Genetics and genomics of prostate cancer

Michael Dean¹ and Hong Lou²

Prostate cancer (PCa) is one of the most common malignancies in the world with over 890,000 cases and over 258,000 deaths worldwide each year. Nearly all mortalities from PCa are due to metastatic disease, typically through tumors that evolve to be hormone-refractory or castrate-resistant. Despite intensive epidemiological study, there are few known environmental risk factors, and age and family history are the major determinants. However, there is extreme heterogeneity in PCa incidence worldwide, suggesting that major determining factors have not been described. Genome-wide association studies have been performed and a considerable number of significant, but low-risk loci have been identified. In addition, several groups have analyzed PCa by determination of genomic copy number, fusion gene generation and targeted resequencing of candidate genes, as well as exome and whole genome sequencing. These initial studies have examined both primary and metastatic tumors as well as murine xenografts and identified somatic alterations in TP53 and other potential driver genes, and the disturbance of androgen response and cell cycle pathways. It is hoped that continued characterization of risk factors as well as gene mutation and misregulation in tumors will aid in understanding, diagnosing and better treating PCa.

Keywords: androgen receptor; cancer progression; chromatin remodeling; metastasis; prostate cancer (PCa); risk factor; somatic mutation; tumor heterogeneity

PROSTATE CANCER (PCa)—GLOBAL IMPACT AND RISK FACTORS

PCa is the second most common cancer of men with 890,000 new cases (14% of all cancers) reported each year.¹,² However, the rate of PCa incidence varies by over 25-fold globally with Australia and New Zealand, Europe and Northern America having the highest incidence rates. Incidence is also high in populations of African descent, the Caribbean, South America and sub-Saharan Africa, and very low in South-Central Asia (4.1 per 100,000).

In 2008 (the last year for which global statistics are available), there was an estimated 258,000 deaths from PCa making it the sixth most common cause of death from cancer in men (Figure 1). The percentage of cancer mortalities in men due to PCa ranges from 0.6%–4.8% in Asia, to 9.7%–16% in North, Central and South America Europe and Australia (Supplementary Table 1). It is generally agreed that the widespread use of the prostate-specific antigen test in developed countries leads to a higher diagnosis and PCa represents 20%–30% of all male cancer diagnoses in these countries (Supplementary Table 1). However, in Asia, PCa represents only 1%–10% of male cancer cases, but the incidence is rising dramatically.³

PCa is a disease of aging with a majority of both cases and deaths occurring in men in their 70s (http://seer.cancer.gov/statfacts/html/prost.html).⁴ The extreme differences in incidence and mortality in Asia as compared to the rest of the world have prompted searches for dietary or environmental factors. For as breast cancer, Japanese have extremely low rates that rise if they are first-generation immigrants to the United States and yet further for subsequent generations, suggesting that there are indeed very strong environmental factors.⁵,⁶ Dietary fat and particular animal fats have been implicated with odds ratios as high as 3.6 for animal fat consumption in the most positive studies. Similar modest associations with body mass index and PCa have been found. Neither smoking or alcohol use show strong associations nor have studies of nutrient levels and supplementation yielded highly positive correlations. Androgens are clearly required as eunuchs and people with androgen deficiency have very low risk, but attempts to correlate testosterone and other androgen levels have not shown dramatic associations.

In conclusion, PCa is one of the few common cancers without a strong known lifestyle or infectious agent as a risk factor. The parallels to breast cancer are striking (i) gender-specific; (ii) hormonally driven; (iii) large geographic difference in incidence; and (iv) evidence for major lifestyle factors from immigration studies, but lack of clear association to specific dietary components.

FAMILY HISTORY AND FAMILIAL PCA

While family history is a risk factor, only 15% of men with the disease have a first-degree relative with PCa compared to 8% of the general population.⁷,⁸ A meta-analysis of multiple studies suggested that the risk is higher for men with a brother affected than a father, and that the familial risk is higher for early onset disease.⁹ In Sweden, a nation-wide registry of cases was used to document 3- to 10-fold higher risks in men with a family history and genetic factors account for 12% of the disease.¹⁰
Linkage studies have been performed on large collections of PCa pedigrees and several candidate familial loci have been identified. The most convincing locus is the \textit{HOXB13} gene, found in a region with a log-of-odds score of 5.5. A G84E mutation was identified in four of the linked pedigrees and this same mutation and others in the \textit{HOXB13} gene were identified in additional families.\textsuperscript{11–13} The G84E mutation is significantly associated with disease in men with a family history and/or early disease onset. The \textit{HOXB13} gene binds to the androgen receptor and plays a role in prostate development.

Other loci implicated in familial studies include HPC1 (chromosome 1q25) with the \textit{RNASEL} gene mutated in some families; PCAP (1q42-43); HPCX (Xq27-28); CAPB (1q36), with the \textit{EPHB2} gene implicated;\textsuperscript{14} and HPC20 (20q13). At least a dozen other loci have been implicated but no specific genes were identified and confirmed in these regions.\textsuperscript{15}

Several studies have demonstrated that in Ashkenazi Jewish populations, the common \textit{BRCA1} and \textit{BRCA2} mutations identified in breast and ovarian cancer families confer risk for PCa. The odds ratios range from 2.1 to 4.8 and generally reach statistical significance for \textit{BRCA2} but not \textit{BRCA1}.\textsuperscript{16–18} Additional studies in multiple populations support a role for \textit{BRCA2} and some studies for \textit{BRCA1} in early-onset, high-Gleason score disease and/or death by PCa.\textsuperscript{18}

**GENOME-WIDE ASSOCIATION STUDIES (GWAS)**

With the advent of microarrays capable of interrogating one million or more single nucleotide polymorphisms (SNPs), GWAS have been performed on most major cancer types including multiple studies on PCa.\textsuperscript{12,19–24} Over 33 loci have been identified that qualify as being genome-wide significant and replicated in more than one study, or internally in multiple validation sets (\textit{Supplementary Table 2}). The majority of these loci are in intergenic regions, many near known genes, and a few have known biological relevance to PCa. The odds ratios of these associations are modest (1.02–1.66) and so no one loci contributes highly to PCa risk.

The strongest associations are to a region on chromosome 8q24 about 500 kb 5' to the \textit{MYC} oncogene. Multiple SNPs in this region are associated, with association being different in different populations and multiple variants in this region are associated with other cancers.\textsuperscript{19,20,23–27} Despite some effort to directly link these variants to \textit{MYC} gene expression levels, no clear correlation could be found, but this region is known to interact at the chromatin level with the \textit{MYC} locus.\textsuperscript{26,27}

One of the strongest SNP associations is to the rs10993994 SNP on chromosome 10, in front of the \textit{MSMB} gene.\textsuperscript{21,23,28,29} \textit{MSMB} encodes beta-microseminoprotein, an immunoglobulin binding factor family protein produced by the epithelial cells of the prostate gland and secreted into the seminal plasma. The \textit{MSMB} protein has inhibin-like activity and may play a role as an autocrine paracrine factor in uterine, breast and other female reproductive tissues. The expression of the \textit{MSMB} protein is decreased in PCa, and antibodies to the protein are used as biomarkers of PCa recurrence and/or progression. Due to the obvious connection of this protein to PCa and to incident metastatic disease,\textsuperscript{30} considerable effort has been made in validating and understanding this result.

The rs10993994 SNP affects a predicted CREB transcription factor binding site\textsuperscript{31} and this activity was validated by functional studies.\textsuperscript{32} This association is reflected in the finding of reduced levels of expression of \textit{MSMB} protein in urine and prostate tissue in multiple populations.\textsuperscript{33–35} The SNP also shows association to PCa in Chinese populations and to \textit{MSMB} levels in serum. Fine-mapping and resequencing confirmed that this SNP is the most significantly associated and that there are no common mutations/variants in the \textit{MSMB} coding region.\textsuperscript{36} However, the \textit{MSMB} gene is also adjacent to the \textit{NCOA4} gene, a known coactivator of androgen receptors and Lou \textit{et al.}\textsuperscript{37} documented that there are hybrid transcripts expressed from the \textit{MSMB} promoter that contain the entire coding region of \textit{NCOA4}. These hybrid transcripts can produce a stable protein and are regulated by CREB binding to the rs10993994 site. However,
Fitzgerald et al. could find no association between SNPs in this region and overall NCOA4 expression and Rinckleb et al. presented data that there is an association of rs10993994 risk homozygosity and DNA repair capacity. Therefore, this PCA GWAS locus is very complex and the association signal may be a combination of very subtle effects on more than one gene/pathway.

Another loci with some functional validation is the rs12653946 polymorphism near the IRX4 (Iroquois homeobox 4) gene on chromosome 5p15. IRX4 is expressed in the prostate and heart, and there is a significant association between the genotype of rs12653946 and IRX4 expression in normal prostate tissues. Knockdown of the IRX4 gene enhances the growth of PCA cell lines and the protein interacts with the vitamin D receptor.

PCa has been studied by GWAS more thoroughly than any other cancer except perhaps breast. Multiple studies in many countries and in all major ethnic groups have been carried out and a group of associated loci with high confidence are now known. However, translating this knowledge into further understanding of PCA development and progression or in use as effective biomarkers will require considerably more functional and clinical studies.

SOMATIC ALTERATIONS IN TUMORS

In order to understand the somatic alterations that take place in the genomes of PCA tumors, several groups have begun sequencing exomes and genomes, carrying out methylation studies and performing RNA sequencing studies. Prior to undertaking this, there had been studies on mutations in prostate tumors of known oncogenes and tumor suppressors, revealing common mutations in genes such as TP53, PIK3CA and PTEN. The identification of high expression of ERG, an ETS-family transcription factor led to the identification of DNA deletions that result in the activation by gene fusion of ERG in many prostate tumors. While the most common lesion is a nearly 3-MB deletion of chromosome 21 resulting in the fusion of TMPRSS2 and ERG, other deletions and translocations have been described resulting in fusions of ERG or other ETS-family genes (ETV1, ETV4 and ETV5) resulting in overexpression of a hybrid ETS-family protein (reviewed in Clark and Cooper). These lesions can be found in up to half of all PCAs in people of European descent; however, there has been little advance in using these lesions as biomarkers for PCA therapy or recurrence, and the function of the ERG protein in the tumor is poorly understood (reviewed in Rosen et al.).

Three studies have been published to date with genome/exome sequence of more than 10 prostate tumors. One challenge with PCA is the high proportion of normal cells and heterogeneity in the primary tumor. To address this, some groups have either passed xenografts in mice or sequenced metastatic tumors. Kumar et al. performed exome capture and sequenced xenografts from 16 lethal metastatic tumors and three primary tumors. They did not have corresponding normal DNA, but used filters to identify the most likely somatic variants and identified recurrent mutations in the TP53, DLK2, GPC6, SFD4 and 19 other genes. Three of these tumors exhibited a dramatic increase in the number of mutations (2500–4000) as compared to the other 20 samples that had 362 ± 147 mutations. The molecular basis for this hyper-mutability phenotype was not determined, but has been seen in other tumors.

Barbieri et al. performed exome capture and sequenced 112 treatment naïve prostate adenocarcinomas and matched normal samples from the United States and Australia. They identified mutations in the known genes, PIK3CA, TP53 and PTEN as well as FOXA1, MEDI2, THSD7B, SCN11A and ZNF595, and CDKN1B was mutated in three and deleted in 16 tumors, indicating that this gene is another common driver. The mutations in the FOXA1 gene were missense variants clustered in the Forkhead domain. FOXA1 regulates androgen receptor-driven transcription providing a potential mechanism for these mutations to influence PCA development. MED12 encodes a subunit of the cyclin-dependent kinase 8 (CDK8) and MED12 mutations are found in up to 70% of uterine leiomyomas. The mutations found in PCA are clustered in the active site. The SPOP gene had been previously shown to be mutated in PCAs and was sequenced in an additional 300 primary tumors and metastases from the United States and Europe and recurrent heterozygous SPOP substitutions were identified in 6%–13% of primary tumors and in 15% of subjects with metastatic disease. These SPOP mutations affect conserved residues in the putative substrate-binding cleft and these altered residues (Tyr87, Trp131 and Phe133) have key roles in substrate interaction, and knockdown of SPOP increases invasiveness of PCA cell lines.

Grasso et al. sequenced the exomes of 50 lethal, metastatic castrate-resistant PCAs and 11 treatment-naïve, high-grade localized tumors. They identified mutations in the known genes TP53, AR, ZFHX3, RB1, PTEN and APC and described three additional significantly mutated genes: MLL2, OR5L1 and CDK12. Through integration of the mutation and copy number data, they implicated the WNT and PTEN pathways as being altered in a large number of tumors. Grasso et al. also described a frequent copy number loss on chromosome 5q21 at the location of CHD1, a gene encoding a chromatin-remodeling enzyme, and documented loss or mutation in multiple chromatin-remodeling genes, as has been found for multiple tumor types. They documented that multiple of these chromatin-modifying genes directly interact with the androgen receptor and along with additional regulators, such as FOXA1, can partially explain the androgen-resistant nature of the tumors they sequenced.

Supplementary Table 3 summarizes the results from all three of these studies. Only two genes were found mutated in all three studies, TP53 and MLL2, with PTEN significantly mutated in two studies. The SPOP, FOXA1, CDKN1B and MED12 genes described above are mutated in two of the three studies. However, there is clearly a great diversity of genes with many genes significantly mutated in one study and not detected in the other three. It is true that the designs of the studies were quite different with xenograft, primary tumors and metastatic tumors being the principle source of DNA for the studies, respectively. Thus part of the difference may be due to the starting material selected. Sensitivity of detection can also be an issue, especially in the Barbieri et al. study using primary tumors as the contamination of normal cells and the heterogeneity of the primary tumor may mask the ability to detect some mutations.

Clearly the mutational landscape of PCA is extremely complex, and the several studies that have also integrated expression and copy number analysis into the picture introduce other complications. The finding of multiple chromatin-remodeling genes mutated and connecting these mutations and that of other genes, such as FOXA1, in hormone signaling is an important advance, as nearly all lethal PCAs are castrate-resistant. However, relatively few genes have emerged from these studies that can be targeted by currently available directed compounds, and the availability of metastatic tumor tissue is very limited in PCa, creating a problem for an individualized medicine strategy. While circulating tumor cells could provide a solution to this problem, it remains to be seen if these cells represent the actual cells with metastatic capacity.

The current studies were all performed on European populations, and will need to be repeated in people of African descent (where
The serum prostate-specific antigen test is the most common diagnostic tool to identify early lesions and is widely used in many countries. However, the test results in up to 80% false-positive results, leading to a high number of prostate biopsies and additional tests that are unneeded. More concerning is the estimate of 17%–50% overdiagnosis (diagnosis of men who will remain asymptomatic during their lifetime). Nearly 90% of these men receive either radiation therapy, surgery or androgen deprivation therapy, all of which have modest benefit and are unneeded. More concerning is the estimate of 17%–50% overdiagnosis during their lifetime. Nearly 90% of these men receive either radiation therapy, surgery or androgen deprivation therapy, all of which have modest benefit and are unneeded.

The apparent large influence of genetic factors has influenced the undertaking of a large number of GWAS in multiple populations. This has led to the discovery of a large number of loci highly significantly associated. The most directly relevant is the rs10993994 SNP in the promoter region of MSMB that regulates promoter activity, expression and serum and urine levels of this important marker. However, clinical application of this result is uncertain. And although the constellation of markers can be used to estimate the risk of a given man, few of these markers are informative for outcome.

The best hope for improving our knowledge of this disease is comprehensive molecular characterization of the tumor. This has begun to be carried out and the first studies have uncovered new genes mutated, amplified, deleted and/or activated in PCa. The studies of castrate-resistant tumors have shed new light on genes implicated in androgen-responsiveness such as MLL2 and other chromatin-remodeling proteins and the FOXA1 transcription factor. However, the sequencing studies of primary tumors, and those of metastatic or xenografts are very different, suggesting that there is considerable heterogeneity and that a sample of the metastatic tumor is needed to identify the major cancer driver genes. The difficulty in accessing metastatic material in patients makes the application of personalized treatment difficult and the lack of targeted agents for the genes altered in PCa needs to be addressed.

The next generation of molecular studies of PCa need to address, not only exome and genome sequencing, but copy number analysis, epigenome analysis and expression and chromatin studies. As in all solid tumors, generating, annotating and combining these results in a meaningful fashion requires large numbers of samples, properly collected and annotated, but this approach will hopefully be informative for reducing the burden of the disease.

CONCLUSIONS AND FUTURE PERSPECTIVES

The high incidence of PCa in much of the world and the very poor prognosis for metastatic disease has placed a high priority on understanding the causes and risk factors, developing effective prevention and detection strategies as well as effective treatments. While immigration studies and the rapidly rising incidence in parts of Asia support a key role of lifestyle factors, identifying those factors has been challenging. Some studies support an influence of ‘Western diet’ and both high fat content and animal protein have been implicated, but either proving this or changing diet sufficiently is unlikely to occur in the near term.

The high incidence of PCa in much of the world and the very poor prognosis for metastatic disease has placed a high priority on understanding the causes and risk factors, developing effective prevention and detection strategies as well as effective treatments. While immigration studies and the rapidly rising incidence in parts of Asia support a key role of lifestyle factors, identifying those factors has been challenging. Some studies support an influence of ‘Western diet’ and both high fat content and animal protein have been implicated, but either proving this or changing diet sufficiently is unlikely to occur in the near term.

COMPETING FINANCIAL INTERESTS

All authors declare that there are no competing financial interests.

ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of the NIH, the National Cancer Institute Center for Cancer Research.

Supplementary Information accompanies the paper on Asian Journal of Andrology’s website (http://www.nature.com/aja).


