Biochip analysis of prostate cancer

M.Q. Fan, P.X. Wang, J.Y. Feng, Y. Xiao and C.B. Huang

Department II of Urology, XinQiao Hospital,
Third Military Medical University, Chongqing, China

Corresponding author: C.B. Huang
E-mail: chenhaigechg@hotmail.com

ABSTRACT. Microarray expression analysis was used to forecast the roles of differentially co-expressed genes (DCG) and DCG and links in the pathogenesis of prostate cancer. In addition, we demonstrate that the relationship between transcriptional factors (TFs) and their targets can be considered a key factor in determining the difference between primary and metastatic prostate cancer. Regulatory impact factors were adopted to calculate the impact of TF. We identified 5 TFs and 29 target genes important in the transition between normal prostate and primary prostate cancer and 2 TFs and 7 target genes important in the transition between primary and metastatic prostate cancer. These results suggest that it may be possible to predict the clinical behavior of prostate cancer based on gene expression analysis.

Key words: Prostate cancer; Differentially co-expressed genes (DCG); DCG and links; Transcription factor; Regulatory impact factor

INTRODUCTION

It is a challenge to predict the prognosis in prostate carcinoma even though it is the cancer most frequently diagnosed. Prostate cancer, a leading cause of cancer death, displays a broad range of clinical behavior, from relatively benign to aggressive metastatic disease (Hsing et al., 2000). Its incidence and mortality vary in different parts of the world, but these values are the highest in western countries, with an estimated 1,479,350 new cancer cases and 562,340 deaths in the United States in 2009 (Jemal et al., 2009). Despite the high death rate, prostate cancer is often a slow-developing disease, and patients can remain asymptomatic for years. Early diagnosis provides an opportunity for curative surgery. The challenge is to identify methods for detecting the symptoms of prostate cancer and the risk of relapse.

Several clinical features of prostate cancer, including tumor stage (Jewett, 1975), degree of tumor cell differentiation [Gleason score (Gleason, 1966)], and serum prostate-specific antigen (PSA; Stamey et al., 1987), are used in routine clinical prognosis. However, these prognostic indicators do not accurately predict clinical outcomes for individual patients. Microarray profiling studies have identified clinically relevant gene-expression subtypes in leukemia (Ramaswamy et al., 2001), lymphoma (Alizadeh et al., 2000), breast cancer (Perou et al., 2000), and lung cancer (Chen et al., 2003). Because there are few DNA microarray studies of prostate cancer, tumor subtypes based on gene expression have not been defined.

Recently, genomic methodologies have been used to discover consistent gene expression patterns associated with a given histological or clinical phenotype (Golub et al., 1999; Perou et al., 2000; Van't Veer et al., 2002). Here, gene expression patterns of tissues from normal prostate, primary prostate cancer, and metastatic prostate cancer were studied in order to analyze changes in biochemical pathways possibly leading to primary and metastatic prostate cancer. Combining this with studies of the regulation of transcription factors and their target genes, we could find those that assist in providing a more accurate picture of the likely course of the disease. Our data suggest that it may be possible to predict the clinical behavior of prostate cancer based upon gene expression analysis.

MATERIAL AND METHODS

Gene expression profiling analysis

One transcription profile of GSE3325 (Su et al., 2011) was obtained from GEO, a public functional genomic data repository (http://www.ncbi.nlm.nih.gov/geo/). This profile is based on Affymetrix GPL571 platform data (Affymetrix Human Genome U133A 2.0 Array). Only 5 primary prostate cancer chips, 4 metastatic prostate cancer chips, and control chips are available.

In this paper, we used differentially co-expressed genes (DCGs) and DCG links (DCGL) to analyze differentially expressed genes. Gene co-expression analysis was developed to explore gene interconnections at the expression level from a system perspective, and differential co-expression analysis, which examines the change in gene expression correlation between two conditions, was accordingly designed as a complementary technique to traditional differential expression analysis (Liu et al., 2010; Yu et al., 2011). The third part, the core of the study, includes five methods for identifying DCGs and DCGLs, which mainly differ in the measure of differential co-expression of a gene. After gene filtration and link filtration, suppose gene i is associated with n_i links whose co-expression values are projected to X = {x_{i1},

x$_{i2}$,..., x$_{ini}$} and Y = {y$_{i1}$, y$_{i2}$,..., y$_{ini}$} for two conditions. The differential co-expression measures of different methods are given in the following equations.

$$dC_i = \sqrt{\frac{(x_{i1} - y_{i1})^2 + (x_{i2} - y_{i2})^2 + ... + (x_{in} - y_{in})^2}{n}}$$ (Equation 1)

In this equation, x_{in} represents the correlation value of gene i and gene 1 at state 1, and y_{in} represents the correlation value of gene i and gene 1 at state 1.

Data analysis of transcription factors (TFs) and TF target genes

In our paper, we used human transcription factors and their chromosomal regions from the UCSC database (http://genome.ucsc.edu) to annotate the message download from the database. The relationships between TF and target genes were obtained after analysis.

Data analysis of biological pathways

In order to understand the metabolic changes occurring in prostate cancer cells and the normal prostate, we obtained the metabolic pathways through the KEGG PATHWAY DATABASE and used the DAVID website to analyze the differentially expressed genes by the KEGGPATHWAY cluster method (Alberts, 1990; Alonso et al., 2009).

Regulatory impact factor (RIF) calculations

When determining if the differential co-expression links were TF-target gene links, or if the target genes were differentially expressed genes, we could make further analyses: If the TF were a DCG, but its target gene was a non-DCG, it suggests that the TF and target gene have synergistic effect. If the TF were a non-DCG, but its target gene was a DCG, it suggests that the TF may interact with its target gene by changing its conformation as, for example, when the TF is phosphorylated to achieve its active state: although the expression does not change, it can lead to target gene profile fluctuations. In this paper, in order to more clearly show the selected impact of the TF, we adopt RIF (Reverter et al., 2010) to calculate the impact, as shown in Equation 2.

$$RIF_i = \frac{1}{n_{de}} \sum_{j=1}^{j=n_{de}} [(e1_j \times r1_{ij})^2 - (e2_j \times r2_{ij})^2]$$ (Equation 2)

In this equation, n_{de} represents the number of differentially expressed genes, the factor $e1_j$ represents the expression value of differentially expressed genes, and $r1_{ij}$ represents the correlation coefficient between differentially expressed genes i and j at state 1. Similarly, the factor $e2_j$ represents the expression value of differentially expressed genes, and $r2_{ij}$ represents the correlation coefficient between differentially expressed genes i and j at state 2.

RESULTS

Selecting the differentially expressed genes

With DCGL, we made a differential co-expression analysis of normal prostate sam-

ples versus primary prostate cancer samples (benign *vs* primary) and metastatic prostate cancer samples *vs* primary prostate cancer samples (primary *vs* metastatic) to obtain the P value and differentially co-expressed link (DCL). Then, we set q < 0.25 as the threshold of DCG, and DCL, as shown in Table 1. It was predicted that DCG and DCL play important roles in disease progression to primary or metastatic prostate cancer.

Table 1. Number of differentially expressed genes between benign, primary and metastatic cancer.

	Benign *vs* primary	Primary *vs* metastatic
DCG	1258	997
DCL	475023	339863

DCG = differentially co-expressed genes; DCL = differentially co-expressed link.

Data analysis of TF and TF target genes

The above differentially expressed genes compared with their TF and its target gene, and the regulatory relationship between TF and the target gene are shown in Table 2. This regulatory relationship was considered a key factor in causing the difference between primary and metastatic prostate cancer.

In our paper, comparisons between normal prostate, primary and metastatic prostate cancer are shown in Figure 1.

Table 2. Number of differentially expressed genes and the regulatory relationship between benign, primary and metastatic cancer.

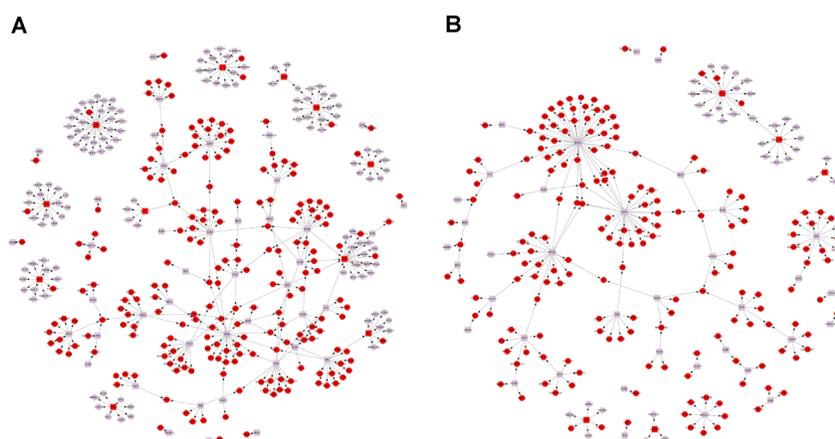|	Benign *vs* primary	Primary *vs* metastatic
TF→target gene	384	252
TF	54	43

TF = transcriptional factor.

Figure 1. Comparison of differentially expressed genes. **A.** Benign *vs* primary cancer. **B.** Primary *vs* metastatic cancer. Circular points = targets genes; square points = transcriptional factor genes; red points = differentially co-expressed genes (DCGs); gray points = non-DCGs.

Analysis of TF impact

From the above analysis, the relationship between TFs and target genes was preliminarily considered as the last step leading to differential expression. To further characterize the size of the impact on TFs and target genes, RIF were adopted to calculate the maximal impact of TF. Results are shown in Tables 3 and 4.

Table 3. Regulatory impact factor (RIF) of transcriptional factor (TF) between normal prostate and primary prostate cancer samples.

TF	RIF score	Rank
HLF	2452584.845	1
ZNF423	2426499.614	2
LMO2	2423023.814	3
FOXF2	1765570.618	4
BPTF	1748870.266	5

Table 4. Regulatory impact factor (RIF) of transcriptional factor (TF) between primary prostate cancer and metastatic cancer samples.

TF	RIF score	Rank
JUND	4507642	1
FOSB	3974856	2
FOXO3	3212461	3
MZF1	2841575	4
EGR1	2122338	5

The maximum values, which result in the greatest differences in gene expression between normal prostate and primary prostate cancer, are displayed in Figure 2A and include values for 5 TFs and 29 target genes. Similarly, the maximum values for gene expression changes between primary and metastatic prostate cancer are displayed in Figure 2B and include values for 2 TFs and 7 target genes.
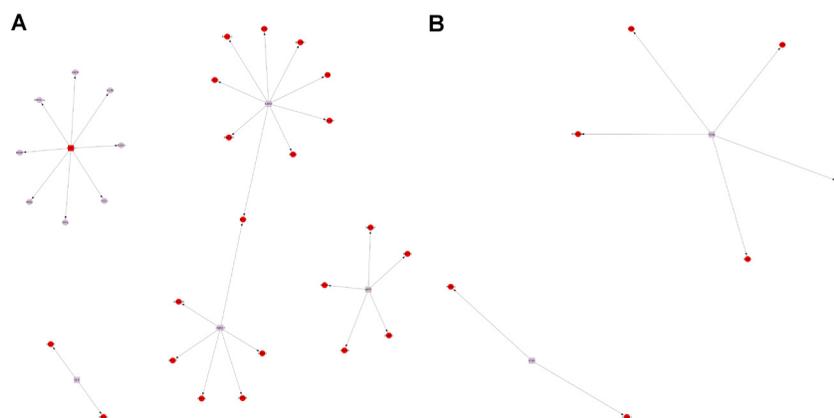

Figure 2. Regulatory impact factor of transcriptional factor (TF) between normal prostate, primary prostate and metastatic cancer samples. **A.** Benign *vs* primary cancer. **B.** Primary *vs* metastatic cancer. Circular points = targets genes; square points = TF genes; red points = differentially co-expressed genes (DCGs); gray points = non-DCGs.

DISCUSSION

The main objective of our study was to survey molecular variations associated with prostate cancer to gain new insight into changes in biochemical pathways that might possibly lead to development of the disease. This was combined with studies of the regulation of TFs and their target genes in the hope of finding some that could be used prognostically. The data suggest that it may be possible to predict the clinical behavior of prostate cancer based on gene expression analysis.

Log ratio of connections is the logarithm of the ratio of the calculated connectivities of a gene between two conditions (Reverter et al., 2006). Weighted gene co-expression network analyses are increasingly used to explore the system-level functionality of genes whose construction is conceptually straightforward: nodes represent genes, and nodes are connected if the corresponding genes are significantly co-expressed across appropriately chosen tissue samples (Mason et al., 2009; Van Nas et al., 2009). Furthermore, the association of a specific gene's expression, such as platelet-derived growth factor receptor (PDGFR), with its outcome raises the possibility that expression analysis may prove useful in selecting patients for emerging mechanism-based therapeutics. In contrast, our method, differential co-expression profile, is based on the exact co-expression relationship between a TF and its targets and identification of co-expression reversal between normal prostate, primary and metastatic prostate cancer. All the methods are able to identify DCGs from microarray datasets.

Since 1988, several investigators have used these clinical parameters, such as biopsy Gleason scores, clinical stage and PSA, to stratify patients into risk groups (low, intermediate, high) and to predict clinical outcomes (nomograms) (Kattan et al., 1998; D'Amico et al., 2000; Partin et al., 2001). Microarray-based tumor classification (Golub et al., 1999; Welsh et al., 2001), as well as treatment response and clinical outcome prediction (Dhanasekaran et al., 2001; Van't Veer et al., 2002), have been employed for many cancer types. In this study, we have further characterized the expression and prognostic value of differential expression. The expression of PDGFRβ is associated with the growth of metastatic prostate tumor cells, which conformed the hypothesis that PDGFRβ contributes to prostate cancer progression (Chott et al., 1999). Attempts to build a gene expression-based predictor of recurrence of prostate tumor following prostatectomy led to build a model that can correctly predict the outcome of DCG and DCL, which play important roles in the progression of primary or metastatic prostate cancer.

Multivariate analysis was performed to examine whether these results were independent of the Gleason score, a well-established histological grading system for prostate cancer (Gleason, 1992). Bibikova et al. (2007) used the expression signature to predict disease relapse. They searched the candidate genes from publicly reported lists of genes differentially expressed in prostate cancer. Therefore, our article served as a detailed extension of previous studies. This study demonstrated that the relationship between TFs and their targets can be considered a key factor determining the difference between primary and metastatic prostate cancer.

It is well known that co-transcriptional regulation (or gene amplification) of PIM1 and MYC possibly mediate a synergistic oncogenic effect on prostate cancer (Buttyan et al., 1987; Abate-Shen and Shen, 2000). Three E-twenty-six (ETS) transcription factors *AR*, *ERG* and *ETV1* were identified as outliers in prostate cancer across prostate cancer-profiling studies, and these results implicate the important role of TFs and their target genes in molecular diagnosis and treatment (Tomlins et al., 2005; Wang et al., 2007). Consistent with the roles

of GATA2 and Oct1 in target gene expression (Tsai and Orkin, 1997), our data indicate the functions of the maximum values of gene expression in normal prostate, primary and metastatic prostate cancer.

Recently, microRNA (miRNA) expression-profiling studies have identified cancer-specific miRNA signatures (Akao et al., 2006; Cummins et al., 2006). Surprisingly, very little is known about the regulation of expression. Transcriptional regulation is complex and varied. In addition to TFs as a part of transcriptional regulation, miRNA was considered as an important common and efficient mechanism (Volinia et al., 2006). Consistent with the roles of GATA2 and Oct1 in androgen target gene expression, silencing these two collaborating transcription factors significantly decreased androgen-induced cell cycle progression (Tsai and Orkin, 1997). If we can select an expression microarray for miRNA, the pathogenesis of prostate cancer may be revealed at the transcriptional level.

In conclusion, we predicted that DCG and DCL play important roles in disease progression to primary or metastatic prostate cancer. In addition, the regulatory relationship was considered a key step in the pathogenesis of prostate cancer. By making direct comparative hybridizations for normal and neoplastic tissues, we identified 5 TFs and 29 target genes important in the transition from normal cells to primary prostate cancer and 2 TFs and 7 target genes important in the transition between primary and metastatic prostate cancer.

REFERENCES

Abate-Shen C and Shen MM (2000). Molecular genetics of prostate cancer. *Genes Dev.* 14: 2410-2434.

Akao Y, Nakagawa Y and Naoe T (2006). MicroRNAs 143 and 145 are possible common onco-microRNAs in human cancers. *Oncol. Rep.* 16: 845-850.

Alberts AW (1990). Lovastatin and simvastatin - inhibitors of HMG CoA reductase and cholesterol biosynthesis. *Cardiology* 77 (Suppl 4): 14-21.

Alizadeh AA, Eisen MB, Davis RE, Ma C, et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403: 503-511.

Alonso R, Defesche JC, Tejedor D, Castillo S, et al. (2009). Genetic diagnosis of familial hypercholesterolemia using a DNA-array based platform. *Clin. Biochem.* 42: 899-903.

Bibikova M, Chudin E, Arsanjani A, Zhou L, et al. (2007). Expression signatures that correlated with Gleason score and relapse in prostate cancer. *Genomics* 89: 666-672.

Buttyan R, Sawczuk IS, Benson MC, Siegal JD, et al. (1987). Enhanced expression of the c-myc protooncogene in high-grade human prostate cancers. *Prostate* 11: 327-337.

Chen G, Gharib TG, Wang H, Huang CC, et al. (2003). Protein profiles associated with survival in lung adenocarcinoma. *Proc. Natl. Acad. Sci. U. S. A.* 100: 13537-13542.

Chott A, Sun Z, Morganstern D, Pan J, et al. (1999). Tyrosine kinases expressed *in vivo* by human prostate cancer bone marrow metastases and loss of the type 1 insulin-like growth factor receptor. *Am. J. Pathol.* 155: 1271-1279.

Cummins JM, He Y, Leary RJ, Pagliarini R, et al. (2006). The colorectal microRNAome. *Proc. Natl. Acad. Sci. U. S. A.* 103: 3687-3692.

D'Amico AV, Whittington R, Malkowicz SB, Schultz D, et al. (2000). Clinical utility of the percentage of positive prostate biopsies in defining biochemical outcome after radical prostatectomy for patients with clinically localized prostate cancer. *J. Clin. Oncol.* 18: 1164-1172.

Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, et al. (2001). Delineation of prognostic biomarkers in prostate cancer. *Nature* 412: 822-826.

Gleason DF (1966). Classification of prostatic carcinomas. *Cancer Chemother. Rep.* 50: 125-128.

Gleason DF (1992). Histologic grading of prostate cancer: a perspective. *Hum. Pathol.* 23: 273-279.

Golub TR, Slonim DK, Tamayo P, Huard C, et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286: 531-537.

Hsing AW, Tsao L and Devesa SS (2000). International trends and patterns of prostate cancer incidence and mortality. *Int. J. Cancer* 85: 60-67.

Jemal A, Siegel R, Ward E, Hao Y, et al. (2009). Cancer statistics, 2009. *CA Cancer J. Clin.* 59: 225-249.

Jewett HJ (1975). The present status of radical prostatectomy for stages A and B prostatic cancer. *Urol. Clin. North Am.* 2: 105-124.

Kattan MW, Eastham JA, Stapleton AM, Wheeler TM, et al. (1998). A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J. Natl. Cancer Inst.* 90: 766-771.

Liu BH, Yu H, Tu K, Li C, et al. (2010). DCGL: an R package for identifying differentially coexpressed genes and links from gene expression microarray data. *Bioinformatics* 26: 2637-2638.

Mason MJ, Fan G, Plath K, Zhou Q, et al. (2009). Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics* 10: 327.

Partin AW, Mangold LA, Lamm DM, Walsh PC, et al. (2001). Contemporary update of prostate cancer staging nomograms (Partin Tables) for the new millennium. *Urology* 58: 843-848.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, et al. (2000). Molecular portraits of human breast tumours. *Nature* 406: 747-752.

Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, et al. (2001). Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci. U. S. A.* 98: 15149-15154.

Reverter A, Ingham A, Lehnert SA, Tan SH, et al. (2006). Simultaneous identification of differential gene expression and connectivity in inflammation, adipogenesis and cancer. *Bioinformatics* 22: 2396-2404.

Reverter A, Hudson NJ, Nagaraj SH, Perez-Enciso M, et al. (2010). Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics* 26: 896-904.

Stamey TA, Yang N, Hay AR, McNeal JE, et al. (1987). Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. *N. Engl. J. Med.* 317: 909-916.

Su H, Hu N, Yang HH, Wang C, et al. (2011). Global gene expression profiling and validation in esophageal squamous cell carcinoma and its association with clinical phenotypes. *Clin. Cancer Res.* 17: 2955-2966.

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.

Tsai FY and Orkin SH (1997). Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood* 89: 3636-3643.

van Nas A, Guhathakurta D, Wang SS, Yehya N, et al. (2009). Elucidating the role of gonadal hormones in sexually dimorphic gene coexpression networks. *Endocrinology* 150: 1235-1249.

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415: 530-536.

Volinia S, Calin GA, Liu CG, Ambs S, et al. (2006). A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. U. S. A.* 103: 2257-2261.

Wang Q, Li W, Liu XS, Carroll JS, et al. (2007). A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Mol. Cell* 27: 380-392.

Welsh JB, Sapinoso LM, Su AI, Kern SG, et al. (2001). Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Res.* 61: 5974-5978.

Yu H, Liu BH, Ye ZQ, Li C, et al. (2011). Link-based quantitative methods to identify differentially coexpressed genes and gene pairs. *BMC Bioinformatics* 12: 315.