RESEARCH ARTICLE

Meta-analysis of Gene Expression Data Identifies Causal Genes for Prostate Cancer

Xiang-Yang Wang, Jian-Wei Hao, Rui-Jin Zhou, Xiang-Sheng Zhang, Tian-Zhong Yan, De-Gang Ding, Lei Shan*

Abstract

Prostate cancer is a leading cause of death in male populations across the globe. With the advent of gene expression arrays, many microarray studies have been conducted in prostate cancer, but the results have varied across different studies. To better understand the genetic and biologic mechanisms of prostate cancer, we conducted a meta-analysis of two studies on prostate cancer. Eight key genes were identified to be differentially expressed with progression. After gene co-expression analysis based on data from the GEO database, we obtained a co-expressed gene list which included 725 genes. Gene Ontology analysis revealed that these genes are involved in actin filament-based processes, locomotion and cell morphogenesis. Further analysis of the gene list should provide important clues for developing new prognostic markers and therapeutic targets.

Keywords: Prostate cancer - meta-analysis - gene expression data - co-expression - protein interaction

Asian Pacific J Cancer Prev, **14 (1)**, 457-461

Introduction

Prostate cancer is the most commonly diagnosed malignancy and a leading cause of cancer-related death among males over the age of fifty in the western world (Stangelberger et al., 2008). Although the age-adjusted rate of cancer deaths has decreased steadily in the past 10 years, prostate cancer remains the second leading cause of cancer deaths in men after lung cancer (Shen and Abate-Shen, 2010). The morbidity and mortality of prostate cancer is principal caused of its propensity to metastasize to other tissue, such as lung, liver and bone (Bubendorf et al., 2000; Logothetis and Lin, 2005). Therefore, understanding the mechanism of prostate cancer onset and metastasis is the key to treat this disease successfully and increasing survivability (Jiang et al., 1994).

Currently, there are many hypothesis of prostate cancer pathogenesis. Some investigators appreciate that tumor microenvironment plays an important role in the initiation and progression of prostate cancer (Tuxhorn et al., 2001; Chung et al., 2005; Alberti, 2006). In a recent study, Dakhova and colleagues found that formation of reactive stroma in prostate cancer result in alterations in a number of pathways including neurogenesis, axonogenesis and the DNA damage/repair pathways (Dakhova et al., 2009). Berger et al. found several prostate tumors contained complex chains of balanced rearrangements that occurred within or adjacent to known cancer genes and suggested genomic rearrangements may arise from transcriptional or chromatin aberrancies and engage prostate tumorigenic mechanisms (Berger et al., 2011). As the high-throughput technologies have been used in many fields, the detection of the expression level across the whole genome is a useful way to find unusual genomic alteration in prostate cancer patients with microarray (Rhea et al., 2011). Microarray gene expression profiling has identified several new biomarkers with diagnostic and possible prognostic value for prostate cancer, including AMACR, HPN, MUC1, AZGP1, CD166/MEMD, CD24, SLC18A2, TEAD1 and SPINK1 (Glinsky et al., 2004). Besides, polymorphisms of VDR gene and MDM2 gene were also reported to be associated with elevated prostate cancer risk (Chen et al., 2012; Guo et al., 2012).

However, there are disagreement among the study results of different microarray experiments and different statistic methods. Meta-analysis focused on contrasting and combining results from different studies, in the hope of identifying patterns among study results of prostate cancer.

Materials and Methods

Gene expression datasets of prostate cancer

Our meta-analysis was based on Chandran's study (Chandran et al., 2007) and Gorlov's study (Gorlov et al., 2009). In the first study, Chandran and colleagues analyzed gene expression profiles of 24 androgen-ablation resistant metastatic samples and 64 localized prostate tumor samples. Differential expression analysis showed that 415 genes were up-regulated and 364 genes were down-regulated in the progression of localized prostate cancer to metastatic prostate cancer. We selected the differentially

*Department of Urology, Henan Provincial People's Hospital, Zhengzhou, China *For correspondence: shanleileishan@hotmail.com*

	GeneA	GeneB	GeneC	⋯
GeneA	X_{aa}	X_{ab}	X_{ac}	
GeneB	X_{ba}	X_{bb}	X_{bc}	
GeneC	X_{ca}	X_{cb}	X_{cc}	
⋯				

(I) $Gene_i=(i1,i2,i3,i4,i5,\cdots)$
(II) $Gene_j=(j1,j2,j3,j4,j5,\cdots)$
(III) $X_{ij}=cor(Gene_i,Gene_j)$

Figure 1. The Work Flow of Extraction of Co-Expression Genes. We calculated the coexpression value between random two genes used the genes from Affy expression array, and marked the value as Xij. Genei represents the expression value of the i gene

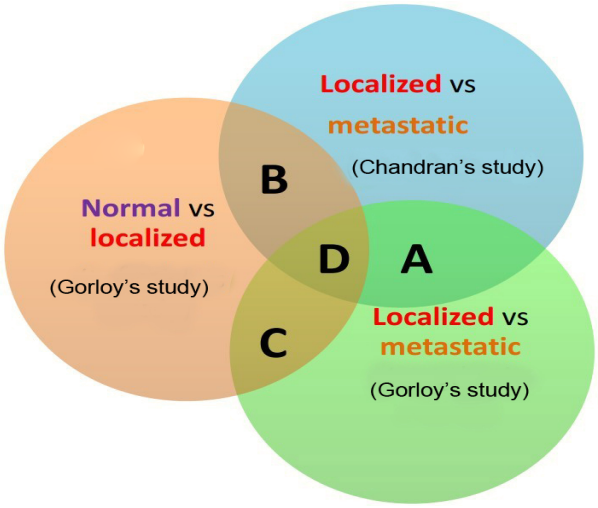

Figure 2. Venn Plot Displays the Gene Datasets from Chandran's Study and Gorloy's Study. The orange panel is the differentially expressed genes between normal prostate tissue and localized prostate cancer tissue from Gorloy's study. The blue panel is the differentially expressed genes between localized and metastatic prostate cancer tissue from Chandran's study. The green panel is the differentially expressed genes between localized and metastatic prostate cancer tissue from Gorloy's study. The A/B/C/D panels reflect the overlapping genes of those three gene datasets

expressed genes with at least two-fold change as our first dataset. This dataset consisted of 174 genes, and 73 of which were up-regulated genes in metastatic prostate cancer tissue.

The other dataset of our study were from Gorlov's meta-analysis. They analyzed 18 gene array datasets , including 11 datasets for the transition from normal prostate to localized prostate cancer and 7 datasets for the transition from localized prostate cancer to metastatic prostate cancer. These datasets can represent the dynamic process of prostate cancer. We extracted the top 500 significant differentially expressed genes between normal prostate and non-metastatic prostate cancer, and those between non-metastatic prostate cancer and metastatic prostate cancer as our second dataset. Overall, we conducted a meta-analysis of 1174 gene expression data.

Extraction of Co-expressed genes

In order to extract co-expressed genes of our target genes, we developed an advanced method to calculate the co-expression relationship. First, we downloaded expression dataset from Gene Expression Omnibus using "Affy U133 v2 platform" and "Human sapient" and a total of 1974 expression datasets were collected. Then, we calculated the D-value of each gene between the max expression level (Emax) and min expression level (Emin) (D-value = Emax - Emin) in each dataset. For each gene, we selected the top 30 D-value datasets and constructed gene expression matrix using those data (Figure 1)., Next, we calculated the Pearson coexpression value between random two genes, and marked the value as Xij. For example, Xba reflected the Pearson correlation between Gene A and Gene B, calculated by equation (III) in Figure 1. We applied this method for all human genes which could detect by Affy expression array, and got the correlation matrix of genes.

Protein-Protein Interaction network construction

Protein-protein interaction (PPI) network construction have become a major method for identifying biological relationships. As there are so many protein-protein interaction database publicly and each database has their own features, we constructed an in-house database with HPRD, IntAct, MIPS, BIND, DIP, MINT, PDZBase and Reactome databases recommended by a previous study (Mathivanan et al., 2006). To demonstrate the potential relationships among the differentially expressed genes, we matched the differentially expressed genes to PPI data that have been collected from the above in-house database. Based on the above relationships, an PPI network was constructed using the Cytoscape (Shannon et al., 2003).

Gene Ontology analysis

The Gene Ontology (GO) project (http://www.geneontology.org/) provides structured, controlled vocabularies and classifications that cover several domains of molecular and cellular biology and are freely available for community use in the annotation of genes, gene products and sequences (Harris et al., 2004). To further understand the functions of the gene list, we performed GO enrichment analysis using BINGO package (Maere et al., 2005) in R. The p-value < 0.01 was considered as significant level.

Biological Pathway Analysis

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a collection about understanding high-level functions and utilities of the biological system. These molecular datasets generated mostly by genome sequencing and other high-throughput experimental technologies (Kanehisa, 2002). To functional annotation of genes in the list, we identified the over-represented KEGG categories in pathways.

Results

Identification of differentially expressed genes

We collected a total of 1174 differentially expressed genes in the whole progression of prostate cancer from two studies. By integrating these genes, we found that

Table 1. The GO Term Enriched by the 725 Genes with High *P* value (P < 1.0E-05)

GO-ID	Description	Gene count	*p*-value
GO:0030029	actin filament-based process	289	6.87E-06
GO:0040011	locomotion	830	6.57E-06
GO:0000904	cell morphogenesis involved in differentiation	449	2.06E-06
GO:0003012	muscle system process	187	7.08E-10
GO:0006936	muscle contraction	170	2.54E-10
GO:0044421	extracellular region part	544	6.47E-06
GO:0008092	cytoskeletal protein binding	306	2.58E-06

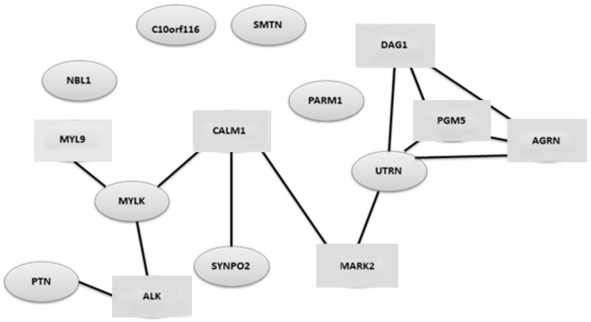

Figure 3. The Interaction Relationships among the Eight Differentially Expressed Genes. The oval genes are the differentially expressed genes. The rectangle genes are the high-related genes with differentially expressed genes according to PPI weight

there were 8 genes differentially expressed in the whole progression (Part D in Figure 2). These genes were NBL1, C10orf116, SMTN, PARM1, UTRN, SYNPO2, MYLK and PTN. We chose these 8 genes for further analysis.

Key Prostate Cancer Gene Network

We further constructed a PPI network using the 8 differentially expressed genes (Figure 3). We tried to find the shortest way to connect the differentially expressed genes using data from Protein-Protein Interaction Database. If the differentially expressed genes can be connected (significant) through one or two genes, we linked them with black line, and colored the linker genes blue (Figure 3). Gene PARM1, C10orf116, NBL1 and SMTN cannot be connected in short way.

Extraction of Co-expressed genes

To deeply analyze these eight genes, the newly developed method (co-expression relationship) was employed to find the potential co-expression relationships among them. For each differentially expressed gene, we got their expression level data from the 1974 datasets and calculated the D-value between max expression level and min expression level. And then, the top 30 D-value datasets for each gene were extracted. By calculating the Pearson correlation between the differentially expressed gene and genes in the selected 30 datasets using the method described in Figure 1, we obtained 8 co-expression gene lists corresponding to the 8 differentially expressed genes. Each list contained the top 500 co-expression genes according to correlation value.

Next, we merged those eight lists and selected the genes co-expressed with over two of the 8 differentially expressed genes, and finally obtained a gene list including

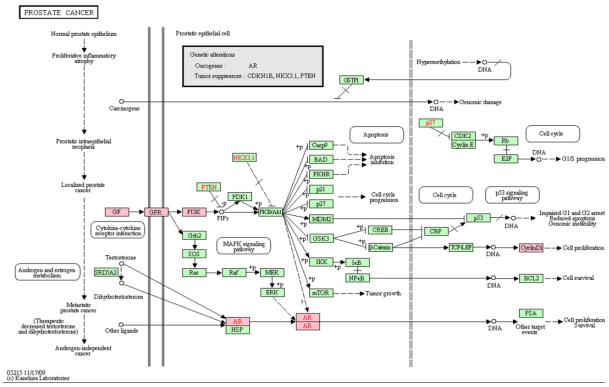

Figure 4. The Most Significant Enriched Pathway of the 725 Co-expressed Gene List. The genes in red box were co-expressed genes in our list

725 co-expressed genes. We believed that this list included most of the potential prostate cancer genes.

GO enrichment analysis

To functionally classify these 725 co-expressed genes, we performed GO analysis and observed significant enrichment of these genes in 7 GO categories (Table 1). The most significant enrichment was the GO category of actin filament-based process with P = 6.87E-06. The other significant GO categories included locomotion (P = 6.57E-06) and cell morphogenesis involved in differentiation (P = 2.06E-06).

Pathway enrichment analysis

To further investigate the function of this 725 co-expressed gene list, we mapped them to the KEGG database. Excitingly, seven of them were found in the prostate cancer pathway (KEGG-ID: hsa05215; Figure 4) with p-value of 0.03 in our KEGG analysis. This pathway contains many molecular alterations in prostate-cancer cells which implicates in carcinogen defenses (GSTP1), growth factor signaling pathways (NKX3.1, PTEN, and p27), and androgens (AR). They are critical determinants of the phenotype of prostate-cancer cells. These genes also covered some other important pathways. For example, six genes are in the pathway of hsa05200 (Pathways in Cancer)

Protein-Protein Interaction Analysis

Although some of the co-expressed genes in our list were mapped to KEGG pathways, there are still many genes that cannot be located in any pathway. Therefore, we used the protein-protein interaction data again to infer their function through their interacted protein. We hypothesized that the gene in our list involved in prostate cancer if its interactive genes involved in prostate cancer. And, if all genes inputted to PPI have similar function, there would be a regulation network among them.

Firstly, the network weight of the 725 genes were calculated basing on the PPI data. We chose the gene with the highest weight - ITGB1 for further analysis. ITGB1 has been reported to be involved in extra-cellular matrix interactions and be also related to many tumor types, including prostate cancer.

To construct a relatively simple network, we used

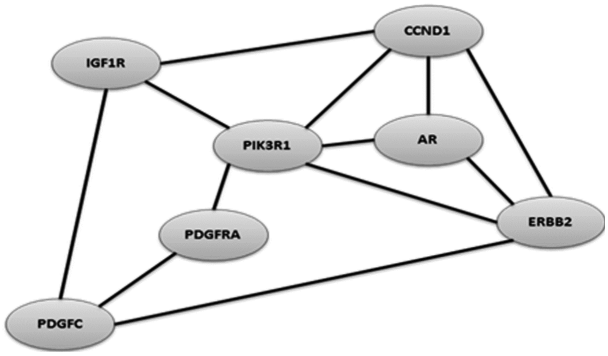

Figure 5. The Core PPI Network Between the 8 Differentially Expressed Genes and Top 7 Co-Expressed Genes. The genes in red color represent the interactive genes. The lines among them represent that those two genes have PPI relationships

a smaller gene list comprised the eight differentially expressed genes and the top 7 genes (high network weight) from 725 list to perform PPI analysis. We are glad to find that eight of them were interacted as a network (Figure 5).

Discussion

The overlapped differentially expressed genes were C10orf116, PARM1, MYLK, NBL1, PTN, SMTN, SYNPO2 and UTRN. Among these identified genes, we found that some of them were classically known biomarkers which are closely related to prostate cancer progression, such as PARM1 and MYLK.

Previous studies have reported that PARM-1 is a novel mucin-like, androgen-regulated gene exhibiting proliferative effects in prostate cancer cells (Fladeby et al., 2008), and it may play a role in prostatic cell immortalization (Cornet et al., 2003). The MYLK gene was also presented significant changes in functional connectivity between normal and tumoral conditions of prostates (Fujita et al., 2008). Moreover, SYNPO2 expression has been reported to suppress tumor growth (Korkola et al., 2009) and found to inhibit proliferation and invasiveness of prostate cancer cells. It is a homolog of myopodin which suppresses tumor growth and metastasis in prostate cancer (Jing et al., 2004). To confirm our result, we queried these eight genes on the Cancer Gene Census provided by Sanger Institute (CGC, http://www.sanger.ac.uk/genetics/CGP/Census/), and found three of them were recorded in CGC prostate cancer list (17 totally). This suggested that our result was reliable.

PPI network construction showed that 4 of the 8 differentially expressed genes can be connected by 7 genes (Figure 3). We focused on the gene CALM1, as it connected MYLK, SYNPO2 directly and connected UTRN and PTN through one or two genes. It may play an important role in the progression of prostate cancer.

Among the 725 co-expressed gene list, four of them were prostate related genes known well. They are PARM1, PMEPA1, PTGDS, and PTGER2. The PMEPA1 gene has been shown to suppress the androgen receptor (AR) and TGF-β signaling pathway and is abnormally expressed in prostate tumors (Liu et al., 2011). Besides, PTGDS and PTGER2 have been reported as AR-regulated or involved in prostate cancer (Love et al., 2009).

As GO analysis result in Table 1, those 725 genes were enriched in several biological processes. The most significant GO category was actin filament-based process. It have been reported that miR-1 had down-regulated target gene in actin filament-based process in human prostate tumors, consistent with the array data in the Hudson's study (Hudson et al., 2011). The second significant GO category in our result was locomotion. A recent study suggested that homotypic collisions between two prostate cancer cells results in contact inhibition of locomotion, which was mediated by EphA-Rho-Rho kinase signaling (Astin et al., 2010). Bonkhoff collected many experience about benign and malignant prostate tissue, and finally found that there are many evidence that cell morphogenesis will affect neuroendocrine differentiation in prostate cancer (Bonkhoff, 1998). Almost all the significant GO function have been verified to have relationship with prostate cancer, except for the fourth category in Table 1- muscle system process. The definition of this category is an organ system process carried out at the level of a muscle. The physical sign of prostate cancer patients always includes acratia phenomena and it's always not easy to detect prostate cancer early, we should pay attention to genes enriched in this category.

From Figure 5, we could easily find that PIK3R1 was a hub gene. To date, activated mutations of PIK3R1 have not been reported in prostate cancer (Boormans et al., 2010). But why this gene located in the center site of the network? Could it imply us the relation between PIK3R1 with prostate cancer? From the Figure 3, we know that AR plays an important role in prostate cancer progression. The silence of IGF1R will enhance DNA damage in prostate cancer, which activated by the PTEN mutation (Rochester et al., 2004). Another study (Savinainen et al., 2002) had verified that ERBB2 plays a role in the development of prostate cancer. McCabe showed that inhibition of PDGFRA reduces proliferation of prostate cancer cells (McCabe et al., 2008). All these studies tell us that the remaining genes in the network of Figure 5 may also related to prostate cancer. It seems that the hub gene of PIK3R1 plays an control role of this network undoubtedly.

In conclusion, Many of the genes in our 725 co-expression list have been reported to be implicated in the oncogenesis of prostate cancer. Therefore, the remaining genes in our list are worthy to be validated in wet lab, especially the genes in Figure 3 and Figure 5.

References

Alberti C (2006). Prostate cancer progression and surrounding microenvironment. *Int J Biol Markers*, **21**, 88-95.

Astin JW, Batson J, Kadir S, et al (2010). Competition amongst Eph receptors regulates contact inhibition of locomotion and invasiveness in prostate cancer cells. *Nat Cell Biol*, **12**, 1194-204.

Berger MF, Lawrence MS, Demichelis F, et al (2011). The genomic complexity of primary human prostate cancer. *Nature*, **470**, 214-20.

Bonkhoff H (1998). Neuroendocrine cells in benign and malignant prostate tissue: morphogenesis, proliferation, and androgen receptor status. *Prostate Suppl*, **8**, 18-22.
Boormans J, Korsten H, Ziel-van Der Made A, et al (2010). E17K substitution in AKT1 in prostate cancer. *Bri J cancer*, **102**, 1491-4.
Bubendorf L, Schopfer A, Wagner U, et al (2000). Metastatic patterns of prostate cancer: an autopsy study of 1,589 patients. *Hum Pathol*, **31**, 578-83.
Chandran U, Ma C, Dhir R, et al (2007). Gene expression profiles of prostate cancer reveal involvement of multiple molecular pathways in the metastatic process. *BMC Cancer*, **7**, 64.
Chen T, Yi SH, Liu XY, Liu ZG (2012). Meta-analysis of Associations between the MDM2-T309G polymorphism and prostate cancer risk. *Asian Pac J Cancer Prev*, **13**, 4327-30.
Chung LW, Baseman A, Assikis V, Zhau HE (2005). Molecular insights into prostate cancer progression: the missing link of tumor microenvironment. *J Urol*, **173**, 10-20.
Cornet AM, Hanon E, Reiter ER, et al (2003). Prostatic androgen repressed message-1 (PARM-1) may play a role in prostatic cell immortalisation. *Prostate*, **56**, 220-30.
Dakhova O, Ozen M, Creighton CJ, et al (2009). Global gene expression analysis of reactive stroma in prostate cancer. *Clin Cancer Res*, **15**, 3979-89.
Fladeby C, Gupta SN, Barois N, et al (2008). Human PARM-1 is a novel mucin-like, androgen-regulated gene exhibiting proliferative effects in prostate cancer cells. *Int J Cancer,* **122**, 1229-35.
Fujita A, Gomes LR, Sato JR, et al (2008). Multivariate gene expression analysis reveals functional connectivity changes between normal/tumoral prostates. *BMC Syst Biol*, **2**, 106.
Glinsky GV, Glinskii AB, Stephenson AJ, et al (2004). Gene expression profiling predicts clinical outcome of prostate cancer. *J Clini Invetst*, **113**, 913-23.
Gorlov I, Byun J, Gorlova O, et al (2009). Candidate pathways and genes for prostate cancer: a meta-analysis of gene expression data. *BMC Med Genomics*, **2**, 48.
Guo YJ, Shi ZM, Liu JD, et al (2012). Meta-analysis of the relation between the VDR gene TaqIpolymorphism and genetic susceptibility to prostate cancer in Asian populations. *Asian Pac J Cancer Prev*, **13**, 4441-4.
Harris M, Clark J, Ireland A, et al (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*, **32**, D258.
Hudson RS, Yi M, Esposito D, et al (2012). MicroRNA-1 is a candidate tumor suppressor and prognostic marker in human prostate cancer. *Nucleic Acids Res*, **40**, 3689-703 .
Jiang WG, Puntis MC, Hallett MB (1994). Molecular and cellular basis of cancer invasion and metastasis: implications for treatment. *Br J Surg*, **81**, 1576-90.
Jing L, Liu L, Yu YP, et al (2004). Expression of myopodin induces suppression of tumor growth and metastasis. *Am J Pathol*, **164**, 1799-806.
Kanehisa M (2002). The KEGG database. *Novartis Found Symp*, **247**, 91-101; discussion -3, 19-28, 244-52.
Korkola JE, Houldsworth J, Feldman DR, et al (2009). Identification and validation of a gene expression signature that predicts outcome in adult men with germ cell tumors. *J Clin Oncol*, **27**, 5240-7.
Liu R, Zhou Z, Huang J, Chen C (2011). PMEPA1 promotes androgen receptor-negative prostate cell proliferation through suppressing the Smad3/4-c-Myc-p21 Cip1 signaling pathway. *J Pathol*, **223**, 683-94.
Logothetis CJ, Lin SH (2005). Osteoblasts in prostate cancer metastasis to bone. *Nat Rev Cancer*, **5**, 21-8.
Love HD, Booton SE, Boone BE, et al (2009). Androgen regulated genes in human prostate xenografts in mice: relation to BPH and prostate cancer. *PLoS One*, **4**, e8384.
Maere S, Heymans K, Kuiper M (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, **21**, 3448-9.
Mathivanan S, Periaswamy B, Gandhi T, et al (2006). An evaluation of human protein-protein interaction data in the public domain. *BMC Bioinformatics*, **7**, S19.
McCabe CD, Spyropoulos DD, Martin D, Moreno CS (2008). Genome-wide analysis of the homeobox C6 transcriptional network in prostate cancer. *Cancer Res*, **68**, 1988-96.
Rhea JM, Singh HV, Molinaro RJ (2011). Next generation sequencing in the clinical molecular diagnosis of cancer. Medical Laboratory Observer.
Rochester MA, Riedemann J, Hellawell GO, et al (2004). Silencing of the IGF1R gene enhances sensitivity to DNA-damaging agents in both PTEN wild-type and mutant human prostate cancer. *Cancer Gene Ther*, **12**, 90-100.
Savinainen KJ, Saramäki OR, Linja MJ, et al (2002). Expression and gene copy number analysis of ERBB2 oncogene in prostate cancer. *Am J Pathol*, **160**, 339.
Shannon P, Markiel A, Ozier O, et al (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, **13**, 2498-504.
Shen MM, Abate-Shen C (2010). Molecular genetics of prostate cancer: new prospects for old challenges. *Genes Dev*, **24**, 1967-2000.
Stangelberger A, Waldert M, Djavan B (2008). Prostate cancer in elderly men. *Rev Urol*, **10**, 111.
Tuxhorn JA, Ayala GE, Rowley DR (2001). Reactive stroma in prostate cancer progression. *J Urol*, **166**, 2472-83.