Prostate cancer genomics by high-throughput technologies: genome-wide association study and sequencing analysis

Hidewaki Nakagawa
Laboratory for Genome Sequencing Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

Abstract

Prostate cancer (PC) is the most common malignancy in males. It is evident that genetic factors at both germline and somatic levels play critical roles in prostate carcinogenesis. Recently, genome-wide association studies (GWAS) by high-throughput genotyping technology have identified more than 70 germline variants of various genes or chromosome loci that are significantly associated with PC susceptibility. They include multiple 8q24 loci, prostate-specific genes, and metabolism-related genes. Somatic alterations in PC genomes have been explored by high-throughput sequencing technologies such as whole-genome sequencing and RNA sequencing, which have identified a variety of androgen-responsive events and fusion transcripts represented by E26 transformation-specific (ETS) gene fusions. Recent innovations in high-throughput genomic technologies have enabled us to analyze PC genomics more comprehensively, more precisely, and on a larger scale in multiple ethnic groups to increase our understanding of PC genomics and biology in germline and somatic studies, which can ultimately lead to personalized medicine for PC diagnosis, prevention, and therapy. However, these data indicate that the PC genome is more complex and heterogeneous than we expected from GWAS and sequencing analyses.

Key Words

- prostate
- molecular genetics

Introduction

Prostate cancer (PC) is the most common malignancy in men and the second leading cause of cancer-related deaths in Western countries. Incidence and mortality rates vary widely across geographic regions and ethnic groups (Gronberg 2003). In particular, Asian populations have a substantially lower incidence rate than Caucasians or African Americans, indicating the contribution of different genetic backgrounds to PC susceptibility (Gronberg 2003, Schaid 2004, Nakagawa et al. 2012). Although the precise mechanisms of prostate carcinogenesis have not been fully elucidated, it is evident that genetic factors play important roles in PC etiology. A positive family history of PC has been recognized as one of the most important risk factors for PC, as well as African ethnicity and older age (Schaid 2004); twin studies have suggested that the contribution of genetic factors to the development of PC is larger than in other types of common human tumors (Lichtenstein et al. 2001). Over the past 20 years, genetic research by linkage analysis on several hereditary PC families had been conducted to clarify PC genomics and to identify genes responsible for PC susceptibility (Schaid 2004). This strategy identified
some candidate genes predisposing to hereditary PC, but they were not replicated in other studies, and this approach was not successful for understanding PC genomics. However, recent genome-wide association studies (GWAS) using high-throughput genotyping technologies on several thousand samples of several ethnic groups have successfully identified more than 70 single-nucleotide polymorphisms (SNPs) of various genes or chromosome loci that are known to be significantly associated with PC susceptibility (Amundadottir et al. 2006, Gudmundsson et al. 2007, Haiman et al. 2007, Eeles et al. 2008, 2013, Nakagawa et al. 2012).

On the other hand, somatic events in the development and progression of PC have been recently explored more comprehensively through genomic technologies. NXX3.1 (NXX3-1) loss and PTEN loss or point mutations were frequently observed in advanced PCs. The androgen/androgen receptor (AR) signaling pathway plays a central role in PC development and progression, and PC growth is usually androgen-dependent. Hence, most of the patients with relapsed or advanced disease respond well to androgen-ablation therapy (castration). Nonetheless, they eventually acquire castration resistance (CR) and progress to a more aggressive phenotype (Scher & Sawyers 2005). In this phase, AR amplification (10–20%) or AR point mutations (<10%) are frequently observed (Taplin et al. 1995, Bubendorf et al. 1999), indicating that somatic AR alterations play a critical role in CRPC progression (Scher & Sawyers 2005). Furthermore, a new link was established in 2005 between the androgen signaling pathway and PC through the discovery of a genetic rearrangement that drives the fusion of two genes: the androgen-regulated gene TMPRSS2 and the E26 transformation-specific (ETS) transcription factor ERG (Tomlins et al. 2005).

Cancer is essentially a disease of the genome that develops and evolves with the accumulation of a variety of somatic mutations with the background of germline variants, and these genomic alterations have now been targeted for cancer treatment and diagnosis. Recent explosive advances of high-throughput genotyping and high-throughput sequencing technologies (next-generation sequencing (NGS)) with bioinformatics approaches enable us to comprehensively analyze a number of cancer genomes. GWAS, whole-genome sequencing (WGS), whole exome sequencing (WES), and RNA sequencing (RNA-Seq) have now been conducted for many types of cancer genomes worldwide, including the International Cancer Genome Consortium (ICGC; Hudson et al. 2010) and the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL Consortium; Kote-Jarai et al. 2008, Eeles et al. 2013), to explore cancer genomic alterations and diversity. The findings from GWAS using high-throughput genotyping contribute to a better understanding of prostate carcinogenesis and potential application for PC risk prediction and prevention, whereas the findings from somatic mutation analysis, such as WGS and RNA-Seq, may contribute to the discovery of novel therapeutic targets and biomarkers for PC and molecular sub-classifications of PC (Fig. 1). This genomic research could ultimately lead to personalized medicine for PC diagnosis, treatment, and prevention. This review describes how recent high-throughput technologies have been applied to PC genomics research, in germline and somatic studies, and what has been achieved to date.

**GWAS by high-throughput genotyping to identify germline PC susceptibility loci**

GWAS is typically based on a case–control design in which around a hundred thousand SNPs across the human genome are genotyped, and it can scan for germline variants that are associated with disease risk. Such associations are consistent with the ‘common disease–common variant’ hypothesis that assumes that genetic influences on susceptibility to common diseases such as PC are attributable to some number of common variants present in more than 5% of the population (Manolio 2010). GWAS for PC have been remarkably successful in identifying more than 70 common genetic variants or loci, which is quite high compared with the corresponding numbers for other types of cancers and other common diseases. A Manhattan plot of a PC GWAS (Fig. 1) typically indicates many peaks of $-\log(P)$ value across the genome, including an extremely high peak (low $P$ value for the association) in chromosome $8q24$. Subsequent replication and follow-up studies have identified additional associated variants that show genome-wide significance with $P$ values $<5\times10^{-8}$ after adjustment for multiple testing. Recent meta-analyses combining several GWAS cohorts in several ethnic groups (more than 50 000 samples) identified additional PC-susceptibility genes or variants (Eeles et al. 2009, 2013, Kote-Jarai et al. 2011); larger meta-analysis are now being conducted worldwide and are expected to identify additional PC-susceptibility loci and reach more than 100 PC-susceptibility loci in the near future. As PC incidence and heritability remarkably differ for each race, it is important to perform meta-GWAS on several ethnic
or racial groups such as Asian and African populations (Takata et al. 2010, Haiman et al. 2011). However, many of the identified PC-susceptibility variants are located in the intergenic regions, as observed for other disease GWAS, and there are variants for which the biological significance has been proved to be associated with PC. The function or biological significance of only a few variants or genes has been elucidated so far. The greatest challenge of the GWAS loci for prostate genomics is to understand the functional consequences of these many PC-associated loci and to attempt to further understand PC biology and prostate carcinogenesis (Freedman et al. 2011).

Identification of additional PC-susceptibility loci in larger multi-ethnic GWAS cohorts and biological insights from such functional analysis on GWAS loci can be translated into clinical benefits, including reliable risk-prediction biomarkers and effective strategies for screening and prevention of PC.

Figure 1
Recent innovations in high-throughput genomic technologies enable us to analyze prostate cancer genomics more comprehensively, more precisely, and on a larger scale. High-throughput genotyping technologies can genotype 60–100 K SNPs and are applicable for GWAS on 10 000 samples. Next-generation sequencing (NGS) can analyze whole genomes (3 Gb) and the whole transcriptome on dozens of samples to identify point mutations, structural changes, and fusion genes comprehensively. Rare germline variants discovered by NGS have also been analyzed by high-throughput genotyping technologies for GWAS targeting of rare variants.

Chromosome 8q24 region
Many GWAS in multiple ethnic groups reported strong and consistent association of multiple variants at chromosome 8q24 with PC susceptibility (Gudmundsson et al. 2007, Haiman et al. 2007, Yeager et al. 2007, Al Olama et al. 2009). This region contains various independent PC-susceptibility loci within an ~1 Mb segment, and some of them were found to be significantly associated with other types of cancers such as colon cancer and breast cancer (Easton et al. 2007, Tomlinson et al. 2007). However, no gene has been annotated in this ~1 Mb region, and its biological significance in cancer remains unclear. There are at least five separate loci (Region1–3 or Block1–5) associated with PC susceptibility in this region (Ghoussaini et al. 2008). The MYC proto-oncogene is located at ~200 kb downstream, and recent studies have indicated that one of the loci at 8q24 (rs6983267
represents Region1/Block5) could be associated with the WNT signaling pathway (Tuupanen et al. 2009) in colorectal cancer and with MYC expression in several cancers (Pomerantz et al. 2009). The association of its most centromeric region (Region2: Chr8: 128.14–128.28 Mb) with PC is stronger in Japanese and African populations, and this region was transcribed as an ~10 kb intron-less long non-coding RNA (lncRNA). The expression of this lncRNA was upregulated in PC cells and its precursor lesions, and it is likely to be involved with transcriptional AR activity (Chung et al. 2011). Another study reported that a variant on Region2 was correlated with serum testosterone levels (Chu et al. 2010), and because Region2 is associated only with PC, not with breast cancer and others, it is likely to be specifically involved with PC through the androgen/AR signaling pathway.

Prostate-specific genes

Several GWAS of PC identified a variant (rs2735839) of KLK3 at chromosome 19q13, encoding prostate-specific antigen (PSA) as an indicator of PC susceptibility (Eeles et al. 2008). However, several groups observed an association between rs2735839 and serum PSA levels (Gudmundsson et al. 2010), and this genetic variant may be merely associated with the likelihood of diagnosis by virtue of its link with serum PSA levels. A SNP at chromosome 10q11 (rs10993994) was validated to be associated with several GWAS in some populations (Eeles et al. 2008, Thomas et al. 2008). It is located at the 5′ region of the MSMB gene, encoding β-microseminoprotein. Fine mapping and functional analysis demonstrated that this SNP could influence the transcriptional activity of the promoter of MSMB (Chang et al. 2009). MSMB is synthesized by epithelial cells in the prostate gland and secreted into the seminal plasma, and its expression progressively decreased during development of PC from early to late stages, suggesting its potential effect as a tumor suppressor (Beke et al. 2007). Some GWAS showed the consistent association of rs1512268 at chromosome 8p21 with PC susceptibility, which represents a 57 kb genetic region in which the NKX3.1 gene is solely annotated (Eeles et al. 2009, Takata et al. 2010). NKX3.1 is exclusively expressed in the prostate tissue, and variants in the 5′-UTR of NKX3.1 regulate the proximal promoter activity of the NKX3.1 gene (Akamatsu et al. 2010). NKX3.1 is an androgen-regulated homeobox gene, which plays a key role in the regulation of growth and differentiation of prostate epithelium in a normal prostate and marks a stem cell population that functions during prostate regeneration (Wang et al. 2009). Iroquois homeobox 4 (IRX4) at chromosome 5p15 was also identified as a PC-susceptibility gene by GWAS (Nguyen et al. 2012), and it is specifically expressed in the adult prostate and heart, suggesting its involvement with the development of the prostate and heart. The associated variants are located at the 5′ region of the IRX4 gene, and regulate the transcriptional activity of IRX4 through direct interaction with the vitamin D pathway (Nguyen et al. 2012). GWAS have demonstrated that these prostate-specific genes are strongly associated with prostate carcinogenesis as well as prostate development.

Metabolism-related genes

It is interesting that GWAS for PC identified several loci or SNPs that were validated to be inversely associated with type 2 diabetes (T2D). Epidemiological studies suggest that men with T2D are less likely to develop PC than non-diabetic men (Bonovas et al. 2004). GWAS identified variants of the hepatocyte nuclear factor 1 homeobox B gene (HNF1B), which were also related to T2D predisposition (Sun et al. 2008). These variants were shown to be associated with decreased risk of PC. HNF1B encodes a transcription factor, which plays a key role in the development and function of the pancreas and kidney by regulating the expression of numerous genes in these tissues (Pontoglio 2000). HNF1B was previously shown to be mutated in individuals with maturity-onset diabetes of the young type 5 (MODY5; Lindner et al. 1999). Interestingly, this variant at HNF1B was also found to be associated with the risk of endometrial cancer in women, which is also thought to be dependent on some metabolic and hormone factors, as well as PC in men (Spurdle et al. 2011). GWAS showed that variants of juxtaposed with another zinc finger 1 (JAZF1) were also significantly associated with T2D and associated with decreased risk of PC (Thomas et al. 2008). JAZF1 encodes a nuclear protein with three zinc fingers and functions as a transcriptional repressor and it may be involved in lipid metabolism in the liver and adipocytes (Li et al. 2011). JAZF1 variants were also shown to be associated with human height, supporting a role for this gene in the regulation of growth and metabolism (Soranzo et al. 2009). Thus, it appears possible that genetic variations in JAZF1 and HNF1B may influence PC risk by changing the levels of hormones or growth-related factors previously suggested to be related to T2D. A GWAS identified that rs339331 at 6q22 was significantly associated with PC susceptibility (Takata et al. 2010), and this locus includes
the G protein-coupled receptor, family C, group 6, member A gene (GPRC6A). Gprc6a-null mice exhibited a metabolic syndrome characterized by impaired bone mineralization, increased fat mass, fatty liver, glucose intolerance, testicular feminization, and abnormal steroidogenesis (Pi et al. 2008), indicating that it can affect PC susceptibility by altering GPRC6A-mediated sex hormone production and metabolic pathways. Diet and lifestyle are now proposed to be critical risk factors for PC development, and indeed, in Asia, there is rapidly increasing incidence of PC partially due to the prevalence of a Western lifestyle, which includes high consumption of a high-fat diet. These interesting observations illustrate a biological phenomenon connecting metabolism and prostate carcinogenesis.

**Inflammation-related gene**

Japanese GWAS on PC identified a PC-susceptibility locus at 19q13.4 (rs103294), and this SNP is located in the FOXP4 gene (Takata et al. 2010). The FOXP family of forkhead transcription factors are essential for normal T-cell lineage development. A new member of the FOXP family, FOXP4, has also been reported to be dispensable for T-cell development, and it may be involved in the immune response to pathogen infection (Wiehagen et al. 2012). GWAS on Chinese PCs identified a PC-susceptibility locus at 19q13.4 (rs103294), and this SNP was tightly linked with a 6.7 kb deletion that removed the first six exons of leukocyte immunoglobulin-like receptor, subfamily A, member 3 (LILRA3) and induced the loss of its expression in T cells (Xu et al. 2012). The LIR family members, including LILRA3, bind to major histocompatibility complex antigens and regulate the immune and inflammatory response. Although the role of LILRA3 in prostate carcinogenesis is largely unknown, these findings suggest a potential role in chronic inflammation or the immune response in prostate carcinogenesis.

**Variants associated with aggressive PCs**

It remains unclear whether these genetic variants are associated with PC aggressiveness. PC is primarily a disease associated with old age, and many indolent PC cases in elderly people are subject to a watch-and-wait strategy without any PC treatment (Albertsen 2010). The rate of latent PC in autopsies may be as high as 80% by the age of 80 years (Konety et al. 2005), and PC is a very ‘common’ disease in older men. Hence, distinguishing between indolent PC and aggressive PC, which requires treatment, including medical castration, surgery, and radiation therapy, and predicting the risk of aggressive PC are very important current goals in PC research and clinics. Recent meta-analysis of GWAS showed that rs4054823 at 17p12 is associated with aggressive disease of PC, but there is no annotated gene around this SNP and no explanation for its biological association with PC aggressiveness (Xu et al. 2010). Another meta-analysis identified rs11672691 at 19q13 as being significantly associated with aggressive PCs, but this association was also observed in indolent PCs, and the responsible genes and biological significance of this locus was not clear (Al Olama et al. 2013). To date, most of the PC-susceptibility loci or SNPs have not been significantly associated with aggressive PC disease specifically.

**PC risk estimation by multiple SNPs**

GWAS have identified more than 50 variants or loci significantly associated with PC risk, but each of them has limited use in the assessment of PC risk in an individual because each of these genetic markers confers a modest effect (OR: 1.1–2.0). It is required for clinical use of these risk variants to combine individual variants for PC risk assessment (Manolio 2010). Zheng et al. (2008) combined five variants for PC risk assessment plus a family history of PC and confirmed a cumulative and significant association with PC. However, several studies using five or more SNPs have since indicated that the area under the curve (AUC) in the ROC curve was around 0.6, and this SNP panel may have limited clinical utility (Nam et al. 2009, Zheng et al. 2009). Even when the sample number was increased, the AUC was predicted to be <0.7 (Chatterjee et al. 2013). The AUC of the PSA test was found to be 0.65–0.7; to apply the risk estimation model for PC screening, the AUC should be larger than that of PSA. Otherwise, the risk estimation model should focus on individuals with gray-zone or normal-range PSA levels, in which the efficiency of PSA testing to detect PC is questionable (Akamatsu et al. 2012). And also it is important to construct a SNP prediction model for each ethnic group because the frequency and OR of each SNP is different in each ethnic group.

**Rare variants associated with PC susceptibility**

GWAS have thus far focused on common SNPs, but most of these disease-associated SNPs have a very small effect. Even when multiple variants are combined to predict PC risk, they can explain only 10–20% of the genetic risk of
the given disease, indicating missing heritability (Cirulli & Goldstein 2010). Furthermore, GWAS signals have rarely been tracked to causal variants. Although it is still not easy to identify rare variants that are assumed to have a larger effect on susceptibility to common diseases (Cirulli & Goldstein 2010), a few rare variants have been identified that confer a substantial disease risk. Several linkage studies on hereditary PCs and early-onset PCs identified the chromosome 7q21–22 region as one of the most critical regions responsible for hereditary PCs. There are 202 genes in this critical region and NGS analysis identified rare variants of the HOXB13 gene, which is highly expressed in the prostate and is likely to be involved with the development of the prostate or urological system. While a rare variant of HOXB13 G84E was observed in 0.1–0.2% of the control population, the patients of hereditary PC or early-onset PC had 1–3% with an OR of 2.7–5.1 (Ewing et al. 2012). WGS in an Icelandic population that identified a rare variant in the PC critical region at chromosome 8q24, whose frequency was 0.5% in the control population, which is independent of the reported SNPs or loci at chromosome 8q24 (Gudmundsson et al. 2012). Now next-generation GWAS are focusing on rare variants. WGS or WES can identify rare variants (frequency ~0.1 or 0.5%) in thousands of the affected population, and these variants on custom SNP chips can be screened for GWAS on ~10,000 case controls. Owing to their low frequency, these analyses are required to scan a large number of samples to produce sufficient statistical power. Hence, it is critical to collect such a large sample size in the bio-banks or consortium worldwide.

**Somatic mutation analysis by NGS technology**

There is increasing evidence that the somatic mutation burden in an individual’s cancer drives tumor formation, influences disease progression, and affects sensitivity to therapy (Fig. 2). Common somatic mutations in PC include loss of NKX3.1 and PTEN, and AR amplification or point mutations. NKX3.1 at chromosome 8p21 is frequently subject to loss of heterozygosity (LOH) in human PCs, and it has been recognized to function as a tumor suppressor of prostate (Bova et al. 1993), although somatic point mutations of NKX3.1 were not frequently detected in PCs (Voeller et al. 1997). PTEN at chromosome 10q23 is also frequently subject to LOH or homozygous deletion in many types of human cancers, and it is inactivated by point mutations in 20–40% of PCs (Li et al. 1997). Pten+/− mice have prostatic hyperplasia and dysplasia; prostatic intraepithelial neoplasia (PIN) develops in Pten+/−/Nkx3.1+/− and Pten+/−/Nkx3.1−/− mice (Kim et al. 2002), which indicate their tumor suppressor feature in prostate. In 2005, fusion transcripts of ETS family transcription factors with androgen-responsive promoters were discovered, which were driven by a genomic rearrangement (Tomlins et al. 2005). The TMPRSS2-ERG gene fusion and other ETS fusion genes have been causally linked to cancer progression because it promotes invasion, and overexpression of the fusion product in mice shows greatly enhanced PC development (Carver et al. 2009). Now recent rapid advances in sequencing technology (NGS) have enabled us to perform more comprehensive analysis such as WGS for human genome and cancer genome for $5000 in a few days, and its cost is expected to become less than $1000 in a few years. WGS can comprehensively detect almost all single-nucleotide variants, copy number variants, structure changes, and foreign genomes such as integrated virus (Fig. 1). Some projects in ICGC and other groups are analyzing PC genomes by WGS or WES integrated with RNA-Seq and DNA methylation analysis (http://www.icgc.org/) now. It is also required to examine somatic mutations of PC in different ethnic groups (African and Asian PCs) as well as germline variants because PC genomics and its biological behavior show some ethnic differences.

**WGS and WES of PC**

Initial WGS analysis on seven PCs with a Gleason grade 7 or higher (Berger et al. 2011) revealed that a number of genomic rearrangements were present in PC genomes and their breakpoints were enriched near open chromatin, AR, and ERG DNA binding site, leading to several types of ETS gene fusions. Other genomic rearrangements were observed to affect PTEN and PTEN-interacting proteins, indicating the effect of the PTEN pathway on prostate carcinogenesis. On average, PCs harbored 13–40 non-silent coding somatic mutations and ~1.0 mutations per Mb. This number is relatively low compared with the mutation numbers of other types of solid tumors. WES on 23 PC xenografts in mice (Kumat et al. 2011) was reported, and this study identified a number of somatic mutations, because they did not analyze the normal control genome and their data were likely to be contaminated by mouse genome sequences. WES on 112 PC samples (Barbieri et al. 2012) identified recurrent mutations in multiple genes, including SPOP, MED12, and FOXA1. SPOP mutations were identified in 6–13% of PCs, and it is likely to be involved in cell invasion. It is interesting that SPOP mutations occurred mutually exclusively in PC with ERG
rearrangements, and it was also detected in PC precursor, high-grade PIN, suggesting that SPOP mutations comprise an early event in prostate carcinogenesis as well as ERG rearrangements. Another WES analysis was performed focusing on 50 CRPCs (Grasso et al. 2012). This study identified recurrent mutations of chromatin regulators, including MLL2, CHD1, and FOXA1. These mutated genes are likely to be involved with androgen/AR signaling and confer deregulated AR signaling and a CR phenotype. WGS on 11 early-onset PCs was also conducted, which showed age-related differences in structure rearrangement formations. Compared with elderly-onset PC, these early-onset PCs demonstrated enrichment for androgen-driven structural rearrangements involving ETS family genes (Weischenfeldt et al. 2013), suggesting that early-onset PC development is predominantly driven by the androgen and AR signaling pathway.

RNA-Seq and fusion genes in PC

RNA-Seq (whole transcriptome sequencing) can provide not only a gene expression profile but also information about splicing changes, non-annotated ncRNA expression, and RNA editing. Most importantly, for PC genomics, RNA-Seq analysis can detect many types of recurrent fusion genes. ETS family gene fusions, including TMPRSS2-ERG, were discovered in 2005, and most PCs are characterized by the presence of recurrent gene fusions primarily involving androgen-regulated upstream genes, TMPRSS2, KLK2, and SLC45A3, fused to one of the genes of the ETS family of oncogenic transcription factors (ERG, ETV1, ETV4, or ETV5) (Maher et al. 2009). These fusion events are usually driven by genomic rearrangements in PC genomes. Recent RNA-Seq analysis detected one read-through chimera, SLC45A3-ELK4, between the fourth exon of SLC45A3 with exon 2 of ELK4, a member of the ETS transcription factor family, and this fusion event was not driven by genomic rearrangements, but cis-splicing between the two adjacent genes (Maher et al. 2009, Zhang et al. 2012). In addition to the ETS family gene fusions, RNA-Seq analysis on 25 PC samples discovered fusion transcripts involving non-ETS genes such as CDKN1A (p21), CD9, and IKBKB (IKKβ), genes known to exhibit key biological roles in cellular homeostasis, as well as the oncogene PIGU and the tumor suppressor gene RSRC2 (Pflueger et al. 2011). Another RNA-Seq analysis on 14 Chinese PCs uncovered a recurrent fusion gene between USP9Y and TTY15 on chromosome Y, and a fusion event that arose from an interchromosomal translocation involving CTAGES and KHDRBS3 (Ren et al. 2012). A recent study by RNA-Seq discovered RAF kinase gene fusions: SLC45A3-BRAF, ESRP1-RAF1, and RAF1-ESRP1 in advanced PCs. Although they were detected in only 1–2% of PCs, RAF kinase fusions represent the first ‘actionable’ fusions in PCs that do not involve an ETS family member (Palanisamy et al. 2010). RNA-Seq
analysis can detect non-annotated transcripts, different from the microarrays, and systematic RNA-Seq analysis on 102 PC samples discovered more than 100 novel PC-specific ncRNAs and demonstrated that a novel lncRNA PCAT1 is a transcriptional repressor implicated in a subset of PC patients (Prensner et al. 2011).

**Neuroendocrine PC genome**

Neuroendocrine PC (NEPC) is a rare subtype of PC affecting only 0.5–2% of patients with PC, but NEPCs more frequently arise after castration therapy for prostate adenocarcinoma, although it can arise de novo (Palmgren et al. 2007). Focal neuroendocrine differentiation is detected in some 10% of localized PCs and increases with disease progression, indicating that these clones are evolved to a NEPC phenotype after castration (Beltran et al. 2011). New AR inhibitors and blockers of androgen production are now available, and ‘complete’ castration will be established and be a standard therapy for recurrent and metastatic CRPCs (Haddad & García 2012). Hence, the number of patients with the NEPC phenotype will increase (Fig. 2), and more research and clinical focus will be put into the study of NEPCs. TMPRSS2-ERG has been reported in ~50% of NEPCs, suggesting that NEPC is clonally derived and evolved from PC (Lotan et al. 2011). RNA-Seq analysis of seven NEPCs revealed overexpression and amplification of AURKA and MYCN in 40% of NEPCs and 5% of PCs (Beltran et al. 2011).

**Conclusions**

Recent innovation of high-throughput genomic technology has enabled us to analyze PC genomics more comprehensively, more precisely, and on a larger scale. Based on this valuable information, our understanding of PC genomics and biology is increasing. Germline research can help to establish the risk prediction and prevention of PC, while somatic research can provide information on novel molecular targets and biomarkers and a rationale for molecular sub-classification. The final goal is personalized medicine for PC diagnosis, treatment, and prevention (Fig. 1). However, through GWAS and sequencing analysis, we have recognized more complexity and heterogeneity in the human PC genome than expected. Therefore, to gain further new insights into PC genomics, we have to analyze a larger sample size in each ethnic or racial group. It is still too early and primitive to translate these high-throughput data into the clinic because the majority of these genomic data lack biological significance; it is difficult to interpret these huge data sets in all biological and clinical aspects. Hence, in addition to discovery by large-scale analysis, we have to employ functional and informatic approaches to better capitalize on the genomic data and to shed light on the biological significance of the genomic alterations or variants that have been identified by high-throughput genomic technologies.

**Declaration of interest**

The author declares that there is no conflict of interest that could be perceived as prejudicing the impartiality of the review.

**Funding**

This work has been financially supported by the BioBank Japan Project of the Ministry of Education, Culture, Sports, Sciences and Technology of the Japanese government, by Research grant #22390306 (H Nakagawa) from the Japan Society for the Promotion of Science, and by the Takeda Science Foundation.

**References**


Ghoshal S, Song H, Koessler T, Al Olama AA, Kote-Jarai Z, Driver KE, Pooley KA, Ramus SJ, Kjaer SK, Hogdall E et al. 2008 Multiple loci with different cancer specificities within the 8q24 gene desert. *Journal of the National Cancer Institute* 100 962–966. (doi:10.1093/jnci/djn190)


Chu LW, Meyer TE, Li Q, Menashe I, Yu K, Rosenberg PS, Huang WY, Quraishi SM, Kaaks R, Weiss JM et al. 2010 Association between genetic variants in the 8q24 cancer risk regions and circulating levels of androgens and sex hormone-binding globulin. *Cancer Epidemiology, Biomarkers & Prevention* 19 1848–1854. (doi:10.1158/1055-9965.EPI-10-1001)


Kumar A, White TA, Mackenzie AP, Clegg N, Lee C, Drumpit RF, Coleman I, Ng SB, Salpantie SJ, Rieder MJ et al. 2011 Exome sequencing identifies a
spectrum of mutation frequencies in advanced and lethal prostate cancers. PNAS 108 17087–17092. (doi:10.1073/pnas.1113740108)


Received in final form 22 April 2013
Accepted 23 April 2013
Made available online as an Accepted Preprint
26 April 2013