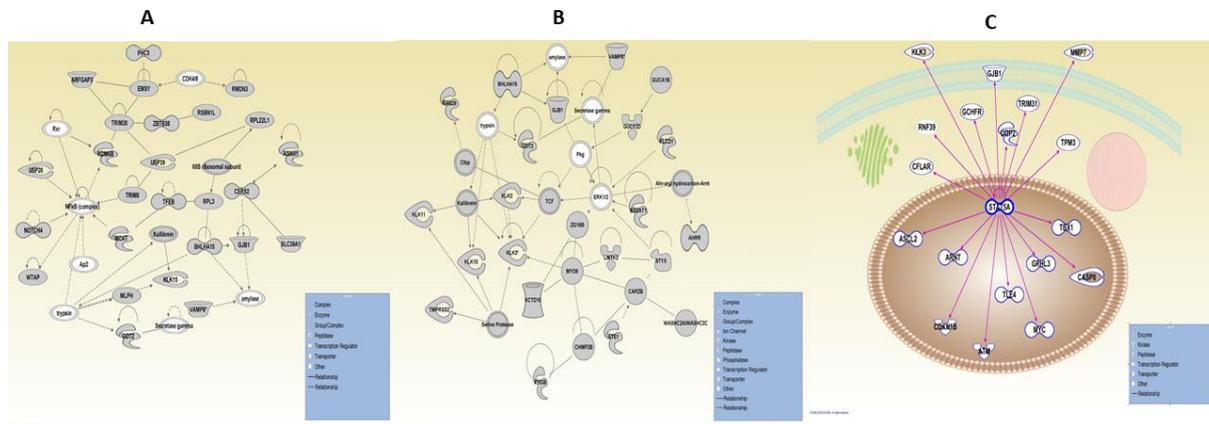


Figure 3. Ingenuity Pathway Analysis (IPA) analysis. A map of top gene network in IPA analysis with the highest number of involved genes **(A)** including HLA genes and **(B)** non-HLA genes: in this analysis, we excluded 14 HLA genes while performing IPA analysis to identify HLA independent key networks/pathways related to the assigned genes (see Supplementary data S1). **(C)** Upstream analysis of the gene lists in (A) and (B) by IPA demonstrated STAT5A as the most significant transcription factor regulating 18 genes, directly. The sub-cellular localisation of the molecules has been illustrated pinpointing a broad network communications of involved molecules in a cell. Arrows have depicted protein-protein interactions of the assigned genes. Arrows between nodes represent direct (solid arrows) and indirect (dashed arrows) interactions between molecules. The arrowheads depict an ‘act on’ relationship towards positive regulation. The blind-ended arrows represent the inhibitory interaction. Bidirectional arrowheads indicate reversible reactions. These networks are compact representations of literature-based knowledge about interactions. Each node represents a protein complex, and each interaction represents a significant number of genetic interactions. Node sizes are proportional to the number of proteins in the complex.

Expression signature of the GWAS identified genes in patient samples. To evaluate the likelihood that functional variants represent clinically related gene(s), we investigated any overlap between post-GWAS assigned genes and dysregulated genes in primary prostate tumour versus normal or metastasis patient samples utilising OncoPrint web-based dataset [22]. In total, there was a higher number of up- or down-expressed overlapped genes in metastatic patient samples compared to primary tumour samples (Figure 4). The Grasso study presented the highest overlapped genes in both comparisons (183 genes and 59 genes in metastasis and primary tumour, respectively). There were no overlapped genes when we compared the post-GWAS genes and dysregulated genes in primary tumour resulted from all three above-mentioned studies (Figure 4A) while the same investigation for metastatic samples identified 8 genes (Figure 4B). The identified genes are well-known genes [in particular, Kallikreins (*KLK2*, 3)] or targets for therapeutic approaches in PrCa such as prostate-specific antigen (PSA, encoded by *KLK3*) inhibitors [45]. Given the urgent need of identifying mechanisms that promote angiogenesis and cell proliferation during PrCa progression from the primary tumour to the bone which is the principal site of PrCa metastasis, focused studies on these genes are required to validate with experiments. For example, there have been studies investigating the potential role of fibroblast growth factor (FGF) in PrCa metastasis [46]. *FGFR2* is required for prostate organogenesis and the acquisition of androgen dependency for tissue homeostasis [32]. Interestingly, *FGFR2* signalling pathway was identified as a significant pathway in our study. Further investigations on findings in this study might help to link the genetic basis of PrCa into the molecular and cellular mechanisms involved and consequently gain insights to urgently needed alternative treatments for this complex disease.

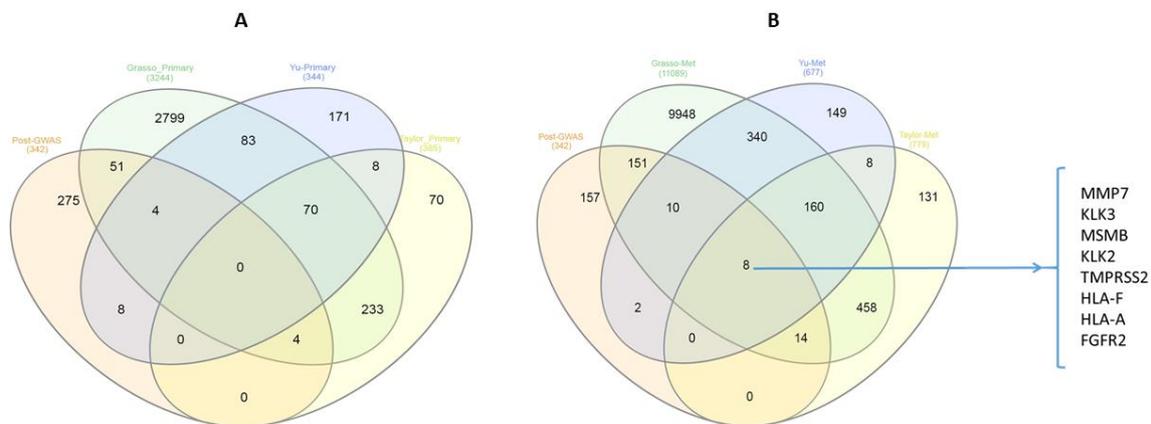


Figure 4. Overlapped genes between three OncoPrint datasets and assigned genes to post-GWAS functional variants in PrCa. (A) Venn diagram of the overlapped genes between post-GWAS assigned genes and primary tumour vs. normal samples **(B)** primary tumour vs. metastatic samples in Taylor, Grasso and Yu prostate datasets. The numbers in parentheses represent the numbers of significantly (P -value <0.05 by more than 1.5 fold change) differentially expressed genes by microarray. There are no overlapped genes in the first comparison while listed 8 genes in (B) are up- or down-expressed in all three studies in metastatic samples and also overlapped with the post-GWAS assigned genes that have been studied here.

Discussion:

In this study, we investigated networks and pathways highlighted by post-GWAS based target genes to demonstrate the biological and clinical relevance of functional variants in PrCa. With increasing success in cancer genomic data interpretation [10] and exponential improvements in post-GWAS studies [47], applying PPWAS may help to reveal the full spectrum of germline variants' role in prostate tumorigenesis. Our analyses confirmed the involvement of several known pathways in PrCa as well as pinpointed other less well-known pathways, which could be important for PrCa and represent novel therapeutic targets. The INFG and antigen processing and presentation pathways were significant biological processes/canonical

pathways demonstrated by four and three of the four tools, respectively (Figure 2). TFs are the main gene family contributing to the involving pathways (Supplementary Data S4.7 and S4.16). In total, there was a higher number of gene networks highlighting cancer-related pathways analysing non-HLA genes in comparison to the gene list including HLA genes (Figure 2A, B). Embryo development, PrCa signalling, pathways in cancer, androgen response and early estrogen response pathways were demonstrated as shared pathways when we included or excluded HLA genes.

With the urgent need for personalised care for PrCa patients, additional screening and treatment approaches can strikingly modify the diagnosis protocol for a better estimation of disease progression [48]. Of note, known pathways in addition to as-yet-unknown pathways may be leveraged to clarify the implication of various gene sets in PrCa and provide an idea for clinical development of pathway inhibitors [49]. For example, modifying HLA antigens which demonstrate frequent alteration in PrCa patients [50] have been suggested to improve the efficacy of immune responses against PrCa [51]. Notably, in this study, the antigen presentation and other immune response pathways were shown to be significantly enriched in PrCa. Similarly, IFNG was a significant pathway that belongs to the type II interferon family and is secreted by activated immune cells [4].

We integrated 70 newly reported genes from a recent prostate tissue transcriptome-wide association study with GWAS-PrCa (known as TWAS) [19]. GO pathway analysis of this study by Mancuso N. et al. on TWAS-PrCa that includes overlaps with previously reported genes represented positive regulation of chromatin binding (GO:0035563), nuclear membrane organization (GO:0071763) and chaperone-mediated protein folding requiring a cofactor (GO:0051085) as top biological processes [19]. These biological processes were identified in our GO analysis, however, other processes such as cell cycle checkpoint (GO:0000075), meiotic telophase I (GO:0007134), negative regulation of meiotic nuclear division (GO:0045835), reciprocal meiotic recombination (GO:0007131) and signal transduction (GO:0007165) were the top-ranked biological processes (Supplementary data S3, S4). This suggests that focusing on functional variants identified by post-GWAS could reveal additional biological information in PrCa. In another study, Schumacher F. et al. introduced a pathway-based approach only for 64 newly found GWAS risk loci including some functional variants

[52]. Their analysis detected the programmed cell death protein 1 (PD-1) signalling pathway as the most significant pathway in addition to the antigen processing, presentation and IFNG mediated signalling pathways [11]. Similarly, in our analysis IFNG pathway was shown as a significant pathway resulted from GSEA, REACTOME, KEGG and GO analyses and PD-1 signalling pathway is demonstrated as the fourth significant pathway in REACTOME analysis. The IFNG mediated signalling pathway rely on other signalling proteins like Janus-activated kinase 1 (JAK1), JAK2 and signal transducer and activator of transcription 1 (STAT-1) thus inducing signal transductions. In this way, assigned genes to post-GWAS loci might actively contribute to molecular and cellular biological processes leading to the overall outcome of prostate cell growth. In fact, their cumulative impacts on several main pathways may be involved in prostate tumorigenesis/progression [53]. Further exploration of discovered pathways can help greatly accelerate our ability to connect the genetic basis of PrCa to the clinic. For example, the AR signalling pathway was significant in GSEA analyses and the AR was demonstrated as one of the upstream regulators by IPA data. This is supporting evidence for the role of functional variants in modulating the genes regulated by AR or implicated in AR pathway which can be incorporated into currently used biomarkers related to the AR axis as the basis of stratifying the metastatic PrCa patients [16]. Moreover, differentially expressed genes in the primary tumour and metastatic samples in Oncomine dataset, which overlaps with the post-GWAS in our study highlights the potential drug targets for metastatic PrCa. Some of these genes have been found associated with progressive PrCa but also associated with nonaggressive PrCa [3]. For example, the *KLK3* gene region was found to be significantly associated with higher Gleason score [54] that is one of the indicators of the risk of progression or metastatic disease. Risk loci in close proximity to the *MMP7* gene contribute to PrCa invasiveness [42].

Identification of critical TFs in PrCa implies the central importance of upstream investigation of functional risk loci resulted from post-GWAS. *STAT5A*, the most significant TF revealed by IPA, regulates 18 of the genes identified by post-GWAS including some critical genes in PrCa such as *ATM*, *CDKN1B* and *ARNT*. *STAT5A/B* is critical for prostate cell survival and prostate tumour growth [11]. Therapeutic potential of STAT in cancer including PrCa is under investigations in several clinical trials targeting by STAT inhibitors [55].

Other TFs such as TDP2, SMAD7, SMARCD1, NLRC5 and HDAC1 that regulate well-known vital genes in PrCa including *KLK3*, *TMPRSS2*, *MYC*, *MMP7* are demonstrated in this study. Given the fact that TF binding formation is the key to reprogramming of genes [55], thus, disrupting TFs might be the potential relevance of a regulatory role of PrCa post-GWAS loci which suggests they are interesting targets to conduct follow-up studies [54]. The role of some of these molecules has been well-studied in PrCa [42], for example, NLRC5 is known to play a role in cytokine response via immune pathways and is suggested as a novel biomarker for cancer patient prognosis and survival [56]. High activity of HDACs has been reported to cause epigenetic alterations associated with malignant PrCa cell behaviour [46]. Consequently, a high rate of HDAC1 expression has been significantly associated with tumour dedifferentiation [47]. HDAC1 exerts an androgen-dependent regulatory effect on prostate cell proliferation and development that might be an interesting therapeutic target in PrCa to study and develop its inhibitors [57]. Interestingly, HDAC inhibitors have entered phase-2 clinical trials as new antineoplastic PrCa drug [3].

Altogether, we believe that utilising post-GWAS data is efficiently warranted and should be considered as the first approach highlighting several reasons: first and more importantly, post-GWAS enable us to focus on functionally involved genes in PrCa risk, therefore, will greatly speed up our ability to translate the functional part of the genome into the clinic. Secondly, reproducibility of GWAS is a valuable advantage emphasising on the high ability to reveal new discoveries from initial GWAS data than any other approach. Thirdly, subsequent dysregulation of the overlapped PrCa-driving molecules such as *MMP7*, *MSMB* and *KLK3* [10,19] based on the Oncomine dataset pinpoints likely promising identified gene networks and pathways to be further investigated in the clinic. However, interpretation of pathway-based analyses largely depends on the existing numerous algorithms that use various criteria to assign a gene to a network/pathway [58]. The observed differences in pathways analyses resulted from different tools may depend on the specific methodologies, and as such generalising the results must be done with caution. In particular, the main challenge in this type of analysis is that the observed outcomes depend on the number of input data (number of genes) mapped by different tools. This situation is common in the field of machine learning, yet finding a way to integrate all the genes in the analyses, in order to reduce bias in the final

model is necessary. For example, in our analysis, the number of recognised gene identifiers by various tools was different and it consequently resulted in a different interpretation. Of particular, this caveat vastly deters our understanding of the critical role of ncRNAs identified by post-GWAS in PrCa given the fact that they participate in a broad range of different mechanisms driving tumorigenesis [24,34,59,60]. The current methods mostly have been developed for integrating protein-coding genes overlooking ncRNAs' contribution to molecular pathways; thus, there is an urgent need to include these molecules in pathway analysis methods. Nevertheless, the least could be the utilisation of the results for deeper investigations to examine whether the post-GWAS identified pathways are promoting tumorigenesis.

Conclusion:

This study investigates the pathways by which many post-GWAS assigned genes may influence PrCa development. The results highlight the importance of complex interplay and crosstalk between different pathways, which will advance the current understanding of PrCa pathogenesis. Consequently, novel drug treatments could be developed for subsets of these identified pathways that are yet to be tested in improving risk stratification.

Abbreviations:

MSMB: Microseminoprotein Beta

NUDT11: Nudix Hydrolase 11

MYC: MYC Proto-Oncogene

AR: Androgen receptor

HLAs: histocompatibility complex genes

INFG: Interferon gamma

MHC: Major Histocompatibility Complex

FGFR2: Fibroblast Growth Factor Receptor 2

PIK3C2B: Phosphatidylinositol-4-Phosphate 3-Kinase Catalytic Subunit Type 2 Beta

KLK3/2: Kallikrein Related Peptidase 3/2

RALB: RAS Like Proto-Oncogene B

NKX3-1: NK3 Homeobox 1

CREB3L4: CAMP Responsive Element Binding Protein 3 Like 4

CDKN1B: Cyclin Dependent Kinase Inhibitor 1B

MAP2K1: Mitogen-Activated Protein Kinase Kinase 1

ATM: ATM Serine/Threonine Kinase

STAT5A/B: Signal Transducer And Activator Of Transcription 5A/B

TMPRSS2: Transmembrane Serine Protease 2

MAPK: Mitogen-activated protein kinase

Akt: AKT Serine/Threonine Kinase 1

JAK1/2: Janus Kinase 1/2

JNK: c-Jun N-terminal kinases

MMP7: Matrix Metalloproteinase 7

NLRC5: NLR Family CARD Domain Containing 5

HDAC1/2: Histone Deacetylase 1/2

Supplementary data:

Supplementary data S1: Prostate cancer functional variants within coding/non-coding regions reported by post-GWAS.

Supplementary data S2: Pathway analysis using the Ingenuity Pathway Analysis (IPA)

algorithm. We excluded the unmapped genes (listed here) in IPA for further analysis. Canonical pathways are compact representations of literature-based knowledge about regulatory interactions. List of implicated networks of affected/dysregulated genes by functional variants is presented. The 10 top canonical pathways in IPA analysis including HLA genes have been listed. We excluded HLA genes to identify HLA independent key pathways related to the assigned genes (see Supplementary data S1). The upstream regulators of the genes involved have been listed.

Supplementary data S3: Gene set analysis using PANTHER. In this analysis, the genes that did not map to Entrez IDs were eliminated. This data includes visualisation for the proportion of gene sets and involving molecules in presented pathways by PANTHER.

Supplementary data S4: Pathway analysis using (GSEA). The genes in overlaps have been shown in separate tables. By default, the result displays the 10 gene sets in the collection that best overlap with the gene set analysed here. P-value is calculated from the hypergeometric distribution for $(k-1, K, N - K, n)$. k is the number of genes in the intersection of the query set with a set from The Molecular Signatures Database (MSigDB, a collection of annotated gene sets for use with GSEA software). K is the number of genes in the set from MSigDB. N is the total number of gene universe (all known human gene symbols). n is the number of genes in the query set (the gene set analysed here).

Supplementary data S5: Pathway analysis using REACTOME including and excluding HLA genes.

References

1. Bell, K.J.; Del Mar, C.; Wright, G.; Dickinson, J.; Glasziou, P. Prevalence of incidental prostate cancer: A systematic review of autopsy studies. *International journal of cancer* **2015**, *137*, 1749-1757, doi:10.1002/ijc.29538.
2. Mucci, L.A.; Hjelmborg, J.B.; Harris, J.R.; Czene, K.; Havelick, D.J.; Scheike, T.; Graff, R.E.; Holst, K.; Moller, S.; Unger, R.H., et al. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *Jama* **2016**, *315*, 68-76, doi:10.1001/jama.2015.17703.
3. Benafif, S.; Kote-Jarai, Z.; Eeles, R.A.; Consortium, P. A Review of Prostate Cancer Genome-Wide Association Studies (GWAS). *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **2018**, *27*, 845-857, doi:10.1158/1055-9965.EPI-16-1046.
4. International HapMap, C. A haplotype map of the human genome. *Nature* **2005**, *437*, 1299-1320, doi:10.1038/nature04226.
5. Penney, K.L.; Sinnott, J.A.; Tyekucheva, S.; Gerke, T.; Shui, I.M.; Kraft, P.; Sesso, H.D.; Freedman, M.L.; Loda, M.; Mucci, L.A., et al. Association of prostate cancer risk variants with gene expression in normal and tumor tissue. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **2015**, *24*, 255-260, doi:10.1158/1055-9965.EPI-14-0694-T.
6. Song, L.; Crawford, G.E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc* **2010**, *2010*, pdb prot5384, doi:10.1101/pdb.prot5384.
7. Furey, T.S. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet* **2012**, *13*, 840-852, doi:10.1038/nrg3306.
8. Smith, A.J.P.; Deloukas, P.; Munroe, P.B. Emerging applications of genome-editing technology to examine functionality of GWAS-associated variants for complex traits. *Physiol Genomics* **2018**, *50*, 510-522, doi:10.1152/physiolgenomics.00028.2018.
9. Luo, Z.; Rhie, S.K.; Lay, F.D.; Farnham, P.J. A Prostate Cancer Risk Element Functions as a Repressive Loop that Regulates HOXA13. *Cell Rep* **2017**, *21*, 1411-1417, doi:10.1016/j.celrep.2017.10.048.

10. Farashi, S.; Kryza, T.; Clements, J.; Batra, J. Post-GWAS in prostate cancer: from genetic association to biological contribution. *Nature reviews. Cancer* **2019**, *19*, 46-59, doi:10.1038/s41568-018-0087-3.
11. Dagvadorj, A.; Kirken, R.A.; Leiby, B.; Karras, J.; Nevalainen, M.T. Transcription factor signal transducer and activator of transcription 5 promotes growth of human prostate cancer cells in vivo. *Clinical cancer research : an official journal of the American Association for Cancer Research* **2008**, *14*, 1317-1324, doi:10.1158/1078-0432.CCR-07-2024.
12. Fabregat, A.; Sidiropoulos, K.; Viteri, G.; Forner, O.; Marin-Garcia, P.; Arnau, V.; D'Eustachio, P.; Stein, L.; Hermjakob, H. Reactome pathway analysis: a high-performance in-memory approach. *BMC bioinformatics* **2017**, *18*, 142, doi:10.1186/s12859-017-1559-2.
13. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S., et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 15545-15550, doi:10.1073/pnas.0506580102.
14. Hindorff, L.A.; Sethupathy, P.; Junkins, H.A.; Ramos, E.M.; Mehta, J.P.; Collins, F.S.; Manolio, T.A. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **2009**, *106*, 9362-9367, doi:10.1073/pnas.0903103106.
15. Kanehisa, M.; Sato, Y.; Furumichi, M.; Morishima, K.; Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic acids research* **2019**, *47*, D590-D595, doi:10.1093/nar/gky962.
16. Wasserman, N.F.; Aneas, I.; Nobrega, M.A. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. *Genome research* **2010**, *20*, 1191-1197, doi:10.1101/gr.105361.110.
17. Pomerantz, M.M.; Li, F.; Takeda, D.Y.; Lenci, R.; Chonkar, A.; Chabot, M.; Cejas, P.; Vazquez, F.; Cook, J.; Shivdasani, R.A., et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nature genetics* **2015**, *47*, 1346-1351, doi:10.1038/ng.3419.
18. Cowen, L.; Ideker, T.; Raphael, B.J.; Sharan, R. Network propagation: a universal amplifier of genetic associations. *Nat Rev Genet* **2017**, *18*, 551-562, doi:10.1038/nrg.2017.38.
19. Mancuso, N.; Gayther, S.; Gusev, A.; Zheng, W.; Penney, K.L.; Kote-Jarai, Z.; Eeles, R.; Freedman, M.; Haiman, C.; Pasaniuc, B., et al. Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. *Nature communications* **2018**, *9*, 4079, doi:10.1038/s41467-018-06302-1.
20. Kramer, A.; Green, J.; Pollard, J., Jr.; Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **2014**, *30*, 523-530, doi:10.1093/bioinformatics/btt703.
21. Mi, H.; Muruganujan, A.; Thomas, P.D. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic acids research* **2013**, *41*, D377-386, doi:10.1093/nar/gks1118.
22. Yu, Y.P.; Landsittel, D.; Jing, L.; Nelson, J.; Ren, B.; Liu, L.; McDonald, C.; Thomas, R.; Dhir, R.; Finkelstein, S., et al. Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **2004**, *22*, 2790-2799, doi:10.1200/JCO.2004.05.158.
23. Sud, A.; Kinnersley, B.; Houlston, R.S. Genome-wide association studies of cancer: current insights and future perspectives. *Nature reviews. Cancer* **2017**, *17*, 692-704, doi:10.1038/nrc.2017.82.
24. Stegeman, S.; Amankwah, E.; Klein, K.; O'Mara, T.A.; Kim, D.; Lin, H.Y.; Permut-Wey, J.; Sellers, T.A.; Srinivasan, S.; Eeles, R., et al. A Large-Scale Analysis of Genetic Variants within

- Putative miRNA Binding Sites in Prostate Cancer. *Cancer discovery* **2015**, *5*, 368-379, doi:10.1158/2159-8290.CD-14-1057.
25. Kichaev, G.; Yang, W.Y.; Lindstrom, S.; Hormozdiari, F.; Eskin, E.; Price, A.L.; Kraft, P.; Pasaniuc, B. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet* **2014**, *10*, e1004722, doi:10.1371/journal.pgen.1004722.
 26. Kote-Jarai, Z.; Amin Al Olama, A.; Leongamornlert, D.; Tymrakiewicz, M.; Saunders, E.; Guy, M.; Giles, G.G.; Severi, G.; Southey, M.; Hopper, J.L., et al. Identification of a novel prostate cancer susceptibility variant in the KLK3 gene transcript. *Hum Genet* **2011**, *129*, 687-694, doi:10.1007/s00439-011-0981-1.
 27. Lu, Y.; Zhang, Z.; Yu, H.; Zheng, S.L.; Isaacs, W.B.; Xu, J.; Sun, J. Functional annotation of risk loci identified through genome-wide association studies for prostate cancer. *Prostate* **2011**, *71*, 955-963, doi:10.1002/pros.21311.
 28. Whittington, T.; Gao, P.; Song, W.; Ross-Adams, H.; Lamb, A.D.; Yang, Y.; Svezia, I.; Klevebring, D.; Mills, I.G.; Karlsson, R., et al. Gene regulatory mechanisms underpinning prostate cancer susceptibility. *Nat Genet* **2016**, *48*, 387-397, doi:10.1038/ng.3523.
 29. Jin, H.J.; Jung, S.; DebRoy, A.R.; Davuluri, R.V. Identification and validation of regulatory SNPs that modulate transcription factor chromatin binding and gene expression in prostate cancer. *Oncotarget* **2016**, *7*, 54616-54626, doi:10.18632/oncotarget.10520.
 30. Chang, B.L.; Zheng, S.L.; Isaacs, S.D.; Wiley, K.E.; Turner, A.; Li, G.; Walsh, P.C.; Meyers, D.A.; Isaacs, W.B.; Xu, J. A polymorphism in the CDKN1B gene is associated with increased risk of hereditary prostate cancer. *Cancer Res* **2004**, *64*, 1997-1999.
 31. Guo, H.; Ahmed, M.; Zhang, F.; Yao, C.Q.; Li, S.; Liang, Y.; Hua, J.; Soares, F.; Sun, Y.; Langstein, J., et al. Modulation of long noncoding RNAs by risk SNPs underlying genetic predispositions to prostate cancer. *Nat Genet* **2016**, *48*, 1142-1150, doi:10.1038/ng.3637.
 32. Lin, Y.; Liu, G.; Zhang, Y.; Hu, Y.P.; Yu, K.; Lin, C.; McKeehan, K.; Xuan, J.W.; Ornitz, D.M.; Shen, M.M., et al. Fibroblast growth factor receptor 2 tyrosine kinase is required for prostatic morphogenesis and the acquisition of strict androgen dependency for adult tissue homeostasis. *Development* **2007**, *134*, 723-734, doi:10.1242/dev.02765.
 33. Mi, H.; Huang, X.; Muruganujan, A.; Tang, H.; Mills, C.; Kang, D.; Thomas, P.D. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic acids research* **2017**, *45*, D183-D189, doi:10.1093/nar/gkw1138.
 34. Anastasiadou, E.; Jacob, L.S.; Slack, F.J. Non-coding RNA networks in cancer. *Nature reviews. Cancer* **2018**, *18*, 5-18, doi:10.1038/nrc.2017.99.
 35. Thomas, P.D.; Campbell, M.J.; Kejariwal, A.; Mi, H.; Karlak, B.; Daverman, R.; Diemer, K.; Muruganujan, A.; Narechania, A. PANTHER: a library of protein families and subfamilies indexed by function. *Genome research* **2003**, *13*, 2129-2141, doi:10.1101/gr.772403.
 36. Holden, M.; Deng, S.; Wojnowski, L.; Kulle, B. GSEA-SNP: applying gene set enrichment analysis to SNP data from genome-wide association studies. *Bioinformatics* **2008**, *24*, 2784-2785, doi:10.1093/bioinformatics/btn516.
 37. Liberzon, A.; Birger, C.; Thorvaldsdottir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems* **2015**, *1*, 417-425, doi:10.1016/j.cels.2015.12.004.
 38. Taylor, B.S.; Schultz, N.; Hieronymus, H.; Gopalan, A.; Xiao, Y.; Carver, B.S.; Arora, V.K.; Kaushik, P.; Cerami, E.; Reva, B., et al. Integrative genomic profiling of human prostate cancer. *Cancer cell* **2010**, *18*, 11-22, doi:10.1016/j.ccr.2010.05.026.
 39. Stepien, E.L.; Costa, M.C.; Enguita, F.J. miRNAtools: Advanced Training Using the miRNA Web of Knowledge. *Non-coding RNA* **2018**, *4*, doi:10.3390/ncrna4010005.
 40. Hazelett, D.J.; Rhie, S.K.; Gaddis, M.; Yan, C.; Lakeland, D.L.; Coetzee, S.G.; Henderson, B.E.; Noushmehr, H.; Cozen, W.; Kote-Jarai, Z., et al. Comprehensive Functional Annotation of 77

- Prostate Cancer Risk Loci. *PLoS Genet* **2014**, *10*, e1004102, doi:10.1371/journal.pgen.1004102.
41. Gusev, A.; Shi, H.; Kichaev, G.; Pomerantz, M.; Li, F.; Long, H.W.; Ingles, S.A.; Kittles, R.A.; Strom, S.S.; Rybicki, B.A., et al. Atlas of prostate cancer heritability in European and African-American men pinpoints tissue-specific regulation. **2016**, *7*, 10979, doi:10.1038/ncomms10979
- <http://dharmastra.live.cf.private.springer.com/articles/ncomms10979#supplementary-information>.
42. Cooper, C.S.; Clark, J.; Brewer, D.S.; Edwards, D.R. Prostate Single Nucleotide Polymorphism Provides a Crucial Clue to Cancer Aggression in Active Surveillance Patients. *European urology* **2016**, *69*, 229-230, doi:10.1016/j.eururo.2015.09.038.
 43. Ghosh, P.M.; Malik, S.N.; Bedolla, R.G.; Wang, Y.; Mikhailova, M.; Prihoda, T.J.; Troyer, D.A.; Kreisberg, J.I. Signal transduction pathways in androgen-dependent and -independent prostate cancer cell proliferation. *Endocr Relat Cancer* **2005**, *12*, 119-134, doi:10.1677/erc.1.00835.
 44. Shtivelman, E.; Beer, T.M.; Evans, C.P. Molecular pathways and targets in prostate cancer. *Oncotarget* **2014**, *5*, 7217-7259, doi:10.18632/oncotarget.2406.
 45. Sotiropoulou, G.; Pampalakis, G. Targeting the kallikrein-related peptidases for drug development. *Trends Pharmacol Sci* **2012**, *33*, 623-634, doi:10.1016/j.tips.2012.09.005.
 46. Bubendorf, L.; Schopfer, A.; Wagner, U.; Sauter, G.; Moch, H.; Willi, N.; Gasser, T.C.; Mihatsch, M.J. Metastatic patterns of prostate cancer: an autopsy study of 1,589 patients. *Human pathology* **2000**, *31*, 578-583, doi:10.1053/hp.2000.6698.
 47. Gallagher, M.D.; Chen-Plotkin, A.S. The Post-GWAS Era: From Association to Function. *American journal of human genetics* **2018**, *102*, 717-730, doi:10.1016/j.ajhg.2018.04.002.
 48. Walsh, P.C. The Search for the Missing Heritability of Prostate Cancer. *Eur Urol* **2017**, *10.1016/j.eururo.2017.04.003*, doi:10.1016/j.eururo.2017.04.003.
 49. Califano, A.; Butte, A.J.; Friend, S.; Ideker, T.; Schadt, E. Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nat Genet* **2012**, *44*, 841-847, doi:10.1038/ng.2355.
 50. Carretero, F.J.; Del Campo, A.B.; Flores-Martin, J.F.; Mendez, R.; Garcia-Lopez, C.; Cozar, J.M.; Adams, V.; Ward, S.; Cabrera, T.; Ruiz-Cabello, F., et al. Frequent HLA class I alterations in human prostate cancer: molecular mechanisms and clinical relevance. *Cancer Immunol Immunother* **2016**, *65*, 47-59, doi:10.1007/s00262-015-1774-5.
 51. Doonan, B.P.; Haque, A. Prostate Cancer Immunotherapy: Exploiting the HLA Class II Pathway in Vaccine Design. *J Clin Cell Immunol* **2015**, *6*, doi:10.4172/2155-9899.1000351.
 52. Schumacher, F.R.; Al Olama, A.A.; Berndt, S.I.; Benlloch, S.; Ahmed, M.; Saunders, E.J.; Dadaev, T.; Leongamornlert, D.; Anokian, E.; Cieza-Borrella, C., et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nature genetics* **2018**, *50*, 928-936, doi:10.1038/s41588-018-0142-8.
 53. Boyle, E.A.; Li, Y.I.; Pritchard, J.K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **2017**, *169*, 1177-1186, doi:10.1016/j.cell.2017.05.038.
 54. Shu, X.; Ye, Y.; Gu, J.; He, Y.; Davis, J.W.; Thompson, T.C.; Logothetis, C.J.; Kim, J.; Wu, X. Genetic variants of the Wnt signaling pathway as predictors of aggressive disease and reclassification in men with early stage prostate cancer on active surveillance. *Carcinogenesis* **2016**, *37*, 965-971, doi:10.1093/carcin/bgw082.
 55. Furqan, M.; Akinleye, A.; Mukhi, N.; Mittal, V.; Chen, Y.; Liu, D. STAT inhibitors for cancer therapy. *Journal of hematology & oncology* **2013**, *6*, 90, doi:10.1186/1756-8722-6-90.
 56. Yoshihama, S.; Vijayan, S.; Sidiq, T.; Kobayashi, K.S. NLRC5/CITA: A Key Player in Cancer Immune Surveillance. *Trends Cancer* **2017**, *3*, 28-38, doi:10.1016/j.trecan.2016.12.003.

57. Wang, X.; Xu, J.; Wang, H.; Wu, L.; Yuan, W.; Du, J.; Cai, S. Trichostatin A, a histone deacetylase inhibitor, reverses epithelial-mesenchymal transition in colorectal cancer SW480 and prostate cancer PC3 cells. *Biochem Biophys Res Commun* **2015**, *456*, 320-326, doi:10.1016/j.bbrc.2014.11.079.
58. Huang, Q.; Whittington, T.; Gao, P.; Lindberg, J.F.; Yang, Y.; Sun, J.; Vaisanen, M.R.; Szulkin, R.; Annala, M.; Yan, J., et al. A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. *Nat Genet* **2014**, *46*, 126-135, doi:10.1038/ng.2862.
59. Matin, F.; Jeet, V.; Srinivasan, S.; Cristino, A.S.; Panchadsaram, J.; Clements, J.A.; Batra, J.; Australian Prostate Cancer, B. MicroRNA-3162-5p-Mediated Crosstalk between Kallikrein Family Members Including Prostate-Specific Antigen in Prostate Cancer. *Clinical chemistry* **2019**, *65*, 771-780, doi:10.1373/clinchem.2018.295824.
60. Stegeman, S.; Moya, L.; Selth, L.A.; Spurdle, A.B.; Clements, J.A.; Batra, J. A genetic variant of MDM4 influences regulation by multiple microRNAs in prostate cancer. *Endocrine-related cancer* **2015**, *22*, 265-276, doi:10.1530/ERC-15-0013.

Acknowledgement:

The authors are grateful for a Queensland University of Technology Postgraduate Award (QUTPRA), and Australian National Health and Medical Research Council (NHMRC) Career Development Fellowship (CDF).

Author Contributions Statement:

S.F. researched data for the article, substantially contributed to discussion of content, wrote the article and reviewed and/or edited it before submission. T.K. substantially contributed to discussion of content and reviewed and/or edited the article before submission. J.B. substantially contributed to the discussion of content and reviewed and/or edited the article before submission.

Competing interests:

The authors declare no competing interests.