

RESEARCH ARTICLE

Identifying Differentially Expressed Genes and Small Molecule Drugs for Prostate Cancer by a Bioinformatics Strategy

Jian Li, Ya-Hong Xu*, Yi Lu, Xiao-Ping Ma, Ping Chen, Shun-Wen Luo, Zhi-Gang Jia, Yang Liu, Yu Guo

Abstract

Purpose: Prostate cancer caused by the abnormal disorderly growth of prostatic acinar cells is the most prevalent cancer of men in western countries. We aimed to screen out differentially expressed genes (DEGs) and explore small molecule drugs for prostate cancer. **Materials and Methods:** The GSE3824 gene expression profile of prostate cancer was downloaded from Gene Expression Omnibus database which including 21 normal samples and 18 prostate cancer cells. The DEGs were identified by Limma package in R language and gene ontology and pathway enrichment analyses were performed. In addition, potential regulatory microRNAs and the target sites of the transcription factors were screened out based on the molecular signature database. In addition, the DEGs were mapped to the connectivity map database to identify potential small molecule drugs. **Results:** A total of 6,588 genes were filtered as DEGs between normal and prostate cancer samples. Examples such as ITGB6, ITGB3, ITGAV and ITGA2 may induce prostate cancer through actions on the focal adhesion pathway. Furthermore, the transcription factor, SP1, and its target genes ARHGAP26 and USF1 were identified. The most significant microRNA, MIR-506, was screened and found to regulate genes including ITGB1 and ITGB3. Additionally, small molecules MS-275, 8-azaguanine and pyvinium were discovered to have the potential to repair the disordered metabolic pathways, and furthermore to remedy prostate cancer. **Conclusions:** The results of our analysis bear on the mechanism of prostate cancer and allow screening for small molecular drugs for this cancer. The findings have the potential for future use in the clinic for treatment of prostate cancer.

Keywords: Prostate cancer - differentially expressed genes (DEGs) - enrichment analysis - small molecules

Asian Pac J Cancer Prev, 14 (9), 5281-5286

Introduction

As a kind of malignancy, prostate cancer is resulting from the pathological changes in men's prostate tissue (DeMarzo et al., 2003). The incidence of prostate cancer has significant geographic and racial differences. In Europe, the United States and other developed countries, prostate cancer is the most common male malignancy, moreover, its mortality rate is in the second place among the various cancers (de Bono et al., 2011). In Asia, its incidence is lower than western countries, but rising rapidly in recent years (Zhang et al., 2011). Therefore, the research and treatment of prostate cancer have a great significance to human health.

In recent years, considerable researches are undertaken around prostate cancer. The researches about the molecular genetics and growth regulation of the prostate cancer have made remarkable progress (Singh et al., 2012). For example, large-scale genomic rearrangements and gene mutation found in prostate cancer may be the molecular mechanism of the cancer (Porkkaand Visakorpi, 2004). Abundant of reports about the regulatory functions of transcription factors and microRNAs in cancer have

been published. Nevertheless, they mostly focused on a certain target or DNA. Therefore, the previous studies are inefficient and inadequate to explore the mechanism of cancer. Additionally, sex-steroid hormone, androgens and estrogens, play regulatory roles in the progression of prostate cancer (Boorjian and Tindall, 2013). Meanwhile, there are some small molecules which can treat prostate cancer, such as CPIC (1-(3-(2-chlorophenoxy) propyl)-1H-indole-3-carbonitrile), AR54 ((2-(pyrimidin-2-ylthio)-1-(2,2,4-trimethyl-4-phenyl-3,4-dihydroquinolin-1(2H)-yl)ethanone) (Shapiro et al., 2012). However, these molecules don't treat the prostate cancer efficiently and selectively. Therefore, the need for new methods to elucidate the mechanism of prostate cancer and new therapeutics for prostate cancer is highlighted.

In this paper, microarrays were utilized to identify differentially expressed genes (DEGs) between cancer and normal prostate cells. Significance of differential expression was tested by Limma and adjusted for multiple testing with the Benjamini and Hochberg (BH) procedure. The functions of DEGs were investigated by Gene Ontology (GO) and pathway enrichment analysis. Additionally, several target sites of the transcription

Department of Urology, the 452nd Hospital of PLA, Chengdu, China *For correspondence: xuyahong452@hotmail.com

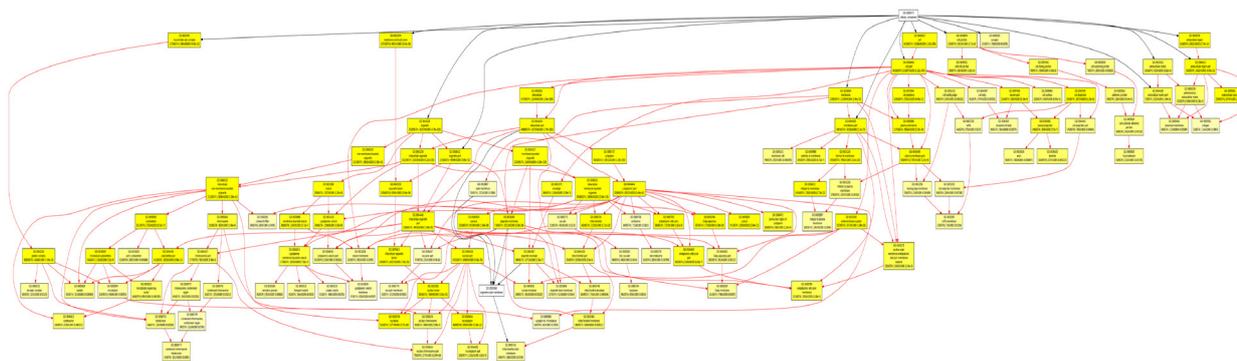


Figure 1. The Enriched Gene Ontology (GO) Terms of the Cellular Compartment of the Differentially Expressed Genes (DEGs). The colored entries are the significant aggregation (FDR < 0.05) ones, and the deeper colored entries represent the more significant ones

factors and some regulatory microRNAs were screened out, helping us to elucidate the mechanism of the prostate cancer on a molecular level. In addition, candidate small molecules were identified for their potential use in the treatment of prostate cancer.

Materials and Methods

Derivation of genetic data

The gene expression profile of GSE38241 (Aryee et al., 2013) were downloaded from a public functional genomics data repository GEO (Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo/>) database. Total 39 specimens, including 21 normal samples and 18 prostate cancer specimens, were available based on the GPL4133Platform.

DEGs analysis

We analyzed the derived genetic data by Geoquery package and Limma package in R language (v.2.13.0) (Team, 2011). Geoquery can quickly access the expression profiling data on the GEO database, while Limma is the most popular method in the statistical analysis to study the DEGs (Diboun et al., 2006; Smyth, 2004). The preprocessed microarray data were obtained by Geoquery package then performed log₂ transformation. We applied the Limma package, a linear regression model, to compare the normal samples and prostate cancer samples. Only the genes with *p*-value < 0.05 were screened out as DEGs.

GO enrichment analysis

GO analysis has become a common approach for functional annotation of large-scale genomic data (Hulsegge et al., 2009). GOEAST (Gene Ontology Enrichment Analysis Software Toolkit) an easy-to-use web-based toolkit that identifies statistically overrepresented GO terms within given gene sets (Zheng and Wang, 2008).

GOEAST was utilized for GO enrichment analysis to identify the locations of DEGs in cellular compartments and the DEGs affected molecular functions, based on the hypergeometric distribution, with the false discovery rate (FDR) less than 0.05.

Biological pathway enrichment analysis

Biological pathways were studied to explore the

prostate cancer cells changes on the molecular level. All metabolic and non-metabolic pathways were downloaded from the open WikiPathways (Pico et al., 2008; Kelder et al., 2012) database, and conducted Wikipathways cluster analysis (Zhang et al., 2005; Duncan et al., 2010) to the DEGs by Gene Set Analysis Toolkit V2 platform. The count number larger than 2 and *p*-value less than 0.00001 were chosen as cut-off criterion.

Exploring potential target sites of transcription factors and potential regulatory microRNAs

Well-annotated gene sets in MsigDB (Molecular Signature Database, <http://www.broadinstitute.org/gsea/msigdb/index.jsp>) were conducted Gene Set Enrichment analysis (GSEA) (Subramanian et al., 2005). Then the GSEA results were statistically accounted with the hypergeometric distribution. Afterwards, the consequences were adjusted for multiple testing with BH. Finally, the target sites with FDR < 0.05 were selected as the potential target sites, which can regulate transcription factors. Similarly, the potential regulatory microRNAs were identified with FDR < 0.05.

Identification of candidate small molecules

CMap (The connectivity map) database contains 7056 gene-expression profiles data involving 6100 small molecules treatment-control pairs (Lamb et al., 2006). The DEGs were divided into up- and down-regulated groups. The top 500 genes with a highly significant correlation in each group were screened out. Then, these genes were conducted GSEA and compared with the DEGs in CMap database. Finally, a correlation score for each perturbagen was calculated, ranging from -1 to +1 (Braconi et al., 2010).

Results

Screening DEGs in prostate cancer cells

Based on the Geoquery and Limma package in R language, at a *p*-value of 0.05, a total of 10611 probes were identified to be differentially expressed in prostate cancer samples compared with normal controls, which corresponded to 6588 DEGs.

Gene function annotation

To determine the function of DEGs in prostate cancer,

Table 3. The Enriched Potential Regulatory microRNA

Target sequence	Potential microRNA	Genes	P-value
hsa_GTGCCTT	MIR-506	ITGB1, ITGA3, IL9R, SP1	1.26E-54
hsa_TGCTGCT	MIR-15A, MIR-16, MIR-15B,		
MIR-195, MIR-424, MIR-497	TRIM2, HDGF, IGF2R	5.56E-53	
hsa_GCACTTT	MIR-17-5P, MIR-20A, MIR-106A,		
MIR-106B, MIR-20B, MIR-519D	ITCH, HAS2, KIF5A	2.31E-47	
hsa_TTTGCAC	MIR-19A, MIR-19B	UBL3, CGN, SBF2	8.60E-45
hsa_TGCCTTA	MIR-124A	SP3, ARG2, AAK1	1.06E-41
hsa_ACCAAAG	MIR-9	TRIM2, MAEA, NFKB1	8.50E-41
hsa_TTGCACT	MIR-130A, MIR-301, MIR-130B	PHF3, TRIM2, FAM53B	4.38E-40
hsa_TGAATGT	MIR-181A, MIR-181B, MIR-181C,		
MIR-181D	PHF3, TRIM2, GRIK2	7.14E-40	
hsa_TGTTTAC	MIR-30A-5P, MIR-30C, MIR-30D,		
MIR-30B, MIR-30E-5P	ITGB3, ITS1N1, HSPA5	9.51E-40	
hsa_TGCACCTT	MIR-519C, MIR-519B, MIR-519A	KIF5A, UBR5, HAS2	4.86E-39
hsa_ACTGTGA	MIR-27A, MIR-27B	AGRN, UBR5, LIF	9.48E-38
hsa_TGGTGCT	MIR-29A, MIR-29B, MIR-29C	IGF1, LIF, SBF2	1.67E-37
hsa_AAGCACT	MIR-520F	UBR3, MLL5, APP	7.70E-37
hsa_CTTTGTA	MIR-524	MLL5, SBF2, EIF5	1.30E-36
hsa_ACATTCC	MIR-1, MIR-206	TRIM2, UBR5, IGF1	1.58E-34
hsa_TTGCCAA	MIR-182	ISL1, EIF5, SP3	5.34E-34
hsa_TAGCTTT	MIR-9	MITF, FAR1, RELT	9.48E-33
hsa_CTCAGGG	MIR-125B, MIR-125A	ABR, CGN, BAK1	4.50E-32
hsa_GTGCAAT	MIR-25, MIR-32, MIR-92, MIR-363,		
MIR-367	NLK, SLX4, PPCS	4.69E-32	
hsa_TTTGTAG	MIR-520D	ADD1, IGF1, SP3	1.10E-31

Table 4. The Enriched Significant Small Molecules

CMap name	Enrichment	P-Value
Trichostatin A	-0.485	0
Tanespimycin	-0.431	0
Sirolimus	-0.348	0.00004
Felbinac	0.903	0.0001
Harmol	0.9	0.0001
Amantadine	0.89	0.00014
LY-294002	-0.277	0.00016
Biperiden	0.85	0.00018
Finasteride	0.792	0.00018
Vorinostat	-0.588	0.0002
Benzocaine	0.882	0.00024
MS-275	-0.989	0.00028
Thioridazine	-0.437	0.00046
Thioperamide	0.815	0.00052
Guanabenz	0.811	0.0006
8-azaguanine	-0.857	0.0007
Pyrvinium	-0.736	0.0007
Nadolol	0.852	0.00072
Khellin	0.797	0.00084
Fludrocortisone	0.649	0.0009

Related small molecule drugs screening

In order to screen out small molecule drugs, we performed computational bioinformatics analysis of DEGs using the CMap. Total 20 related small molecules with a highly significant correlation were listed in Table 4, including 9 negatively related molecules and 11 positively related small molecules (Table 4). Among these molecules, MS-275, 8-azaguanine and pyrvinium with the highest negative correlation had the potential to treat the prostate cancer.

Discussion

The clinical heterogeneity and high prevalence of prostate cancer, raises challenges in the management of

newly diagnosed patients (Taylor et al., 2010). Therefore, there is an urgent need to explore the mechanism of prostate cancer, and develop an effective prevention strategy for it. In the present study, we utilized the gene expression profile downloaded from GEO to explore the mechanism of prostate cancer development. Furthermore, small molecule drugs which had the potential to treat the prostate cancer were screened out. Finally, a total of 6588 DEGs were identified between prostate cancer samples and normal controls.

Current strategies typically study entire pathways, whether by singular enrichment analysis or by gene set enrichment analysis (Sherman and Lempicki, 2009). In this study, we identified 49 pathways and screened out focal adhesion as the most significant pathway in the development of prostate cancer. The focal adhesion is a prominent determinant in cancer initiation, progression and metastasis (Luo and Guan, 2010). The genes in integrin family, such as ITGB6, ITGB3, ITGA2 and ITGAV, are closely related to focal adhesion. Among this family, ITGB6 (integrin-beta 6), which mediates the interactions between adjacent cells and between cells and extracellular matrix, is down-regulated during tumor progression (Pontes-Júnior et al., 2009). Meanwhile, ITGB3 and ITGA2 have been found to play important roles in breast cancer and colorectal cancer (Langsenlehner et al., 2006). In addition, ITGAV is related to many cancer types among which prostate and breast cancer for which it is important in the bone environment to the growth and pathogenesis of cancer bone metastases (Daemen et al., 2008). Therefore, our study indicates that these integrin genes located in focal adhesion play crucial roles in cancer initiation, progression and metastasis.

Great deals of reports have declared that abundant of transcription factors are overactive in human cancer cells, which make them targets for the studies of

cancer mechanisms. In this paper, the most significant transcription factor is SP1 (specificity protein 1), which binds to GC/GT-rich promoter elements, such as GGGCGGR. A number of genes, such as ARHGAP26 and USF1, which contain this sequence, can be identified by SP1. ARHGAP26 has been found shown to be expressed in a subset of ovarian cancer tissues at high levels, while it is absent or present only at low levels in normal tissues (Jarius et al., 2013). USF1 (upstream stimulatory factor 1) is a transcriptional suppressor of human telomerase reverse transcriptase in oral cancer cells (Chang et al., 2005).

MicroRNAs are small regulatory RNAs that regulate the translation and degradation of target mRNAs and are extensively involved in human cancers (Fabbri et al., 2007). The most significant microRNA in our study is MIR-506 and its targeting sequence is GTGCCCTT. Genes, such as ITGB1 and ITGA3, which contain this sequence can be regulated by MIR-506. ITGB1 in focal adhesion has been found play crucial roles in biological pathway. Additionally, it also has been discovered can be regulated by MIR-506. Therefore, ITGB1 may play an essential role in the pathogenesis of prostate cancer. Because this finding has not been reported by any other researchers, our study sets a new insight to explore the mechanism of prostate cancer.

Although there are several useful medicines in prostate cancer treatment, such as dutasteride (Andriole et al., 2010) and microRNA miR-34a inhibitor (Liu et al., 2011), these medicines are far from enough for prostate cancer treatment. Recent studies elucidate that small molecules can enhance the efficacy of cancer drugs and treat cancer (Sugahara et al., 2010). In our study, three molecule drugs (MS-275, 8-azaguanine and pyrvinium) were screened out which may effective for prostate cancer treatment. MS-275, a histone deacetylase inhibitor, has been demonstrated to display antiproliferative activity towards several human cancer cell lines, including breast, colorectal, lung, ovarian and pancreatic cancer cells (Altmann et al., 2010). Meanwhile, the incorporation of 8-azaguanine in m-RNA of tumor cells has been found to inhibit protein synthesis and has been implicated as a lead molecule in cancer therapy (Gogiaand Puranik, 2013). Additionally, pyrvinium, a classical anthelmintic, potently inhibited proliferation and STAT3 Tyr705 phosphorylation of human myeloma/erythroleukemia cells (Harada et al., 2012). Consequently, our present findings conform to the previous studies. In conclusion, our studies shed new light on the mechanism and treatment of prostate cancers. DEGs of prostate cancer were analyzed by a computational bioinformatics approach. Meanwhile, the changed biological pathways in cancer cells were identified, and the potential targets of transcription factors and regulatory microRNAs were enriched. Furthermore, three small molecule drugs (MS-275, 8-azaguanine and pyrvinium) capable of treating the prostate cancer were screened. Our research may provide a new strategy in the medical therapy of prostate cancer. Since the increasing public availability of genomic data, we predict that this approach will be an attractive strategy that could be used in many other researches.

Acknowledgements

The author(s) declare that they have no competing interests.

References

- Altmann A, Eisenhut M, Bauder-Wüst U, et al (2010). Therapy of thyroid carcinoma with the histone deacetylase inhibitor MS-275. *Eur J Nucl Med Mol Imaging*, **37**, 2286-97.
- Andriole GL, Bostwick DG, Brawley OW, et al (2010). Effect of dutasteride on the risk of prostate cancer. *N Engl J Med*, **362**, 1192-202.
- Aryee MJ, Liu W, Engelmann JC, et al (2013). DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. *Sci Transl Med*, **5**, 169ra10.
- Boorjian SA, Tindall DJ (2013). Molecular Progression of Prostate Cancer: Androgens and Estrogens. Management of Prostate Cancer. Springer.
- Braconi C, Swenson E, Kogure T, et al (2010). Targeting the IL-6 dependent phenotype can identify novel therapies for cholangiocarcinoma. *PLoS One*, **5**, e15195.
- Chang JTC, Yang HT, Wang TC, Cheng AJ (2005). Upstream stimulatory factor (USF) as a transcriptional suppressor of human telomerase reverse transcriptase (hTERT) in oral cancer cells. *Mol Carcinog*, **44**, 183-92.
- Daemen A, Gevaert O, De Bie T, et al (2008). Integrating microarray and proteomics data to predict the response on cetuximab in patients with rectal cancer. *Pac Symp Biocomput*, **2008**, 166-77.
- de Bono JS, Logothetis CJ, Molina A, et al (2011). Abiraterone and increased survival in metastatic prostate cancer. *N Engl J Med*, **364**, 1995-2005.
- DeMarzo AM, Nelson WG, Isaacs WB, Epstein JI (2003). Pathological and molecular aspects of prostate cancer. *Lancet*, **361**, 955-64.
- Diboun I, Wernisch L, Orengo CA, Koltzenburg M (2006). Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. *BMC Genomics*, **7**, 252.
- Duncan D, Prodduturi N, Zhang B (2010). WebGestalt2: an updated and expanded version of the Web-based Gene Set Analysis Toolkit. *BMC Bioinformatics*, **11**, 1-.
- Fabbri M, Ivan M, Cimmino A, et al (2007). Regulatory mechanisms of microRNAs involvement in cancer: the strange case of Dr Jekyll and Mr Hyde.
- Gogia S, Puranik M (2013). Solution structures of purine base analogues 6-chloroguanine, 8-azaguanine and allopurinol. *J Biomol Struct Dyn*, 1-9.
- Harada Y, Ishii I, Hatake K, Kasahara T (2012). Pyrvinium pamoate inhibits proliferation of myeloma/erythroleukemia cells by suppressing mitochondrial respiratory complex I and STAT3. *Cancer Lett*, **319**, 83-8.
- Hulsegge I, Kommadath A, Smits MA (2009). Globaltest and GOEAST: two different approaches for Gene Ontology analysis. *BMC Proc*, **3**, S10.
- Jarius S, Martínez-García P, Hernandez AL, et al (2013). Two new cases of anti-Ca (anti-ARHGAP26/GRAF) autoantibody-associated cerebellar ataxia. *J Neuroinflammation*, **10**, 1-6.
- Kelder T, van Iersel MP, Hanspers K, et al (2012). WikiPathways: building research communities on biological pathways. *Nucleic Acids Res*, **40**, D1301-7.
- Lamb J, Crawford ED, Peck D, et al (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, **313**, 1929.
- Langsenlehner U, Renner W, Yazdani-Biuki B, et al (2006).

- Integrin alpha-2 and beta-3 gene polymorphisms and breast cancer risk. *Breast Cancer Res Treat*, **97**, 67-72.
- Liu C, Kelnar K, Liu B, et al (2011). The microRNA miR-34a inhibits prostate cancer stem cells and metastasis by directly repressing CD44. *Nat Med*, **17**, 211-5.
- Luo M, Guan JL (2010). Focal adhesion kinase: a prominent determinant in breast cancer initiation, progression and metastasis. *Cancer Lett*, **289**, 127-39.
- Pico AR, Kelder T, Van Iersel MP, et al (2008). WikiPathways: pathway editing for the people. *PLoS Biol*, **6**, e184.
- Pontes-Júnior J, Reis ST, Dall'Oglio M, et al (2009). Evaluation of the expression of integrins and cell adhesion molecules through tissue microarray in lymph node metastases of prostate cancer. *J Carcinog*, **8**, 3.
- Porkka KP, Visakorpi T (2004). Molecular mechanisms of prostate cancer. *Eur Urol*, **45**, 683-91.
- Shapiro DJ, Katzenellenbogen B, Hergenrother PJ, Bolton EC (2012). Small molecule inhibitors of steroid receptors for breast and prostate cancer. *J Biol Chem*, **286**, 4043-8.
- Sherman BT, Lempicki RA (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, **37**, 1-13.
- Singh PK, Doig CL, Dhiman VK, et al (2012). Epigenetic distortion to VDR transcriptional regulation in prostate cancer cells. *J Steroid Biochem Mol Biol*, **136**, 258-63.
- Smyth GK (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*, **3**, 3.
- Subramanian A, Tamayo P, Mootha VK, et al (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, **102**, 15545-50.
- Sugahara KN, Teesalu T, Karmali PP, et al (2010). Coadministration of a tumor-penetrating peptide enhances the efficacy of cancer drugs. *Science*, **328**, 1031-5.
- Taylor BS, Schultz N, Hieronymus H, et al (2010). Integrative genomic profiling of human prostate cancer. *Cancer Cell*, **18**, 11-22.
- Team RC (2011). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008. ISBN 3-900051-07-0. Available at: <http://www.R-project.org>.
- Zhang B, Kirov S, Snoddy J (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res*, **33**, W741-8.
- Zhang L, Yang BX, Zhang HT, et al (2011). Prostate cancer: an emerging threat to the health of aging men in Asia. *Asian J Androl*, **13**, 574-8.
- Zheng Q, Wang XJ (2008). GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic Acids Res*, **36**, W358-63.