European Review for Medical and Pharmacological Sciences 2014; 18: 1354-1360

Transcriptome profiling of prostate tumor and matched normal samples by RNA-Seq

W. ZHAI, X.-D. YAO, Y.-F. XU, B. PENG, H.-M. ZHANG, M. LIU, J.-H. HUANG, G.-C. WANG, J.-H. ZHENG

Department of Urology, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, China

Abstract. – BACKGROUND: RNA-Sequencing (RNA-Seq) has greatly influenced cancer researches, and it provides an unprecedented resolution in estimating gene expression and has less signal noises compared to cDNA microarray.

AIM: We aimed to identify a list of protein-coding genes and lincRNAs that are expressed differentially between tumor and normal tissues.

MATERIALS AND METHODS: In this study, we analyzed including 10 human prostate tumor tissues and their matched normal tissues transcriptome dataset generated by recently developed RNA-Seq technology.

RESULTS: By aligning short reads to human RefSeq genes and lincRNAs, we identified 10 RefSeq genes that were differentially expressed between tumor and normal samples with a p-value < 0.05, which were sufficiently enough to distinguish these two groups. Further loosing the p-value cutoff to 0.1 identified an lincRNA which is antisense to Cullin-associated and neddulation-dissociated 1 (CAND1), whose expression is repressed in prostate tumor cells. By examining the expression of CAND1 and its antisense lincRNA in the transcriptome dataset, we found an interaction between them as high expression of CAND1 and low expression of lincRNA is normal samples, and verse visa in tumor samples.

CONCLUSIONS: These findings suggest the important usage of RNA-Seq in cancer studies for biomarker development and functional investigation.

Keywords

RNA-Seq technology, Prostate tumor, RefSeq genes, lincRNA.

Introduction

Prostate cancer is one of the most prevalent male malignancies in developed countries[1], with ~ 220,000 new diagnosed cases and 32,000 deaths occurred in 2010[2]. Although the underlying mechanism for prostate cancer still remains largely unknown, a number of factors, such as genetic background, age, ethnicity and a family history have an impact on the risk of this cancer[3]. In the last ten years, there are a lot of studies that demonstrated gene expression profiling is a useful approach to identify prostate cancer candidate genes that behave differently between tumor tissues and adjacent normal tissues[4-6]. The main method used in these studies is cDNA microarray, which, however, is subject to various sources of noise. Array effect, efficiency in reverse transcription and hybridization, etc., could introduce variability to the result and complicate later on analysis[7-9], which may result in a certain proportion of false positive result.

The rapid development of next generation sequencing (NGS) technologies in recent years has greatly influenced cancer researches, as it becomes available to perform a whole transcriptome profiling at an affordable cost. This method, termed as RNA-Seq, provides an unprecedented resolution in estimating gene expression and has less signal noises compared to cDNA microarray[10,11]. Hence, it is commonly applied to identify differentially expressed genes[12,13], though RNAseq can also be used to detect fusion genes in tumor tissue[14,16], allele-specific expression[17,18].

It has been displayed that a large proportion of human genome is transcribed. Besides long protein coding mRNA and short regulatory RNAs, a new class of RNA has been characterized as long non-coding RNA (lincRNA), which is longer than 200 base pairs[19-21]. Similar to short regulatory RNAs, lincRNAs have been showed to be functioning during cellular development and their dysfunctions are suggested in prostate cancer development[22,23].

Here we presented a study that used RNAseq to characterize the transcriptome profile of coding genes and lincRNAs in a cohort of prostate cancer samples. By comparing the expression pattern to adjacent normal tissues, we identified a list of protein-coding genes and lincRNAs that

Corresponding Author: Jun-Hua Zheng, MD; e-mail: jacky_zw2002@hotmail.com

are expressed differentially between tumor and normal tissues, which could be helpful for further functional dissection or biomarker development.

Materials and Methods

Transcriptome sequencing data

We retrieved from European Nucleotide Archive (ENA) the transcriptome sequencing data of 10 prostate cancer and matched benign prostate tissues (C02, C03, C06, C08, C09, C11, C13, C15, C19, C23, N02, N03, N06, N08, N09, N11, N13, N15, N19, and N23), where the accession number is SRP002628. According to the original study[24], all samples were radical prostatectomy tissue and obtained from the Baylor Prostate Specialized Programs of Research Excellence (SPORE) Tissue Core. Samples were collected from fresh radical prostatectomy specimens and informed consent was obtained under an institutional review board-approved protocol. Before sample processing, the pathological status was further validated, and all tumor samples were reported to have >80% of tumor cells with Gleason scores of 6-9. Then RNAs were extracted from each sample to prepare paired-end sequencing libraries and were sequenced by Illumina Genome Analyzer II.

Human coding gene and lincRNA transcripts

We downloaded transcripts of the Reference Sequence (RefSeq) genes under current human genome assembly (hg19) from http://genome.ucsc.edu/. This gene set is manually curated and represents a comprehensive, integrated, non-redundant, well-annotated set of sequences. Sequences for human lincRNAs were also downloaded from http://genome.ucsc.edu/, using the "lincRNA Transcripts" track.

Estimate of transcript expression

To estimate the transcription level of protein-coding genes and lincRNAs, we first mapped RNA-Seq reads back to the reference genome (hg19) by TopHat[25], which is a fast splice junction mapper using the ultra high-throughput short read aligner Bowtie[26]. We provided a combined annotation file of both RefSeq genes and lincRNAs, and set the minimal intron size as 20[27], and remained all other parameters as default. Next, we feed TopHat result to Cufflinks[28] with transcript annotation file, and estimated the relative abundances of each transcript by counting the number of reads that mapped to the genomic location of that transcript. Biases in library preparation have been taken into account and the transcription level is measured by the number of fragments per kilobase of transcript per million fragments mapped (FPKM)[28].

Identification of differentially expressed RefSeq genes and lincRNAs

To find genes and lincRNAs which are differentially expressed between tumor and normal samples, we used t-test with unequal variation. Multiple-test correction was performed using Benjamini and Hochberg's method[29].

Supervised clustering analysis

Hierarchical clustering of gene expression was performed by R. For each row (gene), we subtracted each value by the mean of the row and then divided by the standard deviation. Distance between samples was calculated by Euclidean method, and clusters were decided by the complete linkage method, which identifies similar clusters.

Functional annotation and pathway analysis

We used the Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics resource to annotate gene functions and pathways[30-31]. Enrichment tests were performed using tools embedded in DAVID.

Results

Data summary

The prostate cancer RNA-Seqtranscriptome dataset contains 347 million reads in 10 tumor samples and 10 normal samples, averaging ~17.3 million reads in each sample. To align these reads back to the human reference genome, we used TopHat[25], which can efficiently map reads that span exon junctions is thus specifically help in RNA-Seq analysis. The default parameter set was used, except for setting a small intron size as 20 bp[27]. About 311 million reads were successfully aligned to the human reference genome, and the alignment percentage for each sample ranges from 82% to 93.8% (Table I). There is no discrepancy observed for alignment percentage between tumor and normal samples, so all samples were included in further expression analysis.

Table I. Read alignment summary.

SampleID	Tumor			Normal		
	Total (Million reads)	Aligned (Million reads)	Ratio	Total	Aligned	Ratio
2	10.75	9.70	0.90	8.17	7.36	0.90
3	8.10	7.34	0.91	8.01	7.25	0.91
6	13.14	12.08	0.92	8.00	7.17	0.90
8	8.38	7.34	0.88	5.29	4.41	0.83
9	7.56	6.63	0.88	5.64	4.93	0.88
11	21.99	18.03	0.82	29.52	25.89	0.88
13	31.06	28.55	0.92	29.50	26.59	0.90
15	32.55	30.54	0.94	28.47	25.67	0.90
19	32.61	30.59	0.94	23.83	21.10	0.89
23	30.43	28.55	0.94	29.35	27.25	0.93

Expression analysis

To estimate the expression level for a given transcript, we next processed the TopHat alignment result using Cufflinks[28], which derives likelihood for the abundance of given transcripts from aligned reads based on a statistical model of paired-end sequencing experiments, and gives a fragment per kilobase of exon per million (FPKM) value for each transcript. For reads that were mapped to more than one transcript, a weighted number is given for counting the final expression level. To avoid bias, we combined coding gene transcripts and lincRNA transcripts and estimated the expression at the same time. Among 60723 transcripts in total, 18914 (31.1%) transcripts have a FPKM value > 1 in more than 10 samples, which were used as the working dataset to identify differentially expressed transcripts. After multiple test correction by Benjamini and Hochberg's method[29], we found 10 transcripts with a p-value < 0.05, all of which were RefSeq genes. Supervised clustering analysis showed that these genes were sufficient to distinguish tumor and normal samples (Figure 1). Since the resulting number is small and may lack of statistical power in further analysis, we decided to loose the p-value cutoff to 0.1, which led to 55 differentially expressed transcripts, with 50 RefSeq genes and 5 lincRNAs, and were either up-regulated or down-regulated in tumor samples compared to normal samples (Figure 2). And all downstream analyses were performed on transcripts with $p < 0.1$.

Functional and pathway analysis of differentially expressed RefSeq genes

To understand the significance of these statistically differentially expressed RefSeq genes, we used DAVID tools to annotate genes and tested for enrichment. We first interrogated the enrichment of gene ontology (GO) categories of these 50 RefSeq

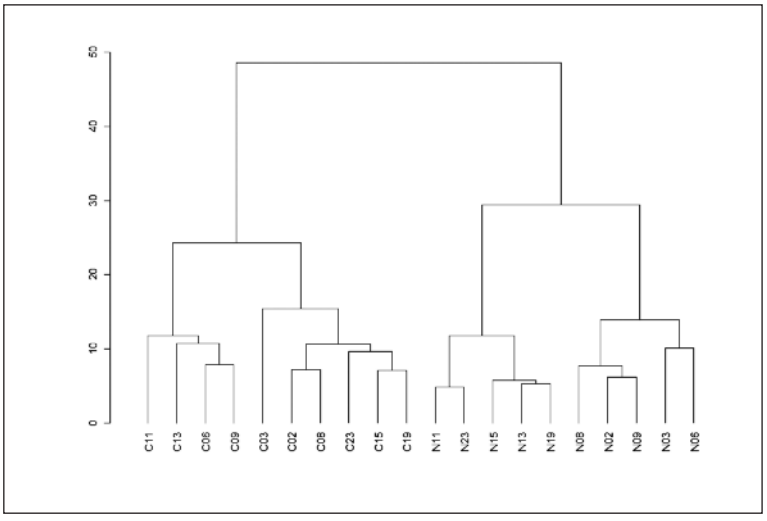

Figure 1. Supervised clustering based on differentially expressed genes. Samples starting with 'C' are tumor samples, and samples starting with 'N' are normal samples. 10 most significantly differentiated RefSeq genes (with p-value < 0.05) were used for this supervised clustering, and tumor and normal samples are clearly separated as two distinguish clusters.

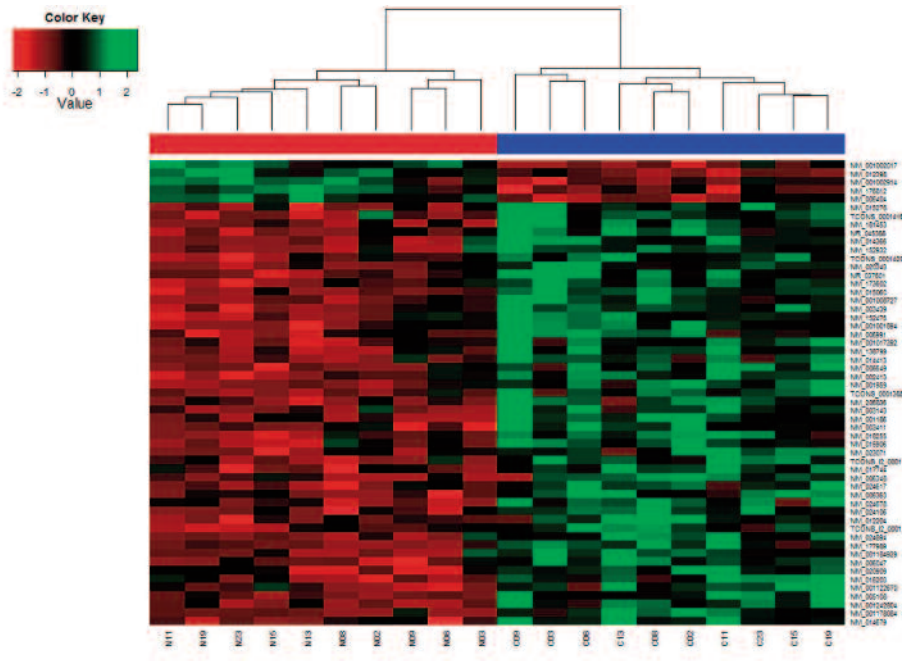

Figure 2. Heatmap of gene expression profile. 55 most differentially expressed RefSeq genes and lincRNAs (with *p*-value < 0.1) were demonstrated. For each gene, expression signals in samples were normalized to obey a standard normal distribution. Color bars above the heatmap represented sample groups, as red for normal and blue for tumor. The dendrogram above was constructed the same as Figure 1. Gene IDs were plotted on the right of the heatmap.

genes. 174 terms of biological process were found among these RefSeq genes, and the most enriched terms are transport related, such as "vesicle-mediated transport" and "Golgi vesicle transport", however, none of the terms reached the statistical significance level ($p < 0.05$). In comparison, there were 68 terms of molecular function found in these RefSeq genes, and the most significant ones are related to binding, such as nucleic acid binding, nucleotide binding and ribonucleotide binding. Although no terms reached a statistical significance, probably due to the small number of genes used here, we still found some interesting cases. On top of this list is the term "translation factor activity, nucleic acid binding", genes in this category were reported to play a critical role in ovarian cancer development[32], suggesting that the differentially expressed genes we identified here may also be related to cancer. We further investigated pathway annotations of our differentially expressed genes, and 4 pathways were identified. Interestingly, 2 of 4 identified pathways were signaling pathway ("Adipocytokine signaling pathway" and "Insulin signaling pathway"), both were annotated as pathways in cancer by Kyoto Encyclopedia of Genes and Genomes (KEGG) and reported to be involved in different types of cancers[33-36]. Due to the data limitation, these pathways were not shown to be statistically enriched in our candidate genes, but our detailed analysis on roles of these pathways in cancer development may suggest certain genes in our list are cancer related.

Functional analysis of differentially expressed lincRNAs

As lincRNAs are commonly involved in tumor development, it is also of interest to study the functional consequence of the 5 lincRNAs that were differentially expressed between tumor and normal samples in this study. We first used Basic Local Alignment Search Tool (BLAST)[37] to search for similar hits of our candidate lincRNAs in a comprehensive non-coding RNA sequence database (http://www.ncrna.org/frnadb/index.html), using an E-value 0.0001, and found hits for 2 lincRNAs (TCONS_00013855 and TCONS_l2_00018657). Further investigation showed that both hits were annotated as human antisense RNAs, and one (TCONS_l2_00018657 OR *FR321700* IN ncRNA database) is overlapped with cullin-associated and neddylation-dissociated 1 (*CAND1*). A few recent studies have shown that *CAND1* gene plays an important role in prostate cancer carcinogenesis, as its expression frequently suppressed in prostate cancer[38-39]. In our dataset, we found the expression of the *CAND1* antisense RNA *FR321700* is higher in tumor samples compared to normal samples, and the expression of *CAND1* is reversed, with a significantly different *CAND1*/*FR321700* ratio between two groups (Figure 3). Considering that antisense RNAs regulate can couple with their sense RNAs and thus down-regulate their expression, the over-expression of *FR321700* in tumor samples is consistent with previous findings, and

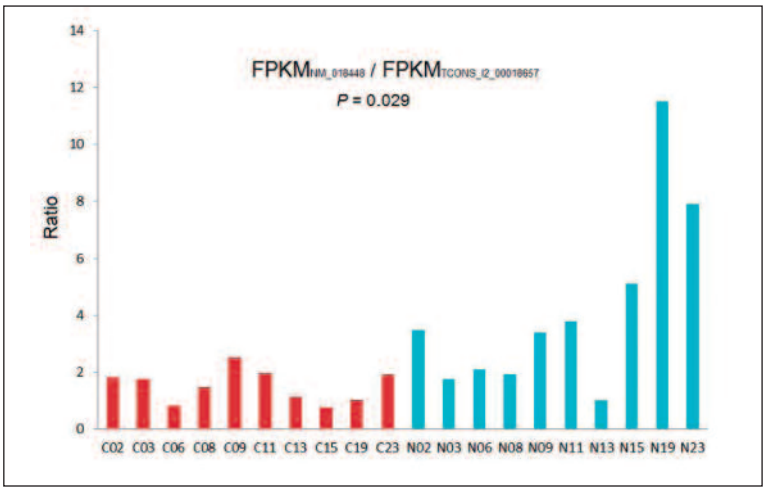

Figure 3. The expression pattern of CAND1 and its antisense lincRNA. The ratio was plotted as the expression of *CAND1* versus its lincRNA. Tumor samples were represented as red bars, and normal samples were represented as blue bars. Expression was estimated as FPKM value, and t-test was used to compare these two groups with the assumption of unequal variation.

suggests an important role that lincRNAs play in human prostate cancer, as well as other cancers.

Discussion

Prostate cancer, along with other cancers, is caused by a series of genetic and environmental factors. Thus, a better understanding could be achieved for the genetic mechanism of prostate cancer by detailed genetic analyses. For a long time, whole-gene expression profiling has been used for molecular diagnosis, clinical outcome prediction or candidate gene identification[40-41]. However, data generated from micro-arrays are subject to several limitations, such as signal saturation, hybridization efficiency or array effects, which introduce additional difficulties to data analysis and may mask informative signals. Recently, RNA-Seq technology has provided a superior method to characterize the transcriptional activities of an entire set of genes, and is widely applied in cancer studies[13,23,42]. In this study, we sought to characterize whole-gene expression pattern between 10 prostate tumor samples and their matched normal samples, using data generated by RNA-Seq. By carefully examining the dataset, we found most genes were not differentially expressed between these two groups, but there were weak signals for handful genes, which could clearly distinguish tumors from normal controls. Further functional dissection also found that these genes were involved in functional categories or pathways related to carcinogenesis, suggesting that these genes may play a possible role in tumor development and may worth further investigation.

Emerged recently and still largely unknown, lincRNAs have become a new aspect of current biological research. Among limited knowledge, lincRNAs have been shown to be an essential contributor to numerous systems and play an important role in cancer biology[43-45]. By cooperatively repressing epigenetic gene expression through chromatin-modification mechanisms[46,47], lincRNAs can interact with known cancer genes in tumorigenesis[45]. Due to the advantage of RNA-Seq, information of most lincRNAs is preserved in the dataset. Therefore, we further compared the expression pattern of more than ten thousands lincRNAs available in University of California Santa Cruz (UCSC) database between tumor and normal populations. At a relatively low statistical level, we identified a few lincRNAs with distinguishing expression pattern. By comparing to a list of annotated lincRNAs, 2 lincRNAs were identified as antisense RNA. Of them, one is complementary to *CAND1* gene, which has been reported as an important gene in prostate cancer development. In one study, Murata et al[39] found that growth of prostate cells is promoted when the expression of *CAND1* is repressed. And in our dataset, we found that the *CAND1* expression is abundant and the expression of its corresponding lincRNA is scarce in normal samples, while in tumor samples an opposite pattern is observed, which is consistent with previous findings and confirm the functional role of lincRNAs in cancer biology.

Declaration of interest

All authors have no conflict of interest to declare.

References

1) Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011; 61: 69-90.
2) Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin 2010; 60: 277-300.
3) Schaid DJ. The complex genetic epidemiology of prostate cancer. Hum Mol Genet 2004; 13(Spec 1): R103-121.
4) Chandran UR, Dhir R, Ma C, Michalopoulos G, Becich M, Gilbertson J. Differences in gene expression in prostate cancer, normal appearing prostate tissue adjacent to cancer and prostate tissue from cancer free organ donors. BMC Cancer 2005; 5: 45.
5) Dakhova O, Ozen M, Creighton CJ, Li R, Ayala G, Rowley D, Ittman M. Global gene expression analysis of reactive stroma in prostate cancer. Clin Cancer Res 2009; 15: 3979-3989.
6) Ernst T, Hergenhahn M, Kenzelmann M, Cohen CD, Bonrouhi M, Weninger A, Klären R, Gröne EF, Wiesel M, Güdemann C, Küster J, Schott W, Staehler G, Kretzler M, Hollstein M, Gröne HJ. Decrease and gain of gene expression are equally discriminatory markers for prostate carcinoma: a gene expression analysis on total and microdissected prostate tissue. Am J Pathol 2002; 160: 2169-2180.
7) Jaluria P, Konstantopoulos K, Betenbaugh M, Shiloach J. A perspective on microarrays: current applications, pitfalls, and potential uses. Microb Cell Fact 2007; 6: 4.
8) Quackenbush J. Computational analysis of microarray data. Nat Rev Genet 2001; 2: 418-427.
9) Simon R, Radmacher MD, Dobbin K. Design of studies using DNA microarrays. Genet Epidemiol 2002; 23: 21-36.
10) Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 2011; 12: 87-98.
11) Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009; 10: 57-63.
12) Zhang LQ, Cheranova D, Gibson M, Ding S, Heruth DP, Fang D, Ye SQ. RNA-seq reveals novel transcriptome of genes and their isoforms in human pulmonary microvascular endothelial cells treated with thrombin. PLoS One 2012; 7: e31229.
13) Ren S, Peng Z, Mao JH, Yu Y, Yin C, Gao X, Cui Z, Zhang J, Yi K, Xu W, Chen C, Wang F, Guo X, Lu J, Yang J, Wei M, Tian Z, Guan Y, Tang L, Xu C, Wang L, Tian W, Wang J, Yang H, Sun Y. RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings. Cell Res 2012; 22: 806-821.
14) Ju YS, Lee WC, Shin JY, Lee S, Bleazard T, Won JK, Kim YT, Kim JI, Kang JH, Seo JS. A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. Genome Res 2012; 22: 436-445.
15) Kohno T, Ichikawa H, Totoki Y, Yasuda K, Hiramoto M, Nammo T, Sakamoto H, Tsuta K, Furuta K, Shimada Y, Iwakawa R, Ogiwara H, Oike T, Enari M, Schetter AJ, Okayama H, Haugen A, Skaug V, Chiku S, Yamanaka I, Arai Y, Watanabe SI, Sekine I, Ogawa S, Harris CC, Tsuda H, Yoshida T, Yokota J, Shibata T. KIF5B-RET fusions in lung adenocarcinoma. Nat Med 2012; 18: 375-377.
16) Lee CH, Ou WB, Marino-Enriquez A, Zhu M, Mayeda M, Wang Y, Guo X, Brunner AL, Amant F, French CA, West RB, McAlpine JN, Gilks CB, Yaffe MB, Prentice LM, McPherson A, Jones SJ, Marra MA, Shah SP, van de Rijn M, Huntsman DG, Dal Cin P, Debiec-Rychter M, Nucci MR, Fletcher JA. 14-3-3 fusion oncogenes in high-grade endometrial stromal sarcoma. Proc Natl Acad Sci U S A 2012; 109: 929-934.
17) Gregg C, Zhang J, Butler JE, Haig D, Dulac C. Sex-specific parent-of-origin allelic expression in the mouse brain. Science 2010; 329: 682-685.
18) Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, Dulac C. igh-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science 2010; 329: 643-648.
19) Bernstein E, Allis CD. RNA meets chromatin. Genes Dev 2005; 19: 1635-1655.
20) Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 2009; 458: 223-227.
21) Mattick JS. The genetic signatures of noncoding RNAs. PLoS Genet 2009; 5: e1000459.
22) Fu X, Ravindranath L, Tran N, Petrovics G, Srivastava S. Regulation of apoptosis by a prostate-specific and prostate cancer-associated noncoding gene, PCGEM1. DNA Cell Biol 2006; 25: 135-141.
23) Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, Laxman B, Asangani IA, Grasso CS, Kominsky HD, Cao X, Jing X, Wang X, Siddiqui J, Wei JT, Robinson D, Iyer HK, Palanisamy N, Maher CA, Chinnaiyan AM. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nat Biotechnol 2011; 29: 742-749.
24) Kannan K, Wang L, Wang J, Ittmann MM, Li W, Yen L. Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. Proc Natl Acad Sci U S A 2011; 108: 9172-9177.
25) Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 2009; 25: 1105-1111.
26) Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009; 10: R25.
27) Zhang Q, Edwards SV. The evolution of intron size in amniotes: a role for powered flight? Genome Biol Evol 2012; 4: 1033-1043.

28) Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010; 28: 511-515.

29) Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 1995; 57: 12.

30) Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4: 44-57.

31) Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009; 37: 1-13.

32) Vui-Kee K, Mohd Dali AZ, Mohamed Rose I, Ghazali R, Jamal R, Mokhtar NM. Molecular markers associated with nonepithelial ovarian cancer in formalin-fixed, paraffin-embedded specimens by genome wide expression profiling. Kaohsiung J Med Sci 2012; 28: 243-250.

33) Belfiore A, Frasca F. IGF and insulin receptor signaling in breast cancer. J Mammary Gland Biol Neoplasia 2008; 13: 381-406.

34) Slomiany MG, Black LA, Kibbey MM, Tingler MA, Day TA, Rosenzweig SA. Insulin-like growth factor-1 receptor and ligand targeting in head and neck squamous cell carcinoma. Cancer Lett 2007; 248: 269-279.

35) Weiss JM, Huang WY, Rinaldi S, Fears TR, Chatterjee N, Chia D, Crawford ED, Kaaks R, Hayes RB. IGF-1 and IGFBP-3: Risk of prostate cancer among men in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. Int J Cancer 2007; 121: 2267-2273.

36) Housa D, Housova J, Vernerova Z, Haluzik M. Adipocytokines and cancer. Physiol Res 2006; 55: 233-244.

37) Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215: 403-410.

38) Korzeniewski N, Hohenfellner M, Duensing S. CAND1 promotes PLK4-mediated centriole overduplication and is frequently disrupted in prostate cancer. Neoplasia 2012; 14: 799-806.

39) Murata T, Takayama K, Katayama S, Urano T, Horie-Inoue K, Ikeda K, Takahashi S, Kawazu C, Hasegawa A, Ouchi Y, Homma Y, Hayashizaki Y, Inoue S. miR-148a is an androgen-responsive microRNA that promotes LNCaP prostate cell growth by repressing its target CAND1 expression. Prostate Cancer Prostatic Dis 2010; 13: 356-361.

40) Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. Nat Genet 2003; 33: 49-54.

41) van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002; 415: 530-536.

42) Sinicropi D, Qu K, Collin F, Crager M, Liu ML, Pelham RJ, Pho M, Rossi AD, Jeong J, Scott A, Ambannavar R, Zheng C, Mena R, Esteban J, Stephans J, Morlan J, Baker J. Whole transcriptome RNA-Seq analysis of breast cancer recurrence risk using formalin-fixed paraffin-embedded tumor tissue. PLoS One 2012; 7: e40092.

43) Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T, Rinn JL. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell 2010; 142: 409-419.

44) Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, Guigo R, Shiekhattar R. Long noncoding RNAs with enhancer-like function in human cells. Cell 2010; 143: 46-58.

45) Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 2007; 129: 1311-1323.

46) Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, Xiong Y. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. Oncogene 2011; 30: 1956-1962.

47) Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY. Long noncoding RNA as modular scaffold of histone modification complexes. Science 2010; 329: 689-693.