Comparative analysis of protein-coding genes and long non-coding RNAs of prostate cancer between Caucasian and Chinese populations

Fu-Yan Hu
Department of Mathematics,
Institute of Systems Biology,
Shanghai University,
Shanghai 200444,
China

Xing-Ming Zhao
Institute of Systems Biology,
Shanghai University,
Shanghai 200444,
China

Nelson L.S. Tang
Department of Chemical Pathology,
Faculty of Medicine,
The Chinese University of Hong Kong,
Hong Kong SAR,
China

Yan Zhang*
Key Laboratory of Systems Biology,
Shanghai Institutes for Biological Sciences,
Chinese Academy of Sciences,
Shanghai 200031, China,
*Corresponding author,
e-mail:
yanzhang01@sibs.ac.cn

Luonan Chen*
Key Laboratory of Systems Biology,
Shanghai Institutes for Biological Sciences,
Chinese Academy of Sciences,
Shanghai 200031, China,
*Corresponding author,
e-mail:
lnchen@sibs.ac.cn

Abstract—**Prostate cancer is one of the most important public health problems in developed countries. To date, a systematic understanding of the pathogenesis of prostate cancer is still lacking. In this work, we identified differentially expressed protein-coding genes and long non-coding RNAs (lncRNAs) between normal and cancer tissues based on a recent RNA-seq study from Caucasian population. We then investigated the relationship between differentially expressed genes and lncRNAs. Furthermore, based on a recently published prostate cancer study on Chinese population, we identified differentially expressed genes between Caucasian and Chinese populations to investigate racial difference. Moreover, for the first time, we compared the correlation of lncRNA-gene across populations. In the end, a lot of differentially expressed genes and lncRNAs were identified. Our results revealed that most of the lncRNA-gene pairs were positively correlated especially for the lncRNA-host gene pairs, indicating the probable mechanism of lncRNA. And 320 genes were differentially expressed in prostate cancer across populations, which may help us to investigate the ethnic differences of prostate cancer. In addition, our results suggested that lncRNAs regulate genes in different manners across populations. Our findings may help understand molecular events underlying prostate cancer development.**

***Keywords*—transcriptome; long non-coding RNAs; prostate cancer**

I. INTRODUCTION

Prostate cancer is one of the most common cancers in men [1]. It is also reported that prostate cancer is more prevalent in Western countries than in Asian countries [2, 3]. Many factors, including genetics and diet, have been implicated in the development of prostate cancer. However, little is known about the molecular mechanisms involved in its development and progression. The lack of knowledge on the biology of prostate cancer has resulted in many disputes on the treatment of its early stages and on the utility of population study.

With the rapid development of high-throughput biotechnologies, gene expression microarrays have been used in numerous applications, including measuring the expression levels of large numbers of genes to identify novel genes associated with certain cancers such as prostate cancer. For example, Reams et al. compared the microarray datasets of prostate tumor in African-American and Caucasian American males to explain the reasons of heterogeneity for prostate cancer [4]. Sorensen et al. discovered some promising molecular markers of prostate cancer by gene expression profiling [5]. Despite microarray-based methods have achieved remarkable success, they may have several major defects, e.g., the dependence on precise probe design for all genes. In recent years, RNA-seq has become a revolutionary tool to investigate gene expression, which uses massively parallel sequencing to allow transcriptome analyses of genomes at a much higher resolution than is available with microarray-based methods. It may help people get a better understanding of the pathogenesis of cancer. Montgomery et al. have showed that high throughput sequencing technologies contribute to reveal new properties of genetic effects on the transcriptome [6]. Ha et al. used RNA-seq to discover gene fusions in breast cancer transcriptome [7]. Very recently, Ren et al. used RNA-seq to reveal multiple aspects of the transcriptome of prostate cancer in the Chinese population, including gene fusions, alternative splicing, the expression of genes and long non-coding RNAs (lncRNAs), which provides insights into the pathogenesis of prostate cancer in Chinese [8]. These studies have made great progress in investigating the cancer-related processes.

In the last decades, researchers have focused on 2% of the human genome that codes proteins. However, it may be also important to study the non-protein-coding sequences [9]. Recently, non-coding RNAs have gained more attention as their diverse roles in many cellular processes are discovered. For example, lncRNAs whose mature transcripts are longer than 200-nt have been shown to regulate gene expression and

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

other important biological processes [10, 11]. Moreover, lncRNAs have also been linked to many human diseases including cancer [12, 13], which shed light on the importance of this emergent field.

In this paper, we first chose a previous RNA-seq study on prostate cancer in the Caucasian population and analyzed the expression profiles of both protein-coding genes and lncRNAs. In addition, we investigated the relationship between differentially expressed lncRNAs and genes. We also found the differentially expressed genes between Caucasian and Chinese populations. More importantly, we compared the gene-lncRNA correlation between Caucasian and Chinese populations. Our study revealed that a lot of genes and lncRNAs were differentially expressed, implying that they may participate in the development of the prostate cancer. Most of these lncRNAs were positively correlated with a variety of differentially expressed genes, implying the important regulatory mechanism between lncRNAs and their target genes. 320 genes were found to be differentially expressed between populations. Interestingly, we found that lncRNAs may regulate the same genes in different manners between Caucasian and Chinese populations.

II. MATERIALS AND METHODS

A. RNA-seq resource and lncRNA dataset

In this work, RNA-seq resource was obtained from Gene Expression Omnibus (GEO) with the accession number of GSE22260, which includes 10 prostate cancer samples and 10 matched normal tissues from the Caucasian population. The human lncRNA database was download from NONCODE (http://www.noncode.org/NONCODERv3/download.htm) [14], which contains 33,818 lncRNAs.

B. Estimating expression levels for protein-coding genes and lncRNAs

We first aligned original reads to the human reference genome (hg18) by TopHat (2.0.1) [15] with default parameters. After that, we estimated the expression abundance of protein-coding genes by running Cufflinks [16] in its expression abundance estimation mode across our 20 samples of Caucasian. In the end, gene expression level was measured by FPKM which stands for fragments per kilobase of transcript per million mapped reads.

With regard to the expression of lncRNAs, we used MAQ [17] to map the RNA-seq reads to all known human lncRNAs coming from NONCODE [14]. No more than 2 mismatches were allowed during the alignment for each read. Reads that could be uniquely mapped to lncRNA were used to calculate the RPKM (Reads Per Kilobase of transcript per Million mapped reads) for each lncRNA. Here, RPKM was calculated through the formula as follows:

$$RPKM = \frac{10^6 U}{NL/10^3},$$

where U was the number of reads uniquely mapped to a given lncRNA; N was the number of reads uniquely mapped to all lncRNAs; L was the length of the given lncRNA.

C. Differentially expressed genes and lncRNAs analysis

We used paired samples t-tests to find differentially expressed genes and lncRNAs by comparing the expression levels of genes and lncRNAs in 10 prostate cancer samples with their expression levels in 10 benign prostate tissues.

D. Correlation analysis of lncRNAs and protein-coding genes

We calculated Pearson's correlation coefficient of lncRNA-gene using the expression levels of lncRNAs and genes in 10 prostate cancer samples. The significant lncRNA-gene correlation was defined by the cutoff value of an absolute correlation coefficient as 0.85.

More importantly, for the lncRNA contained in a protein-coding gene, the host protein-coding gene was identified by locating the sequence of the lncRNA on the human reference genome (hg18). After we got the lncRNA-host gene pairs, Pearson's correlation coefficient was used to explore the relationship between these pairs.

E. Identification of differentially expressed genes between Caucasian and Chinese populations

First, we identified the non-differentially expressed genes by comparing their expression levels in normal Caucasian population samples with their expression levels in normal Chinese population samples through t-test. And then we detected the differentially expressed genes by comparing their expression levels in cancer Caucasian population samples with their expression levels in cancer Chinese population samples with the same method. Finally, we focused on the overlap of the two gene sets.

F. Comparing the gene-lncRNA correlation between Caucasian and Chinese populations

We used lncRNAs that were differentially expressed in both populations to analyze their correlation with the genes that were differentially expressed in both populations. The expression values of lncRNAs and genes in all cancer samples were used to calculate the Pearson's correlation coefficient between each lncRNA and gene for both populations. After we got the coefficient matrix, we used the clustergram function of matlab to create a heat map with dendrograms to show hierarchical clustering of coefficient matrix.

III. RESULTS AND DISCUSSION

A. Differentially expressed genes in prostate cancer

We used paired samples t-test to find statistically differentially expressed genes by comparing their expression levels between 10 pairs of prostate cancer samples and benign prostate tissues from Caucasian population. 2369 genes were identified as significantly differentially expressed genes in prostate cancer ($p<0.01$), with prostate cancer genes CD82, the decreased expression of which may result in the malignant progression of prostate cancer [18].

Furthermore, we investigated the functions that were enriched in the 2369 differentially expressed genes, where the

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

biological process category from Gene Ontology was considered [19, 20]. Table I lists the top 20 most enriched functions for these genes.

TABLE I. FUNCTION ENRICHMENT ANALYSIS FOR THE 2369 DIFFERENTIALLY EXPRESSED GENES.

Function	*P*-value	Count
transcription	7.62E-09	335
vesicle-mediated transport	2.48E-06	108
regulation of transcription	3.11E-06	384
protein amino acid phosphorylation	5.83E-05	115
actin filament-based process	7.58E-05	51
cell junction organization	7.92E-05	19
actin cytoskeleton organization	1.16E-04	48
intracellular transport	1.18E-04	112
cell-substrate junction assembly	1.63E-04	11
cell junction assembly	1.93E-04	15
endosome transport	3.40E-04	18
RNA biosynthetic process	3.77E-04	57
Golgi organization	3.98E-04	10
transcription, DNA-dependent	4.74E-04	56
phosphorus metabolic process	5.37E-04	152
phosphate metabolic process	5.37E-04	152
phosphorylation	8.37E-04	127
modification-dependent protein	1.11E-03	95
modification-dependent macromolecule catabolic process	1.11E-03	95
proteolysis involved in cellular protein catabolic process	2.07E-03	97

Beyond that, the enriched pathways were detected for the 2369 differentially expressed genes in prostate cancer (Table II). Among these enriched pathways, eleven pathways which were highlighted in bold (Table II) were reported as key pathways involved in prostate cancer by Ning [21]. Especially, the abnormal of the insulin signaling pathway may confer risk to prostate cancer [22]; Focal adhesion kinase controls prostate cancer progression[23]; The members of MAPK signaling pathway may become potential therapeutic targets for prostate cancer [24].

TABLE II. PATHWAY ENRICHMENT ANALYSIS OF THE 2369 DIFFERENTIALLY EXPRESSED GENES.

Pathway	*P*-value	Count
Insulin signaling pathway	8.72E-05	32
Focal adhesion	5.20E-04	40
Tight junction	9.70E-04	29
Vascular smooth muscle contraction	1.53E-03	25
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	1.80E-03	19
Sphingolipid metabolism	1.03E-02	11
Glycerophospholipid metabolism	1.95E-02	15
Gap junction	2.19E-02	18
Propanoate metabolism	2.42E-02	9
Adherens junction	2.57E-02	16
ECM-receptor interaction	2.64E-02	17
Pathways in cancer	2.69E-02	50
Hypertrophic cardiomyopathy (HCM)	2.92E-02	17
Dilated cardiomyopathy	2.95E-02	18
Melanogenesis	2.97E-02	19
MAPK signaling pathway	4.14E-02	41
Vibrio cholerae infection	4.91E-02	12

B. *Differentially expressed analysis of lncRNAs*

To better understand the roles played by lncRNAs in prostate cancer, we analyzed the expressions of 33,818 lncRNAs in 10 pairs of prostate cancer and benign prostate tissues, and detected an average of 13605 lncRNAs (range from 9321 to 16489) that were expressed in each sample. We used paired samples t-test to find statistically differentially expressed lncRNAs by comparing the expressions of lncRNAs between 10 pairs of prostate cancer and benign prostate tissues, and indentified 733 differentially expressed lncRNAs ($p<0.01$).

C. *The relationship between lncRNAs and genes*

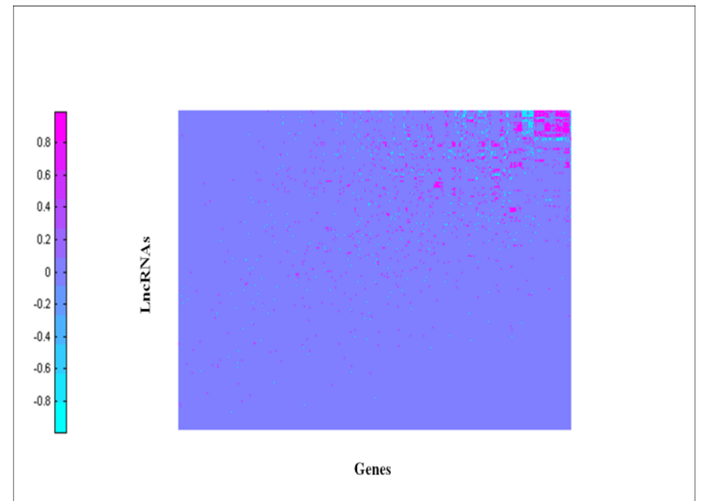

Figure 1. Correlation heatmap between the expression of lncRNAs and genes. Rows represent differentially expressed lncRNAs, and the columns represent differentially expressed genes. A pink color means a positive correlation, while blue-green indicates a negative correlation (absolute correlation coefficient $|R| > 0.85$). The absolute values less than 0.85 were set to zero, using blue color to represent them; Only the correlation coefficients with absolute values beyond 0.85 were kept.

Since lncRNAs have been reported to influence gene expression by a variety of mechanisms, we investigated the expression relationships between lncRNAs and genes. We used the expression values of 733 differentially expressed lncRNAs in each cancer sample to analyze their correlation with 2369

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

differentially expressed genes. A total of 680 lncRNAs and 2178 genes showed statistically significant correlation (absolute correlation coefficient $|R| > 0.85$) (Fig. 1). There were 25095 positively correlated lncRNA-gene pairs and 10367 negative correlated lncRNA-gene pairs respectively. The result is consistent with what was observed in prostate cancer of Chinese population [8], implying that lncRNAs may activate the expression of their target genes.

To further investigate the relationships between lncRNAs and genes, 361 host protein-coding genes were identified for 361 differentially expressed lncRNAs that are contained by protein-coding genes. Among the 361 host protein-coding genes, 212 genes were also differentially expressed. Moreover, Pearson's correlation coefficients were used to analysis the relationship of these 212 lncRNA-host gene pairs. In normal samples, there were 146 lncRNA-host gene pairs showed statistically significant correlation (absolute correlation coefficient $|R| > 0.85$) and all these pairs were positively correlated. At the same time, there were 127 lncRNA-host gene pairs showed statistically significant correlation (absolute correlation coefficient $|R| > 0.85$) in cancer samples and all these pairs were also positively correlated. It is easy to find that no matter in normal or cancer samples the correlations of lncRNA-host gene pairs tended to be positively correlated, which imply that such lncRNAs might act as enhancer-associated lncRNAs. Although, looking at the numbers, there was no big difference between the statistically significant correlated lncRNA-host gene pairs in normal samples and cancer samples, some pairs showed different correlations in the two conditions. Three criteria were used to identify these pairs. The criteria are: (1) The absolute correlation coefficients of lncRNA-host gene pairs were bigger than 0.85 in one condition; (2) The absolute correlation coefficients of lncRNA-host gene pairs were less than 0.85 in the other condition; (3) The difference of the absolute correlation coefficients in the two conditions was bigger than 0.5. Based on the three criteria, 20 lncRNA-host gene pairs were indentified. Among the 20 host genes, 10 genes have been linked to cancer (Table Ⅲ), indicating that these lncRNA-host gene pairs may play an important roles in the development of prostate cancer.

TABLE III. 10 HOST GENES THAT HAVE BEEN LINKED TO CANCER.

Gene names	Tumour Types	Reference (PMID)
MDM4	GBM; bladder; retinoblastoma	17499002; 20863813; 18335186
POTEKP	basel cell carcinomas	134174
KCNC4	head and neck squamous cell carcinomas; prostate cancer	20593490; 14632936
ACP1	noonan syndrome; diabetes autoimmune; immune disorder; insulin resistance	15240615; 18056891; 18490013; 18925540
ATCAY	ataxia; cerebellar; aayman type	14556008; 8845847; 8595404
NKD1	non-small-cell lung cancer; colorectal cancer	21599923; 19956716; 15498874
PDE4A	pulmonary disease chronic obstructive; chronic obstructive asthma; asthma	11964752; 15912967; 19135348
HOXA10	leukemia; myeloid leukemia; infertility	20022444; 10766757; 19736237
CKAP5	breast cancer; hepatomas; colonic tumors	14603251; 8832653
ACTB	Baraitser-Winter syndrome; aystonia; juvenile-onset	10344733; 15111311; 15555571

D. *Differentially expressed genes between Caucasian and Chinese populations*

We indentified 1729 non-differentially expressed genes in normal samples and 896 differentially expressed genes in cancer samples across Caucasian and Chinese populations. The overlap of the two gene sets contained 320 genes which may influence the prostate cancer in different manners across populations. Furthermore, pathway enrichment analysis on the 320 genes showed that Ubiquitin mediated proteolysis (*P*-value =1.0E-2) and mTOR signaling pathway (*P*-value=1.3E-2) are significantly enriched. Ubiquitin-mediated proteolysis plays an important role in the degradation of proteins [25], the dysregulation of which would possibly result in the development and metastasis of cancers. Interestingly, mTOR signaling pathway has been reported as a novel therapeutic molecular targets for prostate cancer [26].

TABLE IV. THE FUNCTION ANNOTATIONS OF 14 GENES.

Gene names	Functions
FAM50A	molecular function
GTF3C4	contributes_to DNA binding; enzyme activator activity; histone acetyltransferase activity; protein binding
HSP90AB2P	ATP binding；unfolded protein binding
LOC646214	unknown
MOB2	protein binding
NAF1	RNA binding; protein binding; snoRNA binding
NKTR	cyclosporin A binding; peptidyl-prolyl cis-trans isomerase activity
NTN1	protein binding
PJA1	protein binding; ubiquitin-protein ligase activity; zinc ion binding
RARRES2	receptor binding
REXO1L1	exonuclease activity; nucleic acid binding
TRPM7	ATP binding; actin binding; calcium channel activity; metal ion binding; myosin binding; protein serine/threonine kinase activity
WFDC1	molecular function; serine-type endopeptidase inhibitor activity
ZNF814	nucleic acid binding; zinc ion binding

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

Besides, there were 14 genes that were differentially expressed in both populations among the 320 genes, which implied that the 14 genes may play central roles in prostate cancer for both populations. Table Ⅳ lists the detailed functions of the 14 genes.

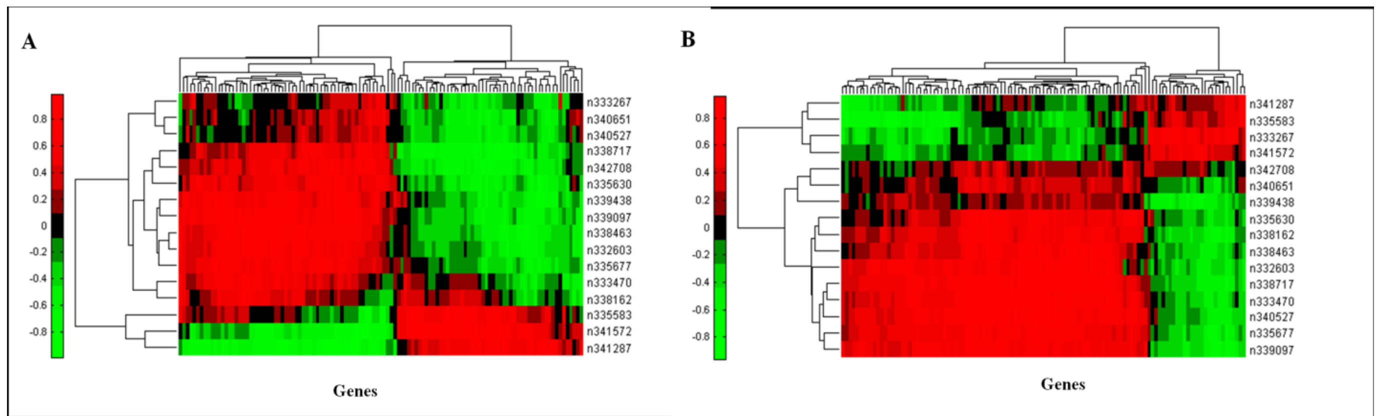

Figure 2. Correlation analysis of lncRNAs and genes for both populations. (A) Correlation analysis of lncRNAs and genes for Caucasian population. (B) Correlation analysis of lncRNAs and genes for Chinese population.

E. Comparing analysis between Caucasian and Chinese populations

Based on the RPKM values of 3185 genes and 2981 lncRNA in 14 pairs of prostate cancer and adjacent normal tissues from Chinese population reported by Ren et al. [8], we identified 587 differentially expressed genes and 157 differentially expressed lncRNAs by using paired samples t-test ($p<0.01$). By comparing the differentially expressed genes and lncRNAs between the Caucasian and Chinese populations, we found that 115 genes and 16 lncRNAs were differentially expressed in both populations. We then analyzed the lncRNA-gene correlationship by calculating the Pearson's correlation coefficient in each population. The heat map with dendrograms of the coefficient matrix for each population is shown in Fig. 2. As shown in Fig. 2, 16 differentially expressed lncRNAs can be divided into two groups for both populations. It is interested that the members in each group for both populations were almost the same except for one lncRNA i.e. n333267. Although the lncRNAs can be divided in the same way for both populations, it is still easy to find that correlationships of lncRNA-gene pairs were different in Caucasian and Chinese populations, which implied that lncRNAs regulate genes in different manners across populations.

IV. CONCLUSION

In this paper, we first found a lot of differentially expressed protein-coding genes and lncRNAs based on a recent RNA-seq study from Caucasian population. Furthermore, we explored the relationship between these differentially expressed lncRNAs and genes. Moreover, we identified 320 differentially expressed genes between Caucasian and Chinese populations. Our results suggested that most of the differentially expressed lncRNAs were positively correlated with differentially expressed genes. We also found that lncRNAs may regulate genes in different manners across populations.

ACKNOWLEDGMENT

This work was partly supported by the Innovation Program of Shanghai Municipal Education Commission (10YZ01), Shanghai Rising-Star Program (10QA1402700), National Natural Science Foundation of China (61103075, 91130032, 31171233)，Shanghai Pujiang Program (11PJ1410500), Chief Scientist Program of SIBS from CAS (2009CSP002), the Knowledge Innovation Program of CAS (2011KIP203, KSCX2-EW-R-01), National Center for Mathematics and Interdisciplinary Sciences of CAS.

REFERENCES

[1] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," CA Cancer J Clin, vol. 61, pp. 69-90, Mar-Apr 2011.

[2] H. G. Sim and C. W. Cheng, "Changing demography of prostate cancer in Asia," Eur J Cancer, vol. 41, pp. 834-45, Apr 2005.

[3] T. Kimura, "East meets West: ethnic differences in prostate cancer epidemiology between East Asians and Caucasians," Chin J Cancer, Nov 15 2011.

[4] R. R. Reams, D. Agrawal, M. B. Davis, S. Yoder, F. T. Odedina, N. Kumar, J. M. Higginbotham, T. Akinremi, S. Suther, and K. F. Soliman, "Microarray comparison of prostate tumor gene expression in African-American and Caucasian American males: a pilot project study," Infect Agent Cancer, vol. 4 Suppl 1, p. S3, 2009.

[5] K. D. Sorensen and T. F. Orntoft, "Discovery of prostate cancer biomarkers by microarray gene expression profiling," Expert Rev Mol Diagn, vol. 10, pp. 49-64, Jan 2010.

[6] S. B. Montgomery, M. Sammeth, M. Gutierrez-Arcelus, R. P. Lach, C. Ingle, J. Nisbett, R. Guigo, and E. T. Dermitzakis, "Transcriptome genetics using second generation sequencing in a Caucasian population," Nature, vol. 464, pp. 773-7, Apr 1 2010.

[7] K. C. Ha, E. Lalonde, L. Li, L. Cavallone, R. Natrajan, M. B. Lambros, C. Mitsopoulos, J. Hakas, I. Kozarewa, K. Fenwick, C. J. Lord, A. Ashworth, A. Vincent-Salomon, M. Basik, J. S. Reis-Filho, J. Majewski, and W. D. Foulkes, "Identification of gene fusion transcripts by transcriptome sequencing in BRCA1-

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

mutated breast cancers and cell lines," BMC Med Genomics, vol. 4, p. 75, 2011.

[8] S. Ren, Z. Peng, J. H. Mao, Y. Yu, C. Yin, X. Gao, Z. Cui, J. Zhang, K. Yi, W. Xu, C. Chen, F. Wang, X. Guo, J. Lu, J. Yang, M. Wei, Z. Tian, Y. Guan, L. Tang, C. Xu, L. Wang, W. Tian, J. Wang, H. Yang, and Y. Sun, "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, vol. 22, pp. 806-21, May 2012.

[9] C. A. Brosnan and O. Voinnet, "The long and the short of noncoding RNAs," Curr Opin Cell Biol, vol. 21, pp. 416-25, Jun 2009.

[10] X. Wang, S. Arai, X. Song, D. Reichart, K. Du, G. Pascual, P. Tempst, M. G. Rosenfeld, C. K. Glass, and R. Kurokawa, "Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription," Nature, vol. 454, pp. 126-30, Jul 3 2008.

[11] T. R. Mercer, M. E. Dinger, and J. S. Mattick, "Long non-coding RNAs: insights into functions," Nat Rev Genet, vol. 10, pp. 155-159, 2009.

[12] S. A. Mitra, A. P. Mitra, and T. J. Triche, "A central role for long non-coding RNA in cancer," Front Genet, vol. 3, p. 17, 2012.

[13] O. Wapinski and H. Y. Chang, "Long noncoding RNAs and human disease," Trends Cell Biol, vol. 21, pp. 354-61, Jun 2011.

[14] D. Bu, K. Yu, S. Sun, C. Xie, G. Skogerbo, R. Miao, H. Xiao, Q. Liao, H. Luo, G. Zhao, H. Zhao, Z. Liu, C. Liu, R. Chen, and Y. Zhao, "NONCODE v3.0: integrative annotation of long noncoding RNAs," Nucleic Acids Res, vol. 40, pp. D210-5, Jan 2012.

[15] C. Trapnell, L. Pachter, and S. L. Salzberg, "TopHat: discovering splice junctions with RNA-Seq," Bioinformatics, vol. 25, pp. 1105-11, May 1 2009.

[16] C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, and L. Pachter, "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation," Nat Biotechnol, vol. 28, pp. 511-5, May 2010.

[17] H. Li, J. Ruan, and R. Durbin, "Mapping short DNA sequencing reads and calling variants using mapping quality scores," Genome Res, vol. 18, pp. 1851-8, 2008.

[18] J. T. Dong, P. W. Lamb, C. W. Rinker-Schaeffer, J. Vukanovic, T. Ichikawa, J. T. Isaacs, and J. C. Barrett, "KAI1, a metastasis suppressor gene for prostate cancer on human chromosome 11p11.2," Science, vol. 268, pp. 884-6, May 12 1995.

[19] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources," Nature Protoc, vol. 4, pp. 44-57, 2009.

[20] W. Huang da, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists," Nucleic Acids Res, vol. 37, pp. 1-13, Jan 2009.

[21] Q. Y. Ning, J. Z. Wu, N. Zang, J. Liang, Y. L. Hu, and Z. N. Mo, "Key pathways involved in prostate cancer based on gene set enrichment analysis and meta analysis," Genet Mol Res, vol. 10, pp. 3856-87, 2011.

[22] A. W. Hsing, L. C. S. Sakoda, and J. Chua, "Obesity, metabolic syndrome, and prostate cancer," Am. J. Clin. Nutr, vol. 86, pp. s843-s857, 2007.

[23] S. Figel and I. H. Gelman, "Focal adhesion kinase controls prostate cancer progression via intrinsic kinase and scaffolding functions," Anticancer Agents Med Chem, vol. 11, pp. 607-16, Sep 2011.

[24] G. Rodriguez-Berriguete, B. Fraile, P. Martinez-Onsurbe, G. Olmedilla, R. Paniagua, and M. Royuela, "MAP Kinases and Prostate Cancer," J Signal Transduct, vol. 2012, p. 169170, 2012.

[25] A. Ciechanover, A. Orian, and A. L. Schwartz, "Ubiquitin-mediated proteolysis: biological regulation via destruction," Bioessays, vol. 22, pp. 442-51, May 2000.

[26] A. W. Tolcher, "Novel therapeutic molecular targets for prostate cancer: the mTOR signaling pathway and epidermal growth factor receptor," J Urol, vol. 171, pp. S41-3; discussion S44, Feb 2004.

978-1-4673-4398-5/12/$31.00 ©2012 IEEE