A Review of Prostate Cancer Genome Wide Association Studies (GWAS)

Sarah Benafif[1], Zsofia Kote-Jarai[1], and Rosalind A. Eeles[1,2]; on behalf of the PRACTICAL Consortium.

1. The Institute of Cancer Research, 15 Cotswold Road, Sutton, SM2 5NG, UK
2. The Royal Marsden Hospital, Fulham Road, London, SW3 6JJ, UK

Corresponding Author: Sarah Benafif; sarah.benafif@icr.ac.uk

Running title: A review of PrCa GWAS

Keywords: Genetics; single nucleotide polymorphism; personalised medicine; prostate cancer; genotype-phenotype correlations

A review of prostate cancer GWAS

Abstract

Prostate cancer (PrCa) is the commonest cancer in men in Europe and the USA. The genetic heritability of PrCa is contributed to by both rarely occurring genetic variants with higher penetrance and moderate to commonly occurring variants conferring lower risks. The number of identified variants belonging to the latter category has increased dramatically in the last 10 years with the development of the genome wide association study (GWAS) and the collaboration of international consortia that have led to the sharing of large-scale genotyping data. Over 40 PrCa GWAS have been reported, with approximately 170 common variants now identified.

Clinical utility of these variants could include strategies for population-based risk stratification to target PrCa screening to men with an increased genetic risk of disease development, while for those who develop PrCa, identifying genetic variants could allow treatment to be tailored based on a genetic profile in the early disease setting. Functional studies of identified variants are needed to fully understand underlying mechanisms of disease and identify novel targets for treatment. This review will outline the GWAS carried out in PrCa and the common variants identified so far, and how these may be utilised clinically in the screening for and management of PrCa.

Introduction

Prostate cancer (PrCa) is the second commonest cancer in men worldwide and the commonest cancer in men in the USA and in Europe (417,000 cases in Europe in 2012).[1, 2] The aetiology of PrCa is multi-factorial and although not yet fully understood, geographical and ethnic variations suggest genetic factors have a role as well as environmental and lifestyle factors. So far, the most well established risk factors are age, family history and ethnicity. These risk factors as well as evidence of the heritability of PrCa from epidemiological as well as twin studies point towards a significant genetic component contributing to PrCa development. The recent Nordic Twin Study of Cancer estimated the variation of PrCa risk attributed to genetic factors to be 57% making it one of the most heritable cancers.[2]

As with other complex diseases, the genetic heritability of PrCa is contributed to by both rarely occurring but higher penetrant genetic variants and moderate to commonly occurring variants conferring lower risks. Current research on PrCa susceptibility variants can explain 34.4% of the familial relative risk (FRR) of PrCa, with approximately 6% accounted for by rarely occurring variants and 28.4% attributed to more commonly occurring (Minor allele frequency (MAF) >1%) SNPs as well as some rarer single nucleotide variants. In the largest PrCa genome wide association study (GWAS) and meta-analysis[3] reported recently, 63 novel PrCa susceptibility loci were identified bringing the total number of known loci to 167 (Table 1, Figure 1).

Evolution of research into prostate cancer heritability

The initial research into the genetic predisposition to PrCa focussed on linkage studies in PrCa families. Genomic regions and candidate genes associated with risk were identified but very few were able to be replicated in subsequent studies. The few significant findings from such linkage studies was the identification of 8q24 as a significant PrCa risk region [4] and the identification of the missense G84E variant in *HOXB13*.[5] The latter mutation has mainly been reported in populations of European descent and carrier frequencies in men with PrCa range between 0.79% (Canada) to 2.99% (Northern Europe), conferring an odds ratio (OR) of 3.6-8.6. [6] In a large international sample of PrCa families recruited from the ICPCG (International Consortium of Prostate Cancer Genetics) consortium, 5% of families carried the G84E variant, with the highest rate observed in families from Finland (22%), followed by Sweden (8.2%). [7] The lack of definitive findings from family based linkage studies and the candidate gene approach reflects the polygenic nature of PrCa susceptibility. Although rare and moderate to highly penetrant risk variants other than *HOXB13* are known, the majority of familial risk remains unexplained. Even within PrCa families known to carry a rare variant such as *HOXB13* or *BRCA2*, there will be men who develop disease in the absence of the germline variant in that family. These cases may be attributed to more commonly occurring genetic variants conferring lower risks of PrCa development.

The development of gene association studies and subsequent modern GWAS has allowed researchers to overcome the limitations of small samples sizes and restricted number of variants investigated in previous linkage mapping studies. The first GWAS reported in 2006 was carried out in PrCa cases and controls and identified SNPs in chromosome 8q24- in fact, 3 separate studies reported around the same time identified and replicated previous association results in this region. With the collaboration of large international consortia recent GWAS have included tens of thousands of cases and controls and greatly increased the power to detect cancer risk variants, including some rarer variants. Since the initial GWAS in 2006,[8] 44 PrCa GWAS have been reported (Listed in NHGRI-EBI Catalogue of published GWASs: http://www.ebi.ac.uk/gwas). The advances in DNA sequencing technology and the ability to genotype large numbers of DNA samples concurrently on custom designed micro-arrays has led to several large PrCa GWAS, some of which have combined their data in meta-analyses and in turn have identified approximately 170 PrCa susceptibility variants.

Variants identified by GWAS

One of the earliest GWAS identified SNPs lying on 8q24; subsequent studies have revealed further SNPs in this region. The biological mechanism leading to prostate carcinogenesis though is unclear as the risk loci lie in non-coding regions of DNA. The nearest gene to this region is *MYC*, a proto-oncogene disrupted in many cancers. Functional studies including chromatin conformation assays, such as 3C, have shown long-range chromatin interactions of the 8q24 SNPs and these are thought to influence the expression of genes such as *MYC*. Other data from 3C experiments incorporating multi-target sequencing identified both intra- and inter-chromosomal interactions of 8q24 loci.[9] An example of an inter-chromosomal interacting gene is *CD96* on chromosome 3q13 with multiple interacting loci or 'hot-spots' in both chromosomal regions. In the same study,[9] 8q24 interactions with genes involved in the Wnt signalling pathway were also observed, suggesting that some risk SNPs have effects on the expression of multiple genes and may influence several cell signalling

pathways. Identifying such interactions will allow better understanding of the biological mechanisms leading to PrCa.

In the examples mentioned above, *CD96* encodes a membrane protein that is thought to play a role in the interactions of activated T and NK cells during the late phase of the immune response. It may also have a role in antigen presentation.[10] Based on the Biograph platform (an automated hypothesis generating platform utilising data integration and data mining to explore biomedical information), CD96 could have a role in PrCa development through its interactions with pathways linked to known cancer associated mi-RNAs such as MIR-127 (BIOGRAPH).[11] The Wnt signalling pathway which was also found to interact with 8q24 loci is involved in prostate bud growth and luminal epithelial differentiation providing a plausible mechanism linked to PrCa development. Other studies have also linked the risk allele of an 8q24 SNP (rs6983267) to enhanced Wnt signalling and other genomic regions harbouring PrCa risk SNPs show significant enrichment of Wnt signalling genes.[9]

As the number of cases and controls included in modern GWAS has increased, so has the yield of new 'hits' from each study. In a study reported by Eeles et al in 2008, a two stage GWAS identified 7 novel PrCa associated variants in addition to confirming previously identified loci. This study included nearly 2000 cases and 2000 controls in stage one, followed by over 3000 cases and 3000 controls in stage two. Some of the SNPs identified in that study were linked to candidate genes that could be involved in PrCa: *MSMB, LMTK2* and *KLK3*.

A later GWAS included a 5-fold higher number of cases and controls to genotype 211,155 SNPs on a custom SNP array, the iCOGS chip. Through the collaborations set up within the PRACTICAL (Prostate Cancer Association group To Investigate Cancer Associated Alterations in the genome) consortium (http://practical.icr.ac.uk; Figure 2), samples from 32 studies (25,074 PrCa cases and 24,272 controls) were included in this GWAS and identified 23 new PrCa susceptibility loci. Although this was the first study to identify susceptibility variants associated with aggressive disease, the 16 SNPs that fit this category also associated with non-aggressive PrCa, and therefore would not be clinically useful in identifying significant PrCa cases, for example those requiring radical treatment rather than non-interventional approaches.

Utilising the iCOGS chip, fine mapping of the *HOXB* region at 17q21 was carried out to analyse the association of common variants of the *HOXB* genes with PrCa, and to investigate the relationship of common variants with the rare but highly penetrant G84E *HOXB13* mutation.[12] By genotyping 700 SNPs in this region, imputing a further 3195 SNPs in cases and controls (approximately 20,000 each), and carrying out haplotype analyses, a common variant was found to tag and partially correlate with the rare G84E variant. This phenomenon of synthetic association may explain some of the missing heritability of PrCa as risk estimates attached to common variants may be significantly underestimated in cases where a GWAS signal is underpinned by a rare causative variant. The investigation of such synthetic associations is becoming more feasible with the increasing access to genomic data from different populations. Rare predisposition variants involved in synthetic association are likely to be limited to certain ethnicities and therefore the lack of signal in a multi-ethnic study

for a tag SNP could suggest that investigation of that locus by sequencing may reveal a rare causal variant.

GWAS and meta-analyses

Recent GWAS projects have combined their data in meta-analyses with other large scale genotyping studies to increase study sample and power and in turn increase variant identification. One of these recent studies by Al-Olama et al [13] analysed more than 10 million SNPs from GWASs carried out in populations of different ethnicities (European, African, Japanese and Latino populations) and analysed GWAS data for ~43,300 cases and ~43,700 controls. This led to the identification of 23 new PrCa associated SNPs.[13] This approach also allows replication in independent sample sets that overcomes subpopulation stratification effects in single population GWASs. Although rs636291 at 1p36 was found to be associated with early onset disease in men of European ancestry (OR 1.18, $P = 2.1 \times 10^{-8}$), this finding was not replicated in the subsequent OncoArray project (discussed below). .

The OncoArray identified prostate cancer risk

The most recent 63 PrCa risk loci identified were a result of a large GWAS and meta-analysis[3] carried out by groups in the OncoArray network.[14] The goal of this network is to gain new insights into the genetic architecture and mechanisms underlying common cancers through the use of a custom designed Illumina array, the OncoArray, to genotype SNPs in cases of the most common cancers (breast, ovary, endometrium, lung, colon, prostate) and cancer-free controls. Individuals that have a genetic predisposition to cancer such as *BRCA1* and *BRCA2* mutation carriers are also included. Each consortium participating in this project contributed between 10,000 to 100,000 cases and controls; inclusion of large numbers of cases and controls has led to the identification of both common and rare variants associated with cancer risk and is likely to identify variants that are shared across cancer types.

The OncoArray project was established, in part, through the efforts of the GAME-ON (Genetic Associations & MEchanisms in ONcology) network set up by the NCI (National Cancer Institute, USA). GAME-ON brings together international collaborators with the long-term goal of providing a rigorous knowledge base to enable clinical translation of GWAS findings. The importance of post GWAS research has been increasingly recognized as the majority of GWAS identified cancer risk SNPs lay in non-coding regions of DNA. Therefore, the biological mechanisms underlying PrCa development is not immediately clear. Computational methods such as enrichment analysis suggest that a large number of the PrCa risk SNPs (as well as other cancer SNPs) are enriched in multiple functional regions such as the binding regions of transcription factors or histone modifiers.[15] Experiments are required to validate these associations and to further characterise the molecular mechanisms underlying variant association with PrCa.

The OncoArray is a high density array comprising nearly 533,000 SNPs (Figure 3), of which approximately 70,000 are PrCa associated SNPs derived from a previous multi-ethnic meta-analysis [13] as well as from fine-mapping of known PrCa loci, and candidate variants. The OncoArray was used to genotype 46,939 PrCa cases and 27,910 controls of European ancestry.[3]

The OncoArray data (46, 939 PrCa cases and 27, 910 controls) were combined with data from other large scale genotyping studies including 32, 225 PrCa cases and 33, 202 controls (also of European ancestry) to carry out a meta-analysis of more than 140,000 men. This led to the identification of 63 novel loci related to PrCa susceptibility. Of these, 52 were identified by imputation of the OncoArray genotyping data. Imputation relies on linkage disequilibrium (LD) of SNPs which describes the inheritance of specific groups of SNPs that correlate with each other due to their relative vicinity in the genome. This allows a single 'tag' SNP to be used to infer the genotypes of other SNPs in the same LD block. Incorporation of the large GWAS backbone (260,000 SNPs) on the OncoArray allowed investigators to utilise LD to increase the power of variant discovery. The large numbers of cases and controls in this study also allowed several sub-analyses of clinical and demographic factors such as age at disease onset and aggressiveness of PrCa. A novel variant at 6q27 (rs138004030) was found to be significantly associated with early onset disease (OR= 1.27; $p=2.85x10^{-8}$). In an analysis of advanced PrCa cases, 4 variants were found to be significantly associated with advanced PrCa ($P<5x10^{-8}$). These were significantly associated with overall PrCa risk but when advanced and non-advanced cases were compared, there was only a marginal statistically significant difference observed($P<1.0x10^{-3}$).[3]

Several candidate genes were identified among the new 63 PrCa variants; one of these is an *ATM* missense variant rs1800057 (OR=1.16; $P=8.15x10^{-9}$; G>C [Pro1054Arg]). Although this missense mutation has been classified as 'benign' in the ClinVar database, *ATM* has been implicated in PrCa development and particularly with aggressive disease. The ATM protein is a key checkpoint kinase that acts as a regulator of a wide range of downstream proteins including TP53 and BRCA1, checkpoint kinase CHK2, checkpoint proteins RAD17 and RAD9, and DNA repair protein NBS1. It is therefore a key player in the DNA damage response pathway. Recent studies have investigated the frequency of germline mutations in DNA repair genes in PrCa. In familial cases, there was a 7% frequency of mutations reported [16] and in sporadic metastatic castration resistant (mCRPC) cases the frequency was even higher at approximately 12%.[17] In both studies, *ATM* was found to be the second most commonly altered gene (after *BRCA2*) and poorer outcomes were observed in mutation carriers. This pattern was confirmed in another recent study, where lethal and indolent PrCa cases were retrospectively compared with respect to germline *BRCA1, BRCA2* and *ATM* mutations. Those with lethal PrCa had a carrier rate of 6% compared to 1.44% in those with indolent disease (p=0.0007).[18]

Another missense variant (rs2066827) was identified in *CDKN1B* (cyclin dependent kinase inhibitor 1B) (OR=1.06; $P=2.31x10^{-9}$; T>G [Val109Gly]) which belongs to the *Cip/Kip* family of cyclin dependent kinase inhibitors. CDKN1B protein controls cell cycle progression at the G1 stage, and in vitro studies have shown levels of CDKN1B to be linked to increased tumour size and grade. This particular variant has previously been implicated in familial PrCa as well as advanced disease.

A third candidate gene was identified by a variant in an intron of *RASSF3* (rs7968403; OR=1.06; $P=3.38x10^{-12}$). RASSF3 is a GTP-binding plasma membrane protein and is a member of the RAS signalling pathway which is aberrant in approximately one third of cancers.

The findings described above are significant as very few risk variants have been found to lie within gene sequences. Functional studies of these variants are still required to further

delineate aberrant biological pathways but with the development of gene editing techniques such as CRISPR, further characterisation of the molecular mechanisms leading a variant to cause PrCa development or progression will be possible.

GWAS in non-European populations

Most large scale GWAS have been carried out in European populations, but there have been studies investigating common risk variants in other ethnic groups and population-specific differences have been reported. This may partially explain some of the geographical differences in PrCa incidence rates. In a study by Marzec et al,[19] 4 SNPs on 8q24 were found to contribute to the risk of PrCa development in Chinese men. Although, these variants did not reach GWAS significance level, their frequency was considerably higher compared with the frequency in European men.[19]

In a meta-analysis of two GWAS from Japanese (1,583 cases and 3,386 controls) and Chinese (1,417 cases and 1,008 controls) populations which included replication in independent samples sets (also of Japanese and Chinese men),[20] 2 independent PrCa susceptibility loci were reported: rs12791447 at 11p15.4, and rs58262369 at 14q23.2. In addition, 33 SNPs of the 100 known (at that time) PrCa SNPs were also found to be significant. rs58262369 is located in the 3'-untranslated region (3'-UTR) of *ESR2* which encodes the oestrogen receptor, ESR2. Animal studies have shown high expression of *ESR2* in normal prostate epithelial cells and *ESR2* knockout mice are found to develop prostate hyperplasia.[20] TCGA (The Cancer Genome Atlas) data shows that *ESR2* is up-regulated in prostate tumours compared to normal prostate tissue, however this didn't correlate with increased mRNA expression therefore further study of protein expression in tumours is needed. The second significant SNP identified, rs12791447 lies in an intron of *PPFIBP2* which encodes the PTPRF-interacting protein, binding protein 2 (Liprin Beta 2). TCGA data shows that the expression level of this gene is significantly lower in tumours compared to normal prostate tissue, and eQTL analysis also linked rs12791447 with expression of *PPFIBP2*. Additionally, this SNP has been shown to map to a strong enhancer region and therefore PPFIBP2 has been suggested as the candidate susceptibility gene associated with this variant.[20]

Epidemiological data have long shown that the risk of PrCa development as well as the mortality rate is increased in men of African descent compared to other ethnicities. Explanation of this increased risk is likely to be due to a number of factors including social and environmental factors as well as hereditary genetics. Studies of the genetic component contributing to the increased risk have identified SNPs in the 8q24 risk region that are specific to African men (rs114798100 and rs111906923)[21]. This may be explained by the finding that African ancestry is over represented in the 8q24 region. A 2017 GWAS meta-analysis[22] of African men combined data from 2 PrCa consortia and 2 large PrCa studies to compare the genotypes of over 10,000 cases and over 10,000 controls. Two novel signals were identified: rs75823044 on 13q34 (OR 1.55, 95% confidence interval (CI) 1.37-1.76), and rs78554043 on 22q12.1 (OR 1.62, 95% CI 1.39-1.89). Interestingly, the latter SNP is a missense variant in the *CHEK2* gene leading to a non-conservative amino acid change (Ile448Ser). With the recent data reporting a correlation between mutations (both germline and somatic) in DNA repair genes and the development of mCRPC, this may be a significant

finding in men of African background who tend to be diagnosed with more aggressive disease compared to non-African ancestry men.

As population specific GWAS continue to be carried out and meta-analysed, further ethnic and population differences in susceptibility variants are likely to emerge and these may partly explain geographical and population differences in PrCa incidence rates as well as phenotypic differences observed. These may suggest alternative biological mechanisms leading to PrCa in different populations.

Radiogenomics GWAS

Definitive treatment of PrCa comprises either radical prostatectomy (RP) or radiotherapy (RT) in the form of external beam radiotherapy (EBRT) or brachytherapy. As discussed above, a better understanding of germline SNPs may guide the degree of intervention offered to a man depending on risk stratification. In those who are in need of definitive treatment of localised disease, factors influencing choice of treatment modality include the staging of localised tumour, age and comorbidities, as well as a patient's preferences. In those that undergo RP and go on to relapse with localised disease or biochemical recurrence only, RT is also often administered as salvage treatment.

Long-term effects observed after prostate RT include urinary frequency, urgency, erectile dysfunction and rectal bleeding. These symptoms are severe in 2-3% of men. [23] Predictive markers of both radiation toxicity and radio-resistance could inform decision making at the time of planning treatment in an effort to avoid the development of long term side effects and optimise outcomes. The relationship between PrCa risk loci and the development of RT side effects has been investigated and there appears to be no increased risk of late radiation effects (e.g. urinary frequency and rectal bleeding) in men with a higher risk genetic profile.[24]

The GWAS approach has been utilised to identify common genetic variants that are associated with the development of radiation side effects after treatment of PrCa as well as other cancer types. These have not been on as large a scale as some of the recent PrCa GWAS but have reported several associated SNPs (Table 2).[25-30] The establishment of the Radiogenomics Consortium in 2009 has brought together existing collaborative groups to allow sharing of expertise, pooling of data, as well as the set-up of replication studies to validate results of GWAS involving small sample sizes.[31] As a result, a consortium led meta-analysis of 4 radiogenomics GWAS was published in 2016 and identified two further SNPs associated with late RT toxicity [25]: rs17599026 on 5q31.2 was associated with urinary frequency (OR 3.12, 95% CI 2.08-4.69, $P=4.16 \times 10^{-8}$) and rs7720298 on 5p15.2 was associated with reduced urine stream (OR 2.71, 95% CI 1.90–3.86, $P=3.21 \times 10^{-8}$).

As with other GWAS identified SNPs, both these variants lie in non-coding regions. rs17599026 is located 23bp downstream of exon 20 of the gene *KDM3B*. This gene encodes the protein lysine-specific demethylase 3B and is highly expressed in bladder tissue, therefore it could feasibly be disrupted by radiation leading to bladder dysfunction.[25] Whether rs17599026 is influencing the expression of *KDM3B* or is in LD with another locus that is influential is yet to be determined as this SNP tags a large LD block containing multiple regulatory regions and therefore may also contain the true causal variant. The second SNP identified, rs7720298, causes reduced urine stream and lies downstream of

exon 30 of *DNAH5*, which encodes the dynein, axonemal, heavy chain 5 protein that is part of a microtubule associated motor protein complex.[25] This complex has an important role in the normal function of cilia in the lung and is also expressed in kidney and bladder tissue therefore its abnormal expression of this gene may lead to dysfunction in the urinary tract. As with rs17599026, this SNP tags an LD block containing transcriptional regulatory regions that may in fact contain the causal SNP.

Utility of GWAS findings

Prostate cancer susceptibility loci and polygenic risk profiles

Based on 147 known prostate cancer susceptibility loci (21 SNPs excluded as did not replicate in OncoArray GWAS) we can explain approximately 28.4% of FRR. Although the OR associated with individual risk variants are low, most range between 1.0-1.6, the PrCa risk conferred by risk alleles is cumulative with increasing number of alleles in the germline of an individual. With no current population screening programs for PrCa, there may well be a unique public health role for using SNP/genetic profiles to stratify men by their risk of PrCa and targeting screening to those at increased risk above a certain threshold. Indeed, in the latest GWAS and meta-analysis reported by Schumacher et al[32], it was reported that by applying the current known PrCa SNPs into a polygenic risk score (PRS), the relative risk (RR) of PrCa for men in the top 1% of the risk distribution is 5.71 relative to men in the 25-75% PRS percentiles (95%CI: 5.04-6.48), and for those in the top 10% the RR is 2.69 (95%CI: 2.55-2.82). Notably, the PRS effect estimates increased with a positive family history or young age of onset (≤55 years). A risk model incorporating a SNP profile along with family history information could be used to target screening to those at highest risk. Several retrospective studies have analysed the use of risk models incorporating SNP profiles as well as other factors such as family history.[33],[34] The Stockholm 3 (STHLM-3) study reported in 2015 by Gronberg et al [35] was the first large prospective PrCa screening study utilising a screening algorithm which incorporated a genetic profile, in addition to plasma biomarkers and a set of clinical variables. This study recruited approximately 48,000 men aged 50-69 and had a paired design so that each study participant had both a PSA test (threshold of ≥3ng/ml) and the STHLM-3 algorithm to assess their PrCa risk. The STHLM-3 model outperformed PSA screening for the detection of moderate to high risk PrCa (Gleason score ≥7) with an AUC for the detection of high risk cancers of 0.76 compared to 0.56 for PSA based screening. Use of the study model also reduced the number of overall prostate biopsies by one-third, the number of benign biopsies by 44% and the number of low grade (Gleason 6) cancers diagnosed by 17%. Interestingly, approximately 20% of Gleason 7 or higher cancers detected by the STHLM-3 model had a PSA under 3 and would have been missed by PSA based screening. It is difficult to analyse the contribution of the 232 SNP profile to the STHLM-3 screening model but all variables in the model contributed to a cumulative improvement in the AUC in the multivariate analysis, and were significantly associated with high risk PrCa. Further prospective studies are underway to assess the use of SNP profiles in the context of population screening. One of these is the BARCODE1 study (NCT03158922) in the UK which is currently recruiting to a pilot study with a plan to expand the full study to 5000 participants. This study is recruiting men aged 55-69 years from the community who will have germline DNA genotyped for the 167 known PrCa risk variants. A PRS will be calculated for each participant and men who fall in the top 10% of the risk distribution will undergo screening for PrCa by prostate biopsy. The primary endpoint of this study will be the association of PRS with prostate biopsy results.

PrCa screening currently relies on PSA measurement with or without digital rectal examination of the prostate. Large screening studies (PLCO[36] in the USA and ERSPC[37] in Europe) have demonstrated the lack of specificity of PSA levels for PrCa detection, with a low positive predictive value of approximately 25% which is reflected in the high proportion of benign tissue or low grade and clinically insignificant tumours biopsied in these studies. PSA levels are known to fluctuate and are influenced by many factors including infections, ejaculation and systemic inflammation. The current consensus from regulatory bodies such as NICE (National Institute for Clinical Excellence) in the UK and the USPSTF (US Preventive Services Task Force) in the USA has been to advise against the use of PSA measurement for population screening as the impact of over-diagnosis and overtreatment of indolent disease outweighs the benefit of detection of PrCa. Therefore, germline DNA and SNP genotypes may be an ideal biomarker to use in a screening program as this is a largely constant marker that is easily accessible and requires a single measurement. The data from ongoing studies such as BARCODE-1 and STHLM3-MR (incorporating targeted biopsy based on MRI of the prostate into the STHLM3 model; NCT02788825) will provide the evidence needed to guide new PrCa screening strategies.

Germline variants influencing somatic alterations

Studies have been carried out to assess the utility of SNPs as biomarkers for clinicopathological features of PrCa. An example of this is a study reported by Luedeke et al linking PrCa SNPs to *TMPRSS2-ERG* fusion status in somatic (tumour) DNA.[38] The *TMPRSS2-ERG* gene fusion occurs in approximately 50% of PrCa cases in European men and is an early event in prostate carcinogenesis. It also appears to be associated with early onset disease, but isn't clearly correlated with aggressive disease. In this study, the germline DNA of PrCa cases that were gene fusion positive and negative were assessed for 27 common risk variants; these two subgroups were additionally compared with controls without PrCa. Two SNPs were identified to be significantly associated with somatic fusion status: rs16901979 (8q24) was enriched in fusion negative cases and rs1859962 (17q24) was enriched in fusion positive cases.[38] Interestingly, an additive effect of the genotypes was observed, for example homozygous carriers of the 8q24 variant had a frequency of 33% for *TMPRSS2-ERG* fusion compared to the overall frequency of 56%.This study went on to functionally assess the interplay between associated germline variants and somatic *TMPRSS2-ERG* using mRNA expression analysis of candidate target genes in fusion-positive and negative tumour tissue. Their results suggest that the 17q24 risk allele promotes *ERG*-mediated changes in expression of *SOX9* in fusion positive tumours, which is not seen in fusion negative tumours. *SOX9* is a transcription factor involved in prostate epithelial development, consequently, the synergistic effect of germline variant and gene fusion could render advantages to precursor cells in tumour formation with up-regulation of transcription factors such as *SOX9*.
Ascertaining somatic fusion status as well as other molecular markers in subsequent GWAS may well identify subtype-specific risk variants in the future. This could lead to their use as biomarkers to guide the risk stratification and/ or management of men with PrCa.

Common variants and precision medicine

In coming years, it is feasible that we may be able to (quickly and accurately) obtain and utilise the germline genetic profile of a man with PrCa at the time of diagnosis to set up a personalised treatment pathway according to the risk category indicated by the SNPs in his

germline, the presence or absence of risk loci for radiotoxicity, as well as the presence or absence of variants (both common and rare) in DNA repair genes and others. Combining this information with a somatic genetic profile could lead to a highly personalised treatment pathway.

SNPs as predictors of aggressive prostate cancer

As well as indicating previously unknown mechanisms of disease, common genetic variants may act as predictors of aggressive PrCa in those that are newly diagnosed. This would enable more accurate risk stratification and in turn tailored management of an individual's PrCa. Current clinico-pathological biomarkers used to assign risk of disease progression or relapse after treatment includes tumour stage, presenting PSA level and Gleason score (GS). Studies have assessed the relationship of germline SNPs with treatment outcomes in an attempt to define better risk stratification models in this context.

In a study of men undergoing either surgery or active surveillance for low grade (GS 6) PrCa, the genotypes of 23 SNPs were assessed with respect to upgrading of pathological grade.[39] In 950 men with a GS 6 PrCa treated with surgery, 3 SNPs were found to be significantly associated with upgrade of GS: rs11568818 on chromosome 11, rs2427345 on chromosome 20 and rs7141529 on chromosome 14. After correction for multiple testing, only the chromosome 11 SNP remained significantly associated with upgrading of GS following surgery (OR 1.46, 95% CI: 1.17–1.82 P=0.0009). The 3 SNPs were also assessed prospectively in a smaller (n=209) cohort of men with GS 6 cancer managed with active surveillance (AS). The hazard ratios analysed for time to upgrade of GS ranged from 3.7 to 5.3 and were statistically significant for all 3 SNPs.[39]

In another study analysing SNPs prospectively in an AS cohort (n=494 with GS≤7 cancers), rs2735839 (KLK3 region on chromosome 19q13) was found to be significantly associated with GS reclassification.[40] There was also a suggestive association of another SNP, rs752822 (chromosome 5q32). Together, these two risk alleles had an OR of 2.71 (95% CI = 1.62‑4.51, $P = 1 \times 10^{-4}$) in the initial SNP identification phase of the study, and a disease reclassification HR of 1.89 (95% CI = 1.13‑3.18, P = 0.016) when genotyping 494 men managed by AS.

Germline SNPs found to be significantly associated with developing higher grade disease which in turn may increase the risk of progression or metastatic disease if not radically treated, may well be more accurate biomarkers for risk stratification at diagnosis than those currently used. Incorporating a validated genetic profile into the work up of men diagnosed with clinically 'low' risk PrCa would allow a more informed approach to decision making by both patient and clinician. Indeed, if similar germline predictors of outcome are identified for men traditionally thought to be at intermediate risk (e.g. GS 7, PSA 10-20) as in the latter study discussed, a proportion of men may be safely managed with less interventional approaches if they are found to have a low risk genetic profile.

In both the studies discussed above, there are plausible biological mechanisms linking the risk variants to PrCa. rs11568818 on chromosome 11 may contribute to PrCa invasiveness as it lies in close proximity to the gene, *MMP7*.[39, 41] This gene encodes matrix metalloproteinase 7, and rs11568818 lies 181bp upstream of its transcription start site. The base change at this locus is thought to affect the binding of transcription factors which in turn

alters *MMP7* expression. MMP7 in addition to other matrix metalloproteinases is involved in extracellular proteolysis and acts as a modulator of the tumour microenvironment.[42] Up-regulation of MMPs promotes tumour invasion and metastasis. MMP7 appears to influence a number of ligands to achieve this, such as Fas ligand which is a mediator of cell death;[41] cleavage of Fas ligand by MMP7 reduces chemotherapeutic induced cellular apoptosis. Investigation of *MMP7* in PrCa is needed to identify whether similar mechanisms are at play when tumours progress and metastasise.

In the second study discussed above,[40] rs2735839 lies within the PSA coding gene *KLK3* and has previously been found to be associated with PrCa risk and aggressive disease in several studies.[43] It may be that the association with aggressive PrCa is linked to higher PSA levels associated with this SNP and increased diagnoses through PSA screening, although the prospective context of the current study suggests an independent association with an aggressive phenotype. Further prospective studies are needed to define this association further. The 5q32 SNP (rs752822) lies within an intron of *CSNK1A1* (Casein Kinase 1, Alpha 1) and forms part of the β-catenin destruction complex.[40] This complex interacts with the Wnt signalling pathway and leads to accumulation of β-catenin in the nucleus which in turn increases expression of *KLK3* and other genes involved in prostate carcinogenesis.

Although there are plausible mechanisms linking these SNPs to PrCa development, the GWAS utility of 'tag' SNPs that lie in LD blocks means that these loci may not be the causative variants but instead may be in LD with SNPs in the same region. Fine mapping studies post GWAS will be necessary to allow further assessment of other variants in these regions.

Decisions related to treatment modality

As well as markers of radiotoxicity, identification of patients less likely to respond to radiation would guide decision making early on in the curative treatment pathway for localised PrCa. In a study of common variants of the *BRCA1* gene in men who underwent definitive treatment for localised PrCa, two *BRCA1* SNPs were found to have a significant association with lethal PrCa after RT:[44] rs4474733 was associated with a 35% lower risk of lethal PrCa and rs8176305 was associated with a two-fold increased risk of lethal PrCa. There was no significant association observed in men treated with surgery.[44]

BRCA1 plays a major role in DNA damage repair and cell cycle control. In PrCa cell lines, *BRCA1* has been shown to be involved in radiosensitivity, which can be explained by its role in repair of double strand DNA breaks induced by radiation. If these results are validated in future studies, these SNPs may act as biomarkers of radio-resistance and if utilised clinically, may lead to modification of the treatment pathway for men with a specific genetic profile. Interestingly, in a retrospective study of outcomes in carriers of rare *BRCA* mutations after radical treatment, men with germline mutations were found to have worse survival rates compared to non-carriers when treated with RT but no difference was seen in those treated with surgery.[45] This study was not designed to compare outcomes in *BRCA* mutation carriers according to treatment modality, but the observations noted may fit with the patterns seen with *BRCA1* common variants. A prospective study of outcomes in *BRCA* mutation carriers as well as carriers of common variants is needed to assess this further.

As the field of radiogenomics advances both in PrCa as well as other cancer types, genetic profiles could be incorporated into already existing normal tissue complication probability (NTCP) models to allow personalisation of radiotherapy delivery.[31] Modified pathways may include isotoxic dose escalation in those found to be radio-resistant (or radiation tolerant), or dose reductions or hyper-fractionation for radiosensitive individuals. In the latter category of patients, adding or increasing systemic therapy may also be considered to optimise treatment.[31] In a set of simulation experiments by Kerns et al [46] using both a 'low penetrance' distribution of radiogenomic SNPs as well as a 'moderate penetrance' distribution, the addition of a set of SNPs to current NTCP models increased the AUC in all models and improved discrimination accuracy. For example, in the 'low penetrance' distribution model, the inclusion of 78 SNPs could increase the AUC to 0.80 compared with 0.70 using NTCP model alone for the prediction of late effects. In a 'moderate penetrance' model, the number of SNPs needed is even lower at 47.[46] By utilising more accurate models of radiotoxicity that incorporate an individual's genetic profile, radiotherapy plans can be modified to optimise treatment dose while avoiding long term side effects that will impact on quality of life.

Future Directions

The rapid advances and falling costs of DNA sequencing, along with the collaboration between international consortia and research groups, have led to an accumulation of data that we have yet to fully understand. Although a large number of common variants associated with PrCa have been identified through GWAS, the mechanisms explaining causality have not been elucidated. Fine mapping studies continue and will help to identify causal SNPs that may be in LD with GWAS identified index SNP. Functional studies and gene editing techniques will allow characterisation of variant interactions with cell signalling and other pathways. This may reveal novel targets for treatment in addition to explaining mechanisms of disease.

The COGS project (Collaborative Oncological Gene-Environment Study) brings together international research groups to promote the study of germline variant interactions with lifestyle factors and understand the impact of this interaction on PrCa risk (as well as other cancers). It is possible that even without fully characterising genetic variants identified thus far, that risk alleles may be used as biomarkers of clinico-pathological features or as risk stratifiers in the assessment of PrCa cases. Therefore the use of genetic variants in clinical practice needs to be established. Indeed, commercial assays incorporating SNPs linked to several conditions are already available, many in the form of 'direct to consumer' tests which utilise DNA extracted from saliva. With the high number of PrCa risk SNPs now known, it is likely that commercial assays incorporating these SNPs will become available.

Approximately 28.4% of the familial relative risk of PrCa can now be explained by 167 known common variants. A further ~6% may be explained by rare but higher risk variants such as *BRCA2* and *HOXB13*. As we gain a better understanding of the hereditary genetics of PrCa, we will be able to utilise this information to generate genetic profiles that can be incorporated into screening programs for PrCa. By targeting screening to those most at risk, the current drawbacks of PSA based testing including over-biopsy of indolent or benign tissue may be avoided. Information from germline genetic profiles related to the risk of aggressive disease and radiogenomics among other factors, can be combined with somatic tumour profiles at

the time of diagnosis to generate a personalised treatment pathway and work towards achieving true precision medicine in the management of PrCa.

Acknowledgements

We would like to thank the collaborators in the PRACTICAL Consortium for their efforts in advancing research into prostate cancer genetic predisposition.

References

1. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/incidence#heading-Nine.
2. Mucci, L.A., et al., *Familial Risk and Heritability of Cancer Among Twins in Nordic Countries.* JAMA, 2016. **315**(1): p. 68-76.
3. Schumacher, F. et al, *Prostate cancer meta-analysis of more than 140,000 men identifies 63 novel prostate cancer susceptibility loci.* Nature Genetics, 2017. **(under review)**.
4. Amundadottir, L.T., et al., *A common variant associated with prostate cancer in European and African populations.* Nat Genet, 2006. **38**(6): p. 652-8.
5. Ewing, C.M., et al., *Germline mutations in HOXB13 and prostate-cancer risk.* N Engl J Med, 2012. **366**(2): p. 141-9.
6. Shang, Z., et al., *Germline homeobox B13 (HOXB13) G84E mutation and prostate cancer risk in European descendants: a meta-analysis of 24,213 cases and 73, 631 controls.* Eur Urol, 2013. **64**(1): p. 173-6.
7. Xu, J., et al., *HOXB13 is a susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG).* Hum Genet, 2013. **132**(1): p. 5-14.
8. Gudmundsson, J., et al., *Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24.* Nat Genet, 2007. **39**(5): p. 631-7.
9. Du, M., et al., *Prostate cancer risk locus at 8q24 as a regulatory hub by physical interactions with multiple genomic loci across the genome.* Hum Mol Genet, 2015. **24**(1): p. 154-66.
10. http://www.genecards.org/cgi-bin/carddisp.pl?gene=CD96.
11. Liekens, A.M., et al., *BioGraph: unsupervised biomedical knowledge discovery via automated hypothesis generation.* Genome Biol, 2011. **12**(6): p. R57.
12. Saunders, E.J., et al., *Fine-mapping the HOXB region detects common variants tagging a rare coding allele: evidence for synthetic association in prostate cancer.* PLoS Genet, 2014. **10**(2): p. e1004129.
13. Al Olama, A.A., et al., *A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer.* Nat Genet, 2014. **46**(10): p. 1103-9.
14. Amos, C.I., et al., *The OncoArray Consortium: a Network for Understanding the Genetic Architecture of Common Cancers.* Cancer Epidemiol Biomarkers Prev, 2016.
15. Chen, H., et al., *Systematic enrichment analysis of potentially functional regions for 103 prostate cancer risk-associated loci.* Prostate, 2015. **75**(12): p. 1264-76.
16. Leongamornlert, D., et al., *Frequent germline deleterious mutations in DNA repair genes in familial prostate cancer cases are associated with advanced disease.* Br J Cancer, 2014. **110**(6): p. 1663-72.
17. Pritchard, C.C., et al., *Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer.* N Engl J Med, 2016. **375**(5): p. 443-53.
18. Na, R., et al., *Germline Mutations in ATM and BRCA1/2 Distinguish Risk for Lethal and Indolent Prostate Cancer and are Associated with Early Age at Death.* Eur Urol, 2016.

19. Marzec, J., et al., *A genetic study and meta-analysis of the genetic predisposition of prostate cancer in a Chinese population.* Oncotarget, 2016. **7**(16): p. 21393-403.
20. Wang, M., et al., *Large-scale association analysis in Asians identifies new susceptibility loci for prostate cancer.* Nat Commun, 2015. **6**: p. 8469.
21. Han, Y., et al., *Prostate Cancer Susceptibility in Men of African Ancestry at 8q24.* J Natl Cancer Inst, 2016. **108**(7).
22. David V. Conti, K.W., Xin Sheng, Jeanette T. Bensen, Dennis J. Hazlett, Michael B. Cook, Sue A. Ingles, Rick A. Kittles, *Two Novel Susceptibility Loci for Prostate Cancer in Men of African Ancestry.* Journal of the National Cancer Institute, 2017. **109**(8).
23. Ohri, N., A.P. Dicker, and T.N. Showalter, *Late toxicity rates following definitive radiotherapy for prostate cancer.* Can J Urol, 2012. **19**(4): p. 6373-80.
24. Ahmed, M., et al., *Common genetic variation associated with increased susceptibility to prostate cancer does not increase risk of radiotherapy toxicity.* Br J Cancer, 2016. **114**(10): p. 1165-74.
25. Kerns, S.L., et al., *Meta-analysis of Genome Wide Association Studies Identifies Genetic Markers of Late Toxicity Following Radiotherapy for Prostate Cancer.* EBioMedicine, 2016. **10**: p. 150-63.
26. Fachal, L., et al., *A three-stage genome-wide association study identifies a susceptibility locus for late radiotherapy toxicity at 2q24.1.* Nat Genet, 2014. **46**(8): p. 891-4.
27. Kerns, S.L., et al., *A 2-stage genome-wide association study to identify single nucleotide polymorphisms associated with development of urinary symptoms after radiotherapy for prostate cancer.* J Urol, 2013. **190**(1): p. 102-8.
28. Kerns, S.L., et al., *Genome-wide association study identifies a region on chromosome 11q14.3 associated with late rectal bleeding following radiation therapy for prostate cancer.* Radiother Oncol, 2013. **107**(3): p. 372-6.
29. Kerns, S.L., et al., *Genome-wide association study to identify single nucleotide polymorphisms (SNPs) associated with the development of erectile dysfunction in African-American men after radiotherapy for prostate cancer.* Int J Radiat Oncol Biol Phys, 2010. **78**(5): p. 1292-300.
30. Kerns, S.L., et al., *A 2-stage genome-wide association study to identify single nucleotide polymorphisms associated with development of erectile dysfunction following radiation therapy for prostate cancer.* Int J Radiat Oncol Biol Phys, 2013. **85**(1): p. e21-8.
31. Barnett, G.C., et al., *Incorporating Genetic Biomarkers into Predictive Models of Normal Tissue Toxicity.* Clin Oncol (R Coll Radiol), 2015. **27**(10): p. 579-87.
32. Schumacher, F., *Prostate cancer meta-analysis of more than 142,000 men identifies 65 novel prostate cancer susceptibility loci.* Nature Genetics, 2017. **(under review)**.
33. Macinnis, R.J., et al., *A risk prediction algorithm based on family history and common genetic variants: application to prostate cancer with potential clinical impact.* Genet Epidemiol, 2011. **35**(6): p. 549-56.
34. Liss, M.A., et al., *Prostate genetic score (PGS-33) is independently associated with risk of prostate cancer in the PLCO trial.* Prostate, 2015. **75**(12): p. 1322-8.

35. Gronberg, H., et al., *Prostate cancer screening in men aged 50-69 years (STHLM3): a prospective population-based diagnostic study.* Lancet Oncol, 2015. **16**(16): p. 1667-76.
36. Andriole, G.L., et al., *Mortality results from a randomized prostate-cancer screening trial.* N Engl J Med, 2009. **360**(13): p. 1310-9.
37. Schroder, F.H., et al., *Screening and prostate-cancer mortality in a randomized European study.* N Engl J Med, 2009. **360**(13): p. 1320-8.
38. Luedeke, M., et al., *Prostate cancer risk regions at 8q24 and 17q24 are differentially associated with somatic TMPRSS2:ERG fusion status.* Hum Mol Genet, 2016.
39. Kearns, J.T., et al., *Associations Between iCOGS Single Nucleotide Polymorphisms and Upgrading in Both Surgical and Active Surveillance Cohorts of Men with Prostate Cancer.* Eur Urol, 2016. **69**(2): p. 223-8.
40. Shu, X., et al., *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**(10): p. 965-71.
41. Cooper, C.S., et al., *Prostate Single Nucleotide Polymorphism Provides a Crucial Clue to Cancer Aggression in Active Surveillance Patients.* Eur Urol, 2016. **69**(2): p. 229-30.
42. Kessenbrock, K., V. Plaks, and Z. Werb, *Matrix metalloproteinases: regulators of the tumor microenvironment.* Cell, 2010. **141**(1): p. 52-67.
43. Eeles, R.A., et al., *Multiple newly identified loci associated with prostate cancer susceptibility.* Nat Genet, 2008. **40**(3): p. 316-21.
44. Sanchez, A., et al., *Common variation in BRCA1 may have a role in progression to lethal prostate cancer after radiation treatment.* Prostate Cancer Prostatic Dis, 2016. **19**(2): p. 197-201.
45. Castro, E., et al., *Effect of BRCA Mutations on Metastatic Relapse and Cause-specific Survival After Radical Treatment for Localised Prostate Cancer.* Eur Urol, 2015. **68**(2): p. 186-93.
46. Kerns, S.L., et al., *The Prediction of Radiotherapy Toxicity Using Single Nucleotide Polymorphism-Based Models: A Step Toward Prevention.* Semin Radiat Oncol, 2015. **25**(4): p. 281-91.

Table 1: Prostate cancer associated variants discovered through GWAS

RAF= Risk allele frequency; OR= Odds ratio

SNP ID	Band	Alleles^	RAF	OR	Reference/ PMID
63 novel SNPs identified on the OncoArray					
rs56391074	1p22.3	A/AT	0.379	1.05	4
rs34579442	1q21.3	C/CT	0.336	1.07	4
rs74702681	2p14	T/C	0.022	1.15	4
rs62106670	2p25.1	T/C	0.379	1.05	4
rs11691517	2q13	T/G	0.741	1.05	4
rs34925593	2q31.1	C/T	0.481	1.06	4
rs59308963	2q33.1	T/TATTCTG TC	0.729	1.05	4
rs1283104	3q13.12	G/C	0.379	1.04	4
rs182314334	3q25.1	T/C	0.895	1.10	4
rs142436749	3q26.2	G/A	0.012	1.23	4
rs10793821	5q31.1	T/C	0.573	1.05	4
rs76551843	5q35.1	A/G	0.991	1.31	4
rs4976790	5q35.3	T/G	0.113	1.08	4
rs4711748	6p21.1	T/C	0.225	1.05	4
rs9469899	6p21.31	A/G	0.357	1.05	4
rs9296068	6p21.32	T/G	0.651	1.05	4
rs12665339	6p21.33	G/A	0.167	1.06	4
rs17621345	7p14.1	A/C	0.741	1.07	4
rs11452686	7p21.1	T/TA	0.558	1.04	4
rs527510716	7p22.3	C/G	0.241	1.07	4
rs10122495	9p13.3	T/A	0.29	1.05	4
rs1048169	9p22.1	C/T	0.379	1.07	4
rs1182	9q34.11	A/C	0.220	1.07	4
rs141536087	10p15.3	GCGCA/G	0.150	1.10	4
rs1935581	10q23.31	C/T	0.623	1.06	4
rs7094871	10q25.2	G/C	0.537	1.04	4
rs547171081	11p11.2	CGG/C	0.470	1.05	4
rs1881502	11p15.5	T/C	0.190	1.06	4
rs61890184	11p15.4	A/G	0.124	1.08	4
rs2277283	11q12.3	C/T	0.313	1.06	4
rs12785905	11q13.2	C/G	0.048	1.09	4
rs11290954	11q13.5	AC/A	0.676	1.07	4
rs1800057	11q22.3	G/C	0.025	1.13	4
rs138466039	11q24.2	T/C	0.01	1.28	4
rs878987	11q25	G/A	0.146	1.07	4
rs2066827	12p13.1	T/G	0.755	1.07	4
rs10845938	12p13.1	G/A	0.551	1.06	4
rs7968403	12q14.2	T/C	0.643	1.07	4

rs5799921	12q21.33	GA/G	0.699	1.08	4
rs7295014	12q24.33	G/A	0.34	1.06	4
rs1004030	14q11.2	T/C	0.59	1.04	4
rs11629412	14q13.3	C/G	0.578	1.06	4
rs4924487	15q15.1	C/G	0.84	1.06	4
rs33984059	15q21.3	A/G	0.978	1.20	4
rs112293876	15q22.31	C/CA	0.289	1.07	4
rs11863709	16q21	C/T	0.960	1.17	4
rs201158093	16q23.3	TAA/TA	0.440	1.05	4
rs28441558	17p13.1	C/T	0.056	1.14	4
rs142444269	17q11.2	C/T	0.788	1.08	4
rs2680708	17q22	G/A	0.605	1.04	4
rs8093601	18q21.2	C/G	0.44	1.04	4
rs28607662	18q21.2	C/T	0.096	1.07	4
rs12956892	18q21.32	T/G	0.30	1.05	4
rs533722308	18q21.33	CT/C	0.412	1.05	4
rs10460109	18q22.3	T/C	0.42	1.04	4
rs11666569	19p13.11	C/T	0.711	1.06	4
rs118005503	19q12	G/C	0.911	1.11	4
rs61088131	19q13.2	T/C	0.835	1.05	4
rs11480453	20q11.21	C/CA	0.602	1.05	4
rs6091758	20q13.2	G/A	0.464	1.09	4
rs9625483	22q12.1	A/G	0.027	1.17	4
rs17321482	23p22.2	C/T	0.866	1.07	4
rs138004030*	6q27	G/A	0.92	1.28	4

^Reference allele/ risk allele

*Associated with early-onset PrCa

SNPs discovered in European populations

SNP ID	Chromosome	Risk alleles	RAF	OR	Reference/ PMID
rs636291	1	A	0.683	1.04	25217961
rs17599629	1	G	0.218	1.07	25217961
rs1218582	1	G	0.4467	1.05	23535732
rs4245739	1	A	0.738	1.10	20197460, 23535732
rs11902236	2	T	0.269	1.05	23535732
rs9287719	2	C	0.467	1.07	25217961
rs9306895	2	C	0.364	1.08	26025378
rs1465618	2	T	0.2141	1.09	19767753
rs721048	2	A	0.1822	1.10	18264098
rs10187424	2	T	0.5738	1.08	21743467
rs12621278	2	A	0.9414	1.27	19767753
rs2292884	2	G	0.2413	1.06	21743057
rs3771570	2	T	0.1495	1.09	23535732
rs2660753	3	T	0.1028	1.13	18264097
rs7611694	3	A	0.579	1.09	23535732

rs10934853	3	A	0.277	1.10	19767754
rs6763931	3	A	0.442	1.04	21743467
rs10936632	3	A	0.5074	1.10	21743467
rs10009409	4	T	0.311	1.06	25217961
rs1894292	4	G	0.515	1.06	23535732
rs12500426	4	A	0.4632	1.07	19767753
rs17021918	4	C	0.6507	1.09	19767753
rs7679673	4	C	0.592	1.13	19767753
rs2242652	5	G	0.794	1.17	21743467
rs2121875	5	C	0.33	1.05	21743467
rs6869841	5	T	0.209	1.04	23535732
rs4713266	6	C	0.517	1.05	25217961
rs7767188	6	A	0.210	1.06	25217961
rs130067	6	G	0.2021	1.05	21743467
rs3096702	6	A	0.3771	1.06	23535732
rs3129859	6	G	0.670	1.06	25217961
rs2273669	6	G	0.1462	1.07	23535732
rs1933488	6	A	0.5788	1.08	23535732
rs9364554	6	T	0.2826	1.11	18264097
rs12155172	7	A	0.220	1.10	23535732
rs10486567	7	G	0.763	1.14	18264096
rs56232506	7	A	0.451	1.06	25217961
rs6465657	7	C	0.4635	1.11	18264097
rs2928679	8	A	0.437	1.05	19767753
rs11135910	8	T	0.1529	1.08	23535732
rs12543663	8	C	0.295	1.12	19767752
rs10086908	8	T	0.697	1.13	19767752
rs183373024	8	G	0.007	2.91	23104005
rs16901979	8	A	0.032	1.56	17401366
rs620861	8	G	0.631	1.15	19767752
rs6983267	8	G	0.511	1.22	17401363
rs1447295	8	A	0.107	1.41	17401363, 17401366
rs17694493	9	G	0.136	1.08	25217961
rs1571801	9	T	0.268	1.03	18073375
rs76934034	10	T	0.917	1.12	25217961
rs10993994	10	T	0.383	1.23	18264096, 18264097
rs3850699	10	A	0.700	1.07	23535732
rs4962416	10	C	0.2668	1.06	18264096
rs7127900	11	A	0.1985	1.19	19767753
rs7931342	11	G	0.504	1.17	18264097
rs11568818	11	T	0.550	1.08	23535732
rs11214775	11	G	0.709	1.07	25217961
rs80130819	12	A	0.908	1.10	25217961
rs10875943	12	C	0.287	1.07	21743467

rs902774	12	A	0.1526	1.13	21743057
rs1270884	12	A	0.482	1.07	23535732
rs8008270	14	C	0.8139	1.09	23535732
rs7141529	14	C	0.499	1.05	23535732
rs8014671	14	G	0.580	1.05	25217961
rs684232	17	C	0.3534	1.09	23535732
rs11649743	17	G	0.8055	1.13	18758462
rs4430796	17	A	0.5253	1.22	17603485
rs138213197	17	T	0.002	3.85	22236224
rs11650494	17	A	0.0779	1.10	23535732
rs1859962	17	G	0.4813	1.17	17603485
rs7241993	18	C	0.6949	1.08	23535732
rs8102476	19	C	0.5393	1.09	19767754
rs11672691	19	G	0.7368	1.10	19318570, 23065704
rs2735839	19	G	0.8527	1.18	18264097
rs2427345	20	C	0.621	1.05	23535732
rs6062509	20	T	0.6983	1.08	23535732
rs58133635	22	T	0.197	1.07	19117981, 25217961
rs5759167	22	G	0.502	1.15	19767753
rs2405942	23	A	0.7833	1.05	23535732
rs5945619	23	C	0.364	1.11	18264097
rs2807031	23	C	0.182	1.06	25217961
rs5919432	23	T	0.8008	1.04	21743467
rs6625711	23	A	0.1763	1.01	25217961
rs4844289	23	G	0.384	1.04	25217961
SNPs discovered in non-European populations					
rs2055109	3	T	0.7643	1.02	22366784
rs12653946	5	T	0.4246	1.08	20676098
rs1983891	6	T	0.2773	1.09	20676098
rs339331	6	T	0.695	1.09	20676098
rs1512268	8	T	0.4296	1.14	20676098
rs817826	9	T	0.8552	1.00	23023329
rs2252004	10	A	0.1017	1.00	22366784
rs12791447	11	G	0.0747	1.05	26443449
rs1938781	11	G	0.2297	1.03	22366784
rs9600079	13	T	0.443	1.01	20676098
rs58262369	14	C	0.998	1.27	26443449
rs7210100	17	A	0.0001	1.34	21602798
rs103294	19	C	0.7812	1.00	23023329
rs75823044	13	T	0.022	1.55	Reference [22]
rs78554043	22	C	0.015	1.62	Reference [22]
SNPs discovered in multi-ancestry populations					

rs1775148	1	C	0.359	1.04	25217961
rs9443189	6	A	0.857	1.07	25217961
rs7153648	14	C	0.082	1.03	25217961
rs12051443	16	A	0.344	1.03	25217961
rs12480328	20	T	0.928	1.11	25217961
rs1041449	21	G	0.433	1.05	25217961
rs2238776	22	G	0.802	1.05	25217961

Table 2: SNPs associated with radiation toxicity

Locus-Nearest Gene(s)	SNP	Toxicity endpoint	OR (95% confidence interval)	Proposed mechanistic relationship	Reference
5q31.2 – *KDM3B*	rs17599026	Urinary frequency (2 years after RT)	3.12 (2.08–4.69)	*KDM3B* expressed in bladder tissue. Radiation may lead to bladder dysfunction due to effect on this gene.	[25]
5p15.2 – *DNAH5*	rs7720298	Decreased urine stream (2 years after RT)	2.71 (1.90–3.86)	*DNAH5* expressed in kidney and bladder tissue. Radiation may lead to bladder dysfunction due to effect on this gene.	[24]
2q24.1 - *TANC1*	rs264663	Overall toxicity, late toxicity	6.6 (2.2-19.6)	*TANC1* involved in repair of muscle damage	[26]
9p21.2 - *IFNK* and *MOB3B*	rs17779457 (one of 8 SNPs in a haplotype block)	Urinary symptoms e.g. incomplete emptying, intermittency, frequency	No OR published. Beta coefficient 2.4	*IFNK* a member of type 1 IFN family with a role in inflammatory response to radiation induced tissue damage. *MOB3B* essential for mitotic checkpoint regulation.	[27]
11q14.3 - *SLC36A4*	rs7120482 rs17630638	Rectal bleeding	3.1 (1.7-5.6) 2.9 (1.6-5.2	SLC36A4 encodes amino acid transporter needed for cellular proliferation.	[28]
Chr 2 - *FSHR*	rs2268363 (GWAS in African American men)	Erectile dysfunction	OR = 7.03; 95% CI 3.4–14.7	Involved in normal testis development. Disruption of FSH signalling associated with abnormal spermatogenesis and infertility.	[29]
10q26.3 - *GLRX3* 19q13.43 -	rs11017104 rs7245988	Erectile dysfunction	1.5 (0.7-3.0) 2.0 (0.9-4.4)	12 SNPs identified that lie near genes involved in biological activities of erectile	[30]

NLRP11	(two of 12 SNPs discovered in this GWAS)			function.	

Figure Legends

Figure 1. Prostate Cancer Risk Loci

Figure 1 demonstrates the distribution of loci associated with prostate cancer risk identified by GWAS studies (each red arrow represents an individual SNP)

Figure 2. Prostate Cancer Consortia

Figure 2 shows the consortia and studies investigating the role of genetic variants in prostate cancer. Large scale collaborative efforts and data sharing are needed to answer the various scientific and clinical questions related to prostate cancer genetics. Abbreviations and acronyms: BARCODE-1: The Use of Genetic Profiling to Guide Prostate Cancer Targeted Screening (IRAS ID147536; Research Ethics Number: 15/LO/1992); BPC3:Breast Prostate Cancer Cohort Consortium (http://epi.grants.cancer.gov/BPC3/); CAPS: Cancer in the Prostate in Sweden (http://ki.se/en/meb/cancer-of-the-prostate-in-sweden-caps); CGEMS: The Cancer Marker Susceptibility Projects (http://dceg.cancer.gov/research/how-we-study/genomicstudies/cgems-summary); ELLIPSE: Elucidating Loci Involved in Prostate Cancer (http://epi.grants.cancer.gov/gameon/); GENPET: An imaging study of FCH-PET-CT in men with prostate cancer and a BRCA gene mutation (IRAS ID 138894) (Research Ethics Number: 15/20/0242); GENPROS: Analysing outcomes after prostate cancer diagnosis and treatment in carriers of rare germline mutation in cancer predisposition genes. (NCT02705846); ICPCG: International Consortium of Prostate Cancer Genetics (www.icpcg.org); IMPACT: The Identification of Men with a Genetic Predisposition to Prostate Cancer: Targeted Screening in BRCA1/BRCA2 mutation carriers and controls (www.impact-study.co.uk); MADCaP: Men of African Decent and Prostate Cancer (http://epi.grants.cancer.gov/madcap); PRACTICAL: Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (http://practical.ccge.medschl.cam.ac.uk); PROFILE: Germline genetic profiling: Correlation with targeted prostate cancer screening and treatment (NCT02543905); RGC: Radiogenomics Consortium (https://epi.grants.cancer.gov/radiogenomics/).

Figure 3. Composition of the OncoArray BeadChip

Figure 3 displays the proportion of SNPs assigned to each cancer type. The OncoArray genotyping microarray comprises ~533,000 SNPs. SNPs related to genetic modifiers of *BRCA1/2* were also included as well as common cancer susceptibility loci.

Fig. 1

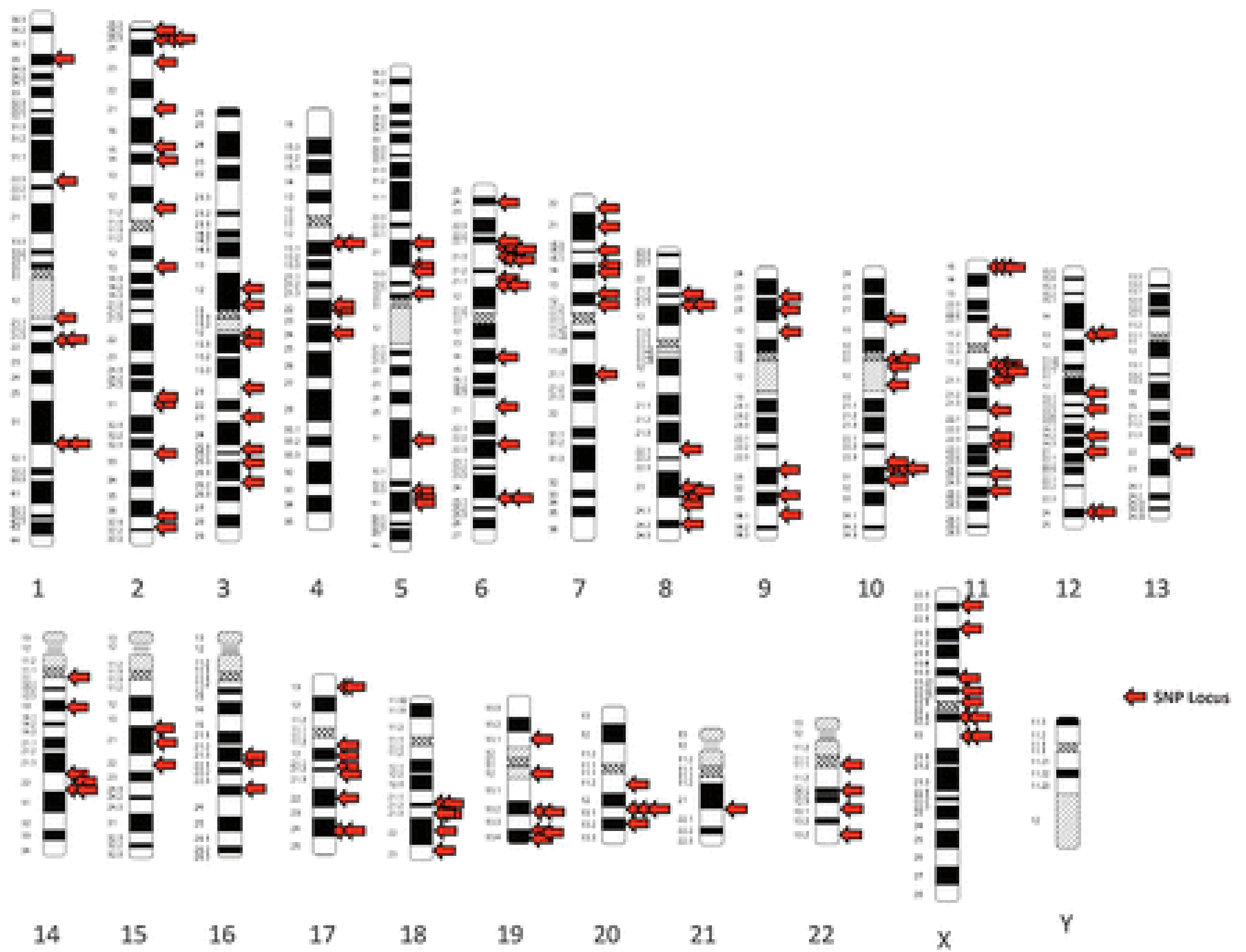

Fig. 2

Fig. 3