Analysis of fusion gene expression in prostate tumors by using single-end reads

D.D. Xie[1], J.Y. Li[2], Y. Wang[1], L. Chen[1] and D.X. Yu[1]

[1]Department of Urology, The Second Affiliated Hospital,
Anhui Medical University, Hefei, China
[2]Institute of Prostate Cancer and LeFrak Center for Robotic Surgery,
James Buchanan Brady Foundation Department of Urology,
Weill Cornell Medical College, Presbyterian Hospital,
New York, NY, USA

Corresponding author: D.X. Yu
E-mail: yudexinsy@hotmail.com

ABSTRACT. Fusion gene expression, a kind of chromosome rearrangement mode, has been strongly linked to prostate cancer diagnosis and prognosis as well as to the Gleason score and the American Joint Committee on Cancer stage assessment. In combination with traditional methods for locating fusion genes and scoring their association with cancer cell growth, proliferation, and invasion through the basement membrane, the emerging high-throughput sequencing technologies offer a panorama of fusion genes in a genome and facilitate the discovery of new fusion modes. We describe here a method for using single-end reads to analyze fusion gene expression in prostate tumors. We obtained the fusion gene expression profiling of prostate tumors, clustered them into several biological pathways, highlighted three "rediscovered" fusion genes (*TMPRSS2-ERG*, *KLK2*, and *KLK3*) and proved the reliability of our method.

Key words: Fusion gene; Prostate cancer; Single-end reads

INTRODUCTION

Prostate cancer (PCa) is the third most common tumor type in men. The appearance of this neoplasia is relevant to age. In the European Union, PCa is directly responsible for 3% of men deaths and accounts for 10% of cancer deaths. The incidence of PCa has risen in recent years, owing primarily to significant increases in life expectancy and secondarily to the introduction of serum prostate-specific antigen (PSA) determination in PCa screening, which has risen diagnostic accuracy in the preclinical phase. In Spain, the epidemiological status of PCa is not significantly different from that in the rest of Europe. Every year, 13,300 cases are diagnosed (13.6% of tumors among Spanish men), of which approximately 65% were 5-year survivals and the deaths with an average age of 75 years (Fernandez-Serra et al., 2011).

Histologically, PCa is composed of a heterogeneous mixture of cells - mainly epithelial and stromal (Nelson et al., 2003). The disease process begins with a dysplasia that starts as a proliferative inflammatory atrophy, which progresses to prostatic intraepithelial neoplasia and in some cases leads to a carcinoma. Evidence suggests that a trigger of tumorigenesis could be prostatic inflammation due to infectious agents or ingestion of carcinogens. In parallel, some cells accumulate genetic alterations that, along with androgenic signaling, stimulate the growth and proliferation of a tumor (Taichman et al., 2007).

Clinically, patients of PCa have been classified into two catalogs: prostate tumors with the capability to spread that end up being lethal, and tumors that are relatively indolent (Taichman et al., 2007). This division raises the problems of distinguishing one kind of tumors from others and determining the best clinical approach in each case.

Currently, serum PSA levels provide highly organ-specific information but little disease-specific information. Thus, in both benign prostatic hyperplasia and prostatitis, serum increases this biomarker occurrence, and many patients with localized PCa have PSA values that overlap with those of healthy subjects, resulting in a gray area of difficult interpretation at levels ranging from 4 to 10 ng/mL (Balk et al., 2003). Moreover, numerous studies have suggested that 30-50% of PCa cases were over-diagnosed, implying that not all patients with elevated PSA have a prostate tumor. After diagnosis, the main prognostic factor is the Gleason score, which assigns a grade of 1-5 in descending differentiation to each of the 2 main foci of the tumor. The sum of both values is the score. Although this parameter is the gold standard in the clinical management of PCa, it presents certain problems: first, the determination is made on tissue obtained through prostate biopsy, a surgical procedure that has comorbidity, especially in elderly patients; second, it is subject to interpretive variation (Evans et al., 2008).

In the prognosis of the disease, the lack of a reliable method for determining the time at which a prostate tumor will become hormone resistant is problematic because from that point forward, patient prognosis worsens, and bone metastases, for which only palliative treatment is currently available, often occurs (Msaouel et al., 2008). Therefore, the identification of new biomarkers as useful tools in the diagnosis and clinical management of PCa is important. These markers should be determinable using objective, quantitative, and mechanism-specific techniques and, as much as possible, be accessible via noninvasive methods.

The recent discovery of fusion of the transmembrane-serine protease gene (*TMPRSS2*) with erythroblast transformation-specific (*ETS*) family members (*TMPRSS2-ETS*) and their possible involvement in the clinical management of patients with PCa makes fusion genes as specific markers for prostate tumor diagnosis and prognosis. *TMPRSS2-ETS* fusion gene

rearrangements constitute a very common and specific alteration in PCa cells. These genetic alterations lead to the overexpression of *ETS* genes, which encode the *E26* family of transcription factors involved in cell proliferation. Of this family, the overexpressed *ERG* (ETS-related gene) oncogene is in almost 50% of PCa cases.

TMPRSS2-ERG overexpresses *ERG* through an androgen-mediated response. Structurally, the rearrangement is mainly due to interstitial deletion and to a lesser extent to reciprocal translocation and plays a key role in cellular metabolism. Almost all fusion gene transcripts produce a truncated ERG protein, and the presence of a specific isoform of this gene suggests clonality in a tumor; hence, metastasis shares the fusion gene status of the primary lesion. Although the prognostic implications of *TMPRSS2-ERG* have not been fully elucidated, *TMPRSS2-ERG* constitute a field of great diagnostic potential, and therefore, the development of techniques to identify and analyze the presence and characteristics of this gene noninvasively deserves attention. Currently, evidence supports the hypothesis that the presence of fusion genes can differentiate two molecular groups within PCa with differential behavior, demonstrating the roles of fusion gene as a potential therapeutic target. In this regard, the use of an anti-histone deacetylase (trichostatin A), antagonists of estrogen receptor alpha, and abiraterone acetate, has shown promised results.

A fusion gene is a hybrid gene formed from 2 previously separate genes. Fusion genes occur as the result of translocation, interstitial deletion, or chromosomal inversion. Fusion genes are often oncogenes, such as *BCR-ABL* (gene encoding the break point cluster region protein-Abelson murine leukemia viral oncogene homolog 1) (Nowell, 1962), *TEL-AML* (translocation ETS leukemia-acute myeloid leukemia), *AML1-ETO* (t(8;21) chromosomal translocation in acute myeloid leukemia-eight twenty-one), and the *TMPRSS2-ERG* fusion gene often results in PCa (Tomlins et al., 2005). Oncogenic fusion gene may lead to a new gene product with special function that is new or different compared with that of the 2 fusion partners. Alternatively, a proto-oncogene and a strong promoter may fuse together. As a result, the oncogenic function may disorder by upregulation via the strong promoter of the upstream fusion partner. Oncogenic fusion transcripts may also be activated by trans-splicing or read-through events (Nacu et al., 2011).

RNA-Seq research applies two main methods (Maher et al., 2009a) to detect gene fusion events (Neuhausen et al., 2005). One method involves a targeted alignment approach, which aligns reads to a set of artificial exon-exon target sequences (Wang et al., 2008). However, for intergenic splicing events in general, targeted alignment is not applicable as a computational strategy. The other method is to align the reads to a reference genome using a program that can split a read to align to different locations in a genome. Many spliced alignment tools can detect gene fusion events. These tools include QPALMA (De Bona et al., 2008), TopHat (Trapnell et al., 2009), SplitSeek (Ameur et al., 2010), and GSNAP (Wu and Nacu, 2010). Some of them provide the capability to find fusion gene events at distant or inter-chromosomal locations in a genome. Spliced alignment is more popular than targeted alignment for detecting gene fusion events, because it can identify novel or distant gene fusion. However, spliced alignment is generally effective only for reads with exon-exon junctions in their middle regions, away from the 14- to 20-nucleotide margins at their ends. Another strategy is to align paired-end reads to a genome separately and search for fusion genes when 2 paired-end reads are distant in different gene locations. Studies suggest that a paired-end strategy is more sensitive to find gene fusion (Maher et al., 2009b). The FusionSeq program is based on paired-end reads (Sboner et al.,

2010). However, the single-end data in fusion genes is precluded. The FusionMap program has been recently developed to align fusion reads directly to a genome without prior knowledge of potential fusion regions. It can detect fusion events in both single- and paired-end datasets from RNA-Seq analysis and characterize fusion junctions at base-pair resolution (Ge et al., 2011).

These methods shed new light on finding fusion gene targets for oncotherapy. A promising approach is to screen featured target genes using treatment targeted at one such fusion gene locus. The overexpression of fusion genes is highly associated with cancer prognosis and diagnosis (Lijovic et al., 2000; Ferretti et al., 2007).

MATERIAL AND METHODS

We downloaded RNA-Seq data of PCa from the Gene Expression Omnibus database with accession No. GSE24283. These data contain 3 human primary prostate tumors (T1, T2, and T3), adjacent matched normal tissue samples (N1, N2, and N3), a human brain reference (HBR) RNA sample, and a universal human reference (UHR) RNA sample. The processed libraries were sequenced via Illumina on a genome analyzer using their single-end protocol (Illumina, Hayward, CA, USA). Read length was 33 nucleotides for samples N3, T1, T2, and T3; 50 nucleotides for N2, HBR, and UHR; and 50 and 75 nucleotides for sample N1. Samples T3 and N1 were each sequenced over 2 Illumina flow cells. A total 226 million single-end reads from 8 sequencing runs for tumors T1, T2, and T3, and matched normal samples N1, N2, and N3 were obtained; 53 million reads for HBR and 60 million reads for UHR were also obtained.

We then cut off bases with qualities of less than 20 and retained reads with lengths greater than 25. The aligner bowtie mapped reads to the human reference genome (hg19) with parameter v. After the bowtie screening, we obtained unmatched reads for subsequent analysis. Last, the unmatched reads were further mapped by using the FusionMap aligner.FusionMap is an efficient fusion aligner that aligns single-end reads spanning fusion junctions directly to hg19 with the RefGene model. Results of each sample were further analyzed in respect of the gene fusion events. When using FusionMap aligner, we selected a minimal parameter length of 15 and other default parameters.

RESULTS

Characters of seed and rescued counts

FusionMap detects fusion junctions based on seed reads, which contain the fusion position in the middle regions of the reads and rescued reads that lose (gain) a few base pairs on one fusion edge and gain (lose) an equal number of base pairs on the other. Thus, the seed and rescued counts represent expression values of fusion genes. We drew each sample of seed counts and rescued counts as the cumulative frequency to show the abundance of gene fusion expression. The green, yellow, and blue circles represent counting of fusion events in each sample, meanwhile, the intercross area illustrates the events that co-exist in the three samples (Figures 1 and 2). To measure the gene fusion event regardless of seed count or rescue count and the sum of seed and rescued reads, fusion gene expression abundance of HBR and UHR is always greater than T and N samples (Figures 1 and 2). In general, the abundance of gene fusion expression in T samples is less than that in N samples.
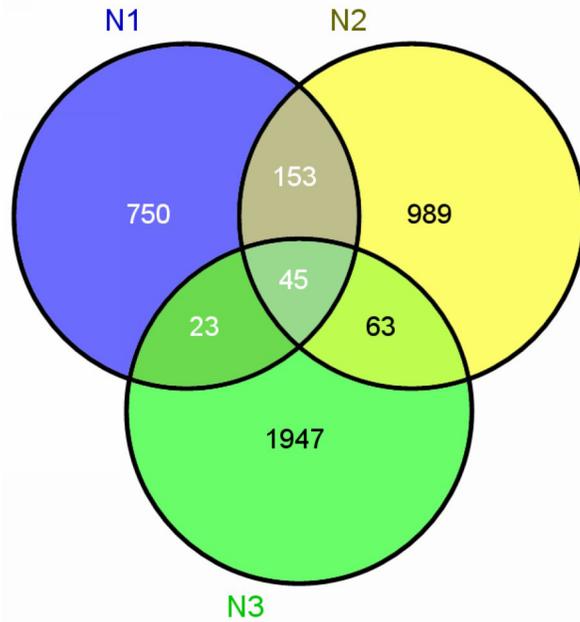
Figure 1. N1, N2, and N3 counts. The blue, yellow, and green circles are the counting of fusion events in normal samples 1, 2, and 3. The intercross area illustrates the events that co-exist in the three samples.
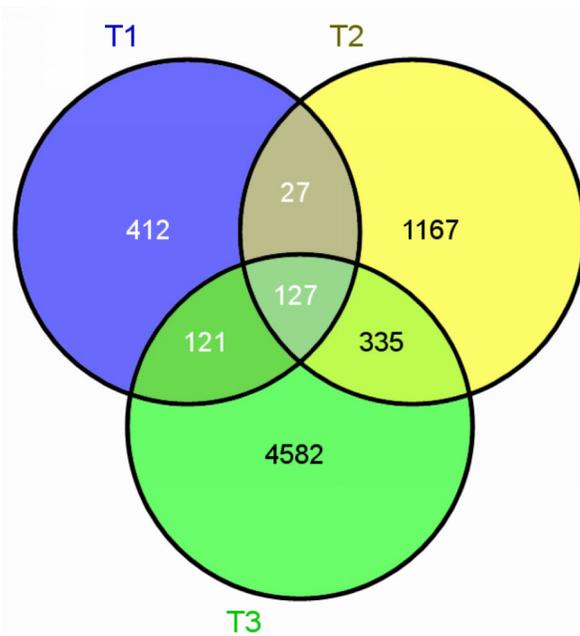

Figure 2. T1, T2, and T3 counts. The blue, yellow, and green circles are the counting of fusion events in tumor samples 1, 2, and 3. The intercross area illustrates the events that co-exist in the three samples.

Characters of gene fusion events

Among a high frequency of gene fusion events in all 3 normal samples (1571, 1250, and 2078), 45 event counts co-existed (see Figure 1), which may be explained by the fact that a basic set of gene fusion events exists among genome regions under evolutionary pressure as well as a certain heterogeneity of fusion gene expression among normal species groups. However, we discovered that gene fusion events introduce large vibrations between different normal samples. This result matches that of the molecular mechanism of gene fusion events.

Up to 127 counts (Figure 2), a somewhat surprisingly high but reasonable frequency of gene fusion events, were found among patients with tumors (687, 1656, and 9412). Researchers should pay more attention to fusion genes because of their relevance to cancer development. Considering this list carefully, we found some familiar genes previously implicated in tumors, such as *TMPRSS2-ERG* and *ETS* fusion (180 publications) (Salagierski and Schalken, 2012), kallikrein-related peptidase 3 (*KLK3*), *KLK2*, insulin response sequence 2 (IRS2), and interleukin 2 fusion genes (Table 1).

Table 1. Gene Ontology cluster results.

Term	Count	P value	Benjamin
Positive regulation of cellular component organization	9	2.20E-03	9.50E-01
Regulation of heart contraction	6	3.00E-03	8.70E-01
Negative regulation of angiogenesis	4	3.00E-03	7.40E-01
Neuron differentiation	13	1.00E-02	9.70E-01
Axon guidance	6	1.20E-02	9.60E-01
Positive regulation of heart contraction	3	1.30E-02	9.50E-01
Neuron projection development	9	1.70E-02	9.60E-01
Cell motion	13	1.80E-02	9.60E-01
Cell projection organization	11	1.90E-02	9.50E-01
Regulation of cell projection organization	5	2.70E-02	9.80E-01
Regulation of neuron differentiation	6	2.70E-02	9.70E-01
Neuron development	10	2.90E-02	9.70E-01
Cellular component morphogenesis	11	3.00E-02	9.60E-01
Muscle cell development	4	3.40E-02	9.70E-01
Axonogenesis	7	3.70E-02	9.70E-01
Muscle contraction	6	4.60E-02	9.80E-01
Regulation of angiogenesis	4	4.60E-02	9.80E-01
Regulation of cell development	7	4.70E-02	9.70E-01
Anti-apoptosis	7	4.90E-02	9.70E-01

Gene Ontology analysis of fusion genes in tumor samples

We chose the David online software to perform Gene Ontology analysis on common fusion genes. The parameters were set to a count of 3 and a Benjamin test, EASE of 0.05. The results were clustered 128 forecasted gene-fusion events (Table 1). From the table, we may infer that the fusion events occur in the positive regulation of cellular component organization, cell motion and cell projection organization, which were likely responsible for cancer formation.

DISCUSSION

At present, the software for predicting gene fusion expression essentially functions

based on paired-end reads in RNA-Seq data. Therefore, a large number of these predictions lack strength and drop information. Using single-end reads, we used the FusionMap software to detect fusion genes and summarize the feasibility of capturing them. In this study, we applied FusionMap to detect gene fusion events in PCa with single-end reads. Forty-five fusion events were found in 3 normal samples, and 127 counts were found in patients with triple tumors. Among the counts, several eminent fusion genes reportedly associated with PCa course were "rediscovered", including *TMPRSS2-ERG*, *KLK2*, *KLK3*, and others.

Androgen stimulation treatment of prostatic adenocarcinoma can increase the expression of *TMPRSS2-ERG*, which then inhibits PSA activity in LNCaP cells and partially inhibits it in VCaP cells at the transcriptional level (Yin et al., 2011). In addition, overexpression of *TMPRSS2-ERG* fusion induces DNA damage, which may be potentiated by poly(ADP-ribose) polymerase-1 inhibition in a manner similar to that of breast cancer and BCRA1/2-deficiency (Brenner et al., 2011). However, *TMPRSS2-ERG* marker cannot serve as a diagnostic or prognostic marker for ovarian cancer (Huang et al., 2011). These fusion genes may be useful as a diagnosis index in both primary and metastatic PCa (Xiao et al., 2011; Guo et al., 2012) and as a PCa treatment target, which will decrease the overtreatment rate when only standard diagnostic tools are used (Salagierski and Schalken, 2012).

The *KLK3* locus on chromosome 19q13.33 and is associated with PCa susceptibility; however, KLK3 variants, which are associated with PSA level, cannot completely explain the association with PCa risk (Penney et al., 2011). Three highly correlated SNPs (rs176325425, rs62113212, and rs62113214) are associated with PCa, and rs11084033 has potential prognostic significance in ovarian cancer (O'Mara et al., 2011).

PSA is activated by *KLK2* when the cells are in physical contact, and *KLK2* is the protease responsible for activating PSA. However, active PSA is insufficient to induce the development of PCa or precursors (Williams et al., 2010). The W allele of the *KLK2* R250W SNP is reportedly less likely to be associated with low Gleason score morphology and hence is not an effective prognostic index (Kohli et al., 2010).

Functional polymorphisms in the genes for insulin, IRS1, IRS2, and insulin-like growth factor 1 may be associated with PCa. IRS1 G972R GR/RR genotypes are associated with a significant 2.8-fold increased risk for PCa and significantly associated with more advanced Gleason score and American Joint Committee on Cancer stage (Neuhausen et al., 2005). Deletion of *Irs2* did not affect the initiation of neoplasia in Pten(+/-) mice but suppressed cancer cell growth, proliferation, and invasion through the basement membrane (Szabolcs et al., 2009).

Although there were limitations in our conclusions, many previous studies supported our results, indicating that chromosome rearrangement was found in more than half of PCa cells and new proteins encoded by fusion genes advanced the development of cancer (Perner et al., 2006; Saramaki and Visakorpi, 2007; Teixeira, 2008; Demichelis et al., 2009). A recent finding in cancer cells is that the 3-dimensional structure of DNA displays ultrahigh compression in the changed chromosome, which will manipulate thousands of downstream genes. The turning on (or off) of gene expression and activities forges the abnormal phenotype and activity of cancer cells. Conversely, the chromosome ectopic mode destabilizes the genome and raises the hazard of pre-cancerous lesions and mutations. When fusion genes overexpress, these mutations and changes occur in most types of cancer cells.

Some believe that androgen-regulated *TMPRSS2-ERG* gene fusion is a strong prog-

nostic factor for disease recurrence after prostatectomy and its transcripts are related to tumor proliferation, invasion, and aggressive phenotype (Swanson et al., 2011). In addition to rediscovering fusion genes in prostate tumors, we believe that some of these fusion genes are also highly associated with prostate tumors and hence will advance research on the mechanism of prostate tumor pathogenesis.

However, the role of *TMPRSS2-ERG* gene fusion remains controversial in PCa. Some still believe that *TMPRSS2-ERG* fusion is not prognostic for recurrence after retropubic radical prostatectomy (Toubaji et al., 2011). Considering the varying significance of *TMPRSS2-ERG* gene fusion prevalence and class in PCa (Magi-Galluzzi et al., 2011), larger cohorts must be deeper researched and conducted before clinical application of this fusion gene.

REFERENCES

Ameur A, Wetterbom A, Feuk L and Gyllensten U (2010). Global and unbiased detection of splice junctions from RNA-seq data. *Genome Biol.* 11: R34.

Balk SP, Ko YJ and Bubley GJ (2003). Biology of prostate-specific antigen. *J. Clin. Oncol.* 21: 383-391.

Brenner JC, Ateeq B, Li Y, Yocum AK, et al. (2011). Mechanistic rationale for inhibition of poly(ADP-ribose) polymerase in ETS gene fusion-positive prostate cancer. *Cancer Cell* 19: 664-678.

De Bona F, Ossowski S, Schneeberger K and Ratsch G (2008). Optimal spliced alignments of short sequence reads. *Bioinformatics* 24: i174-i180.

Demichelis F, Setlur SR, Beroukhim R, Perner S, et al. (2009). Distinct genomic aberrations associated with ERG rearranged prostate cancer. *Genes Chromosomes Cancer* 48: 366-380.

Evans AJ, Henry PC, van der Kwast TH, Tkachuk DC, et al. (2008). Interobserver variability between expert urologic pathologists for extraprostatic extension and surgical margin status in radical prostatectomy specimens. *Am. J. Surg. Pathol.* 32: 1503-1512.

Fernandez-Serra A, Rubio-Briones J, Garcia-Casado Z, Solsona E, et al. (2011). Prostate cancer: the revolution of the fusion genes. *Actas Urol. Esp.* 35: 420-428.

Ferretti G, Felici A, Papaldo P, Fabi A, et al. (2007). HER2/neu role in breast cancer: from a prognostic foe to a predictive friend. *Curr. Opin. Obstet. Gynecol.* 19: 56-62.

Ge H, Liu K, Juan T, Fang F, et al. (2011). FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. *Bioinformatics* 27: 1922-1928.

Guo CC, Wang Y, Xiao L, Troncoso P, et al. (2012). The relationship of TMPRSS2-ERG gene fusion between primary and metastatic prostate cancers. *Hum. Pathol.* 43: 644-649.

Huang L, Schauer IG, Zhang J, Mercado-Uribe I, et al. (2011). The oncogenic gene fusion TMPRSS2: ERG is not a diagnostic or prognostic marker for ovarian cancer. *Int. J. Clin. Exp. Pathol.* 4: 644-650.

Kohli M, Rothberg PG, Feng C, Messing E, et al. (2010). Exploratory study of a KLK2 polymorphism as a prognostic marker in prostate cancer. *Cancer Biomark.* 7: 101-108.

Lijovic M, Fabiani ME, Bader J and Frauman AG (2000). Prostate cancer: are new prognostic markers on the horizon? *Prostate Cancer Prostatic Dis.* 3: 62-65.

Magi-Galluzzi C, Tsusuki T, Elson P, Simmerman K, et al. (2011). TMPRSS2-ERG gene fusion prevalence and class are significantly different in prostate cancer of Caucasian, African-American and Japanese patients. *Prostate* 71: 489-497.

Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, et al. (2009a). Transcriptome sequencing to detect gene fusions in cancer. *Nature* 458: 97-101.

Maher CA, Palanisamy N, Brenner JC, Cao X, et al. (2009b). Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 106: 12353-12358.

Msaouel P, Pissimissis N, Halapas A and Koutsilieris M (2008). Mechanisms of bone metastasis in prostate cancer: clinical implications. *Best Pract. Res. Clin. Endocrinol. Metab.* 22: 341-355.

Nacu S, Yuan W, Kan Z, Bhatt D, et al. (2011). Deep RNA sequencing analysis of readthrough gene fusions in human prostate adenocarcinoma and reference samples. *BMC Med. Genomics* 4: 11.

Nelson WG, De Marzo AM and Isaacs WB (2003). Prostate cancer. *N. Engl. J. Med.* 349: 366-381.

Neuhausen SL, Slattery ML, Garner CP, Ding YC, et al. (2005). Prostate cancer risk and IRS1, IRS2, IGF1, and INS polymorphisms: strong association of IRS1 G972R variant and cancer risk. *Prostate* 64: 168-174.

Nowell PC (1962). The minute chromosome (Phl) in chronic granulocytic leukemia. *Blut* 8: 65-66.

O'Mara TA, Nagle CM, Batra J, Kedda MA, et al. (2011). Kallikrein-related peptidase 3 (KLK3/PSA) single nucleotide polymorphisms and ovarian cancer survival. *Twin Res. Hum. Genet.* 14: 323-327.

Penney KL, Schumacher FR, Kraft P, Mucci LA, et al. (2011). Association of KLK3 (PSA) genetic variants with prostate cancer risk and PSA levels. *Carcinogenesis* 32: 853-859.

Perner S, Demichelis F, Beroukhim R, Schmidt FH, et al. (2006). TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. *Cancer Res.* 66: 8337-8341.

Salagierski M and Schalken JA (2012). Molecular diagnosis of prostate cancer: PCA3 and TMPRSS2:ERG gene fusion. *J. Urol.* 187: 795-801.

Saramaki O and Visakorpi T (2007). Chromosomal aberrations in prostate cancer. *Front Biosci.* 12: 3287-3301.

Sboner A, Habegger L, Pflueger D, Terry S, et al. (2010). FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. *Genome Biol.* 11: R104.

Swanson TA, Krueger SA, Galoforo S, Thibodeau BJ, et al. (2011). TMPRSS2/ERG fusion gene expression alters chemo- and radio-responsiveness in cell culture models of androgen independent prostate cancer. *Prostate*.

Szabolcs M, Keniry M, Simpson L, Reid LJ, et al. (2009). Irs2 inactivation suppresses tumor progression in Pten+/- mice. *Am. J. Pathol.* 174: 276-286.

Taichman RS, Loberg RD, Mehra R and Pienta KJ (2007). The evolving biology and treatment of prostate cancer. *J. Clin. Invest.* 117: 2351-2361.

Teixeira MR (2008). Chromosome mechanisms giving rise to the TMPRSS2-ERG fusion oncogene in prostate cancer and HGPIN lesions. *Am. J. Surg. Pathol.* 32: 642-644.

Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310: 644-648.

Toubaji A, Albadine R, Meeker AK, Isaacs WB, et al. (2011). Increased gene copy number of ERG on chromosome 21 but not TMPRSS2-ERG fusion predicts outcome in prostatic adenocarcinomas. *Mod. Pathol.* 24: 1511-1520.

Trapnell C, Pachter L and Salzberg SL (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25: 1105-1111.

Wang ET, Sandberg R, Luo S, Khrebtukova I, et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456: 470-476.

Williams SA, Xu Y, De Marzo AM, Isaacs JT, et al. (2010). Prostate-specific antigen (PSA) is activated by KLK2 in prostate cancer *ex vivo* models and in prostate-targeted PSA/KLK2 double transgenic mice. *Prostate* 70: 788-796.

Wu TD and Nacu S (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26: 873-881.

Xiao L, Zhu XZ, Wang Y, Gong Y, et al. (2011). TMPRSS2-ERG gene fusion in metastatic prostate cancers: a study of fine needle aspiration specimens. *Zhonghua Bing Li Xue Za Zhi* 40: 392-396.

Yin L, Rao P, Elson P, Wang J, et al. (2011). Role of TMPRSS2-ERG gene fusion in negative regulation of PSMA expression. *PLoS One* 6: e21319.