International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 6 Number 9 (2017) pp. 185-193**
Journal homepage: http://www.ijcmas.com

Original Research Article

https://doi.org/10.20546/ijcmas.2017.608.026

Differential Gene Expression Analysis of Prostate Cancer for Biomarkers and Potential Drug Targets Identification

Mujeeb Rahiman Thayyil Kunhumuhammed[1], Ashvini Desai[2], Inamul Hasan Madar[3]* and Iftikhar Aslam Tayubi[4]

[1]Department of Computer Science, Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Jeddah Kingdom of Saudi Arabia
[2]Department of Bioinformatics, School of Biosciences and Technology, Vellore Institute of Technology, Vellore, 632014, Tamil Nadu, India
[3]Department of Biotechnology and Genetic Engineering and Department of Biochemistry, Bharathidasan University, Tiruchirappalli, 620024, Tamil Nadu, India
[4]Faculty of Computing and Information Technology, Rabigh King Abdul-Aziz University, Jeddah, 21911, Saudi Arabia
**Corresponding author*

ABSTRACT

Keywords

Prostate Cancer, Microarray data analysis, Bioconductor, gene ontology, DAVID, Principal Component Analysis.

Article Info

Accepted:
04 July 2017
Available Online:
10 September 2017

Prostate Cancer is one of the leading causes of malignancy among men and is a complex multifaceted and biologically heterogeneous disease. The *in silico* Microarray data analysis produces the gene expression profiling of Prostate Cancer and it has also been shown to provide the molecular phenotyping that determines the characterizing of various stages of the cancer. In the current study, we analyzed the normal Prostate epithelial cells (PrEC) and Prostate cell line (LnCPa) data. For preprocessing we performed data normalization using the Reliability, Maintainability, and Availability (RMA) algorithm in the R (v-3.2.3) and limma package of the Bioconductor was used to identify the differentially expressed genes (DEGs), using the value ≤0.05 and fold change ≥ 2 to ≤ -2. The gene ontology and gene set enrichment analysis was performed using the DAVID online tool. Current study revealed 140 DEGs obtained by the Gene ontology and pathway analyses in DAVID using the filtered genes that had principle pathways targeting Prostate Cancer.

Introduction

Prostate cancer is the most common cancer in men apart from skin cancer. The incidence of Prostate cancer has shown significant variation across the globe. Though the prevalence and characteristics of this disease have been extensively studied in many countries, data regarding the true incidence of Prostate cancer in India is limited. According to the Statistics adapted from the American Society's publication, Cancer Facts and Figures 2017, an estimated 161,360 men in the United States have been diagnosed with

Prostate cancer. Most Prostate cancers (92 %) are diagnosed when the disease is confined to the Prostate and nearby organs. Cancer Patient's lifespan from the on-set of metastasis is about 5 years. The 5-year survival rate for most men with local or regional Prostate cancer is almost 100%. 98 % are alive after 10 years, and 96% live for at least 15 years. For men diagnosed with Prostate cancer that has spread to other parts of the body, the 5-year survival rate is only 29%.

The factors that stimulate the risk of Prostate Cancer include gender, age, family history and additionally alcohol intake, dietary fat, medication exposure, sexual factors and infectious diseases like gonorrhea (Gann *et al.,* 2002). These factors have aided to increase the progression of Prostate Cancer along with individual factors for half a century. The dramatic increase in Prostate Cancer research and its prevention has shown positive approach in the recent years (Carlson *et al.,* 2015). A majority of patients suffering from Prostate cancer can be saved by early diagnosis and effective treatment methods with novel drug target.

There are standardized tests to diagnose Prostate cancer in practice, but they can't specifically guarantee the prognoses of the disease; however, physicians perform certain medical tests, examination and other detection tests for the diagnosis of Prostate cancer. Apart from the conventional diagnostic method, biomarkers may serve as the confirmatory method for Prostate cancer and further add to the development of the new molecular targets for novel targeted drug development. Despite progress in the treatment of advancement in metastatic Prostate cancer (Labrie *et al.,* 1985; Crawford *et al.,* 1989; Denis *et al.,* 1993; Janknegt *et al.,* 1993; Caubet *et al.,* 1997; Dijkman *et al.,* 1997), still there is a lack of well recognized targeted treatment procedures and the only possibility of a significant reduction in Prostate cancer mortality is effective treatment of localized disease. Most of the cancer research has been done on the analysis and comparative studies of screening and diagnostic method of Prostate cancer. Despite years of research, little is known as to the exact cause of Prostate cancer, making it an area of intense research in medicine today. Genetics has also been found to be an important factor in determining who is at risk for Prostate cancer. It is unclear as to whether this genetic disposition has to do with actual gene defects or with similar patterns in diet and lifestyle. The current study successfully contributes to this area of predicting early diagnosis at genetic level (Biomarker) and selecting the best method of choice for the targeted gene which is predominantly expressed in etiology of Prostate cancer. To make a clear notion towards this hypothesis, the current study will highlight the significant genes.

Biological data analysis with advance computing technologies such as in computational biology approach provides diverse platforms to analyze the biomedical data for predicting differential expression level of genes in various diseased condition. Microarray has shown to be promising method of choice, which allows the study of genes in an organism under different conditions within a single experimental setup. The analysis of such data is performed using Bioconductor in R environment, which as a matter of fact has become the standard in the field of biomarker discovery and target identification for the effective treatment (Desai *et al.,* 2017).

With the emergence of microarray technology, gene expression can be measured on a genome-wide scale in cancer research by supplying tools and techniques to identify significant differences in diseases (Nevins *et al.*, 2007). This technology utilizes differential gene expression patterns in diseased and normal cells of various subtypes of cancer to identify the genes that are over-expressed and under-expressed (Kihara *et al.*, 2006). However, this analysis produces a large amount of data, which is challenging to interpret. With the employment of modern computational and statistical analysis packages in Bioconductor and other bioinformatics tools, the data analysis has been greatly flexible in the recent years based on diverse experiments. The microarray technology has been applied to a range of applications, including discovering novel disease subtypes, developing new diagnostic tools, and identifying underlying mechanisms of disease or drug response (Slonim *et al.*, 2009).

There are mainly two objectives of the current study, the first is to identify biomarkers related to the disease and add DEGs related to the pre-existing list of biomarkers and secondly to make an insightful analysis of the statistical tests and ontologies of the obtained biomarkers. To achieve the above objectives gene-wide analysis of the microarray gene expression profiling was analyzed.

Material and Methods

Data and data source

All the gene expression datasets were downloaded from NCBI's GEO database (http://www.ncbi.nlm.nih.gov/sites/GDSbrowser). The dataset GSE19726 downloaded consisted of samples taken from cancer patients and normal cells. Out of total 4 samples, 2 samples were collected from Prostate cancer cell line (LnCaP) and 2 samples were from Normal Prostate epithelial cells (PrEC) respectively. The datasets were downloaded in.CEL format and were analyzed on R environment (3.2.3). Most of the functionality in R is in the well-established extension packages. Most of the MA analysis packages can be found on Bioconductor (https://www.bioconductor.or g/) it is the largest growing platform for the biological data analysis and comprehension of high-throughput genomic data. R statistical programming language supports most of all the Bioconductor packages and is open source for its development.

Data Quality Check for the samples in the dataset

To check the quality and detect the outlier within the samples in the dataset, diagnostic plots such as box plots were plotted. These plots give a quick view of the normalized log2 intensities. In this work, the data normalization was performed using Reliability, Maintainability, and Availability (RMA) (Nirusha *et al.*, 2015).

Typically preprocessing methods, such as RMA, consist of several steps: background correction, normalization of probes, and summarization where individual probes are combined into a probe set. RMA is useful for highly precise estimates of expression.

Identification of differentially expressed genes

The original data was classified as PrEC and LNCaP groups and were analyzed using R v 3.2.3 and Bioconductor packages. The multichip normalization method Robust Multi-array Average was used for background correction, normalization across the chips, and summarization of probe level data (Irizarry *et al.*, 2003).

To identify the differentially expressed genes in PrEC and LNCaP an adjusted P-Value ≤ 0.05 was used as cut-off criteria.

Gene ontology of DEGs

To investigate the DEGs at a functional level, Primarily, Database for Annotation, Visualization and Integrated Discovery (DAVID)-v 6.7 was used to interpret functionality of gene lists (Huang *et al.*, 2009), to analyze GO classification for identification of biological processes, cellular components, molecular function, visualizing the genes and mapping the genes intro respective pathways, Kyoto Encyclopedia of Gene and Genomes (KEGG) (Kanehisa *et al.*, 2014) was used. The DEG were subjected to gene enrichment analysis using an EASE enrichment analysis with an EASE Score Threshold (minimum count) of 2 for including the minimum number of genes for corresponding GO term.

Results and Discussions

Quality analysis of samples in the dataset

The quality control analysis involves the assessment of the data and detection of the outliers. In this work, the data normalization was performed using Reliability, Maintainability, and Availability (RMA). Typically preprocessing methods, such as RMA, consist of several steps: background correction, normalization of probes and summarization where individual probes are combined into a probe set. RMA is useful for highly precise estimates of expression. The boxplot of the raw data (Fig. 1) represents the distribution of log2 intensities across all the samples. The boxplot of normalized signal intensities across all samples provide a certainty that the normalization step was accomplished (Fig. 2).

Identification of DEGs

The limma package was used to build the model matrix with defined contrasts and an adjusted false discovery rate to analyze the gene. Expression analysis profiles of PrEC and LnCap dataset led to identification of 140 genes, namely SFN, PTGFR, CLCA2, GBP6, IFI16, LAMC2, CFH, F3, CNN3, S100A2, S100A16, S100A14, PTGS2, LAMB3, AKT3, GPR158, DKK1, PLAU, MYOF, COL17A1,CD44, GLYATL1, GSTP1, ARRB1, RAB38, IL18, ETS1, CCND2, EMP1, CNTN1, KRT7, GLIPR1, ARHGDIB, PTHLH, TFCP2, KRT6A, KRT5, LIN7A, PPFIA2, DUSP6, SCEL, SLAIN1, GJB2, MMP14, TTC6, FRMD6, PRKD1, SLC27A2, ANXA2, BNC1, TNS4, KRT23, KRT15, KRT19, KRT17, DSG3, SLC14A1, SERPINB13, SERPINB7, SERPINB2, DSC3, GNA15, KLK5, SPTBN1, GALNT5, INPP1, EFEMP1, IL1A, IL1B, RND3, ITGB6, RBMS1, GCG, FAP, TMEFF2, FLRT3, MIR99AHG, NCAM2, TMPRSS2, TGFBR2, EPHA3, UPK1B, STXBP5L, CSTA, NLGN1, B3GNT5, TP63, HGD, GPR87, CLDN1, EPGN, EREG, AREG, ANXA3, DAPP1, SPON2, FGFBP1, TMEM156, UGT2B17, CLGN, FST, IQGAP2, RANBP3L, GCNT4, ALDH7A1, FBN2, FAT2, SPARC, FAM83B, CD109, NT5E, GJA1, SERPINB1, ADGRF1, COL12A1, ECHDC1, ITGB8, CFAP69, ZNF655, SERPINE1, CAV2, CAV1, MET, PTPRZ1, MACC1, INHBA, SEMA3A, CDK6, TFPI2, AKR1B1, NRG1, RALYL, ANXA2P2, ANXA1, MIR31HG, PRUNE2, LPAR1, TIMP1, SLC6A14 and TARP that were found to be differentially expressed with an adjusted an $P \leq 0.05$ and $FC \geq 2$ and ≤ -2 based on t-test on the normalized resultant data. The volcano plot arranged genes along dimensions of biological and statistical significance (Fig. 3).

GO clustering and pathway enrichment of DEGs

The functional classification of the obtained 140 DEGs was performed with the online biological classification tool-DAVID. The gene list was submitted with Affymetrix Human U133 Chip as background which provides for enrichment calculation. An EASE Score Threshold (maximum probability), a modified Fisher exact p-value ≤ 0.01 was used for strong gene enrichment. The count threshold (minimum count) of 2 was used to retrieve gene counts belonging to GO term with its categories (classifications). The functional annotations of gene classification with their GO terms, p-value count in the present study were identified and detailed in table 1. The DAVID analysis revealed that 140 genes were significantly associated with GO terms and pathways. The KEGG pathway associations for the obtained genes were reported in table 2. Further investigations on these pathways pave a novel way of developing new therapies for treating patients with Prostate cancer.

Prostate cancer is one of the most alarming health problems in the world as the principle cause of death among men globally with varying incidence rates.

Fig.1 The boxplot showing the summarized log2 intensities on the y-axis and the distribution of 2 Normal Prostate epithelial cells and 2 Prostate cancer cell line samples for the raw data

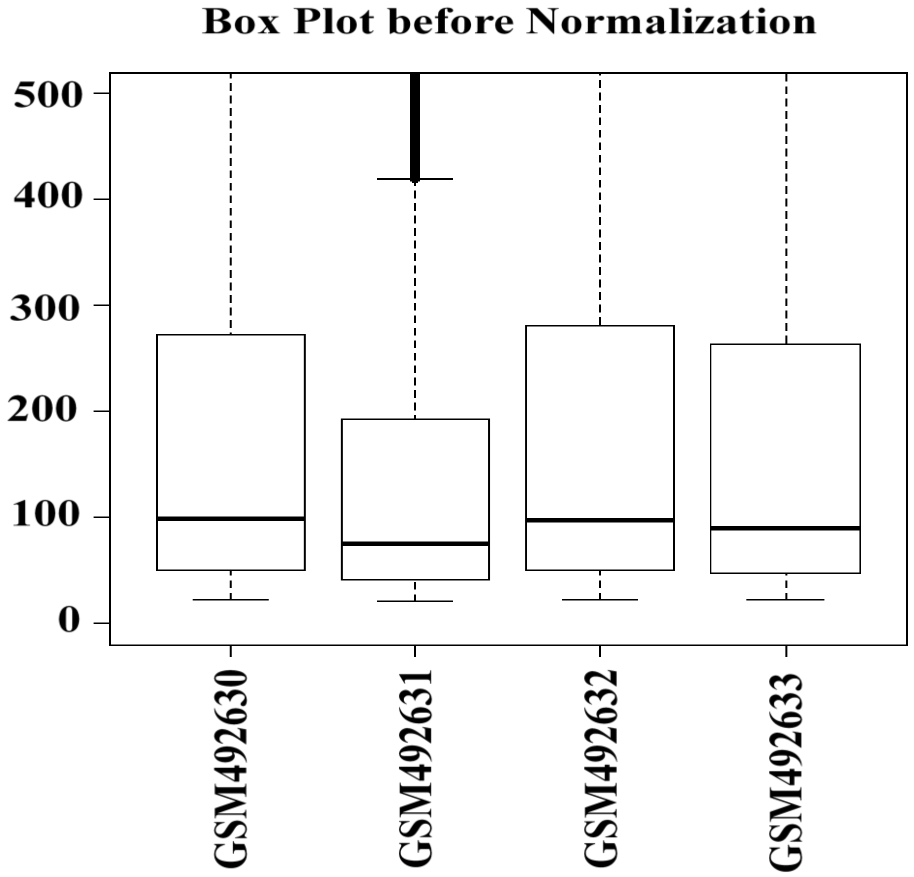


Fig.2 The boxplot showing the normalized log2 intensities on the y-axis and the distribution of 2 Normal Prostate epithelial cells and 2 Prostate cancer cell line samples

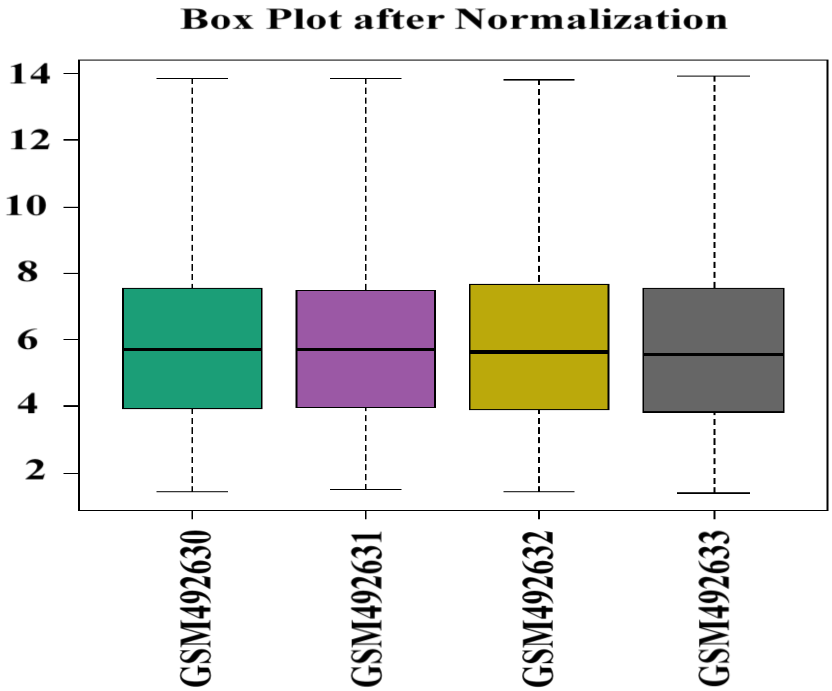


Fig.3 A comparison of two-gene selection method in a volcano plot. Each circle corresponds to one gene. The figure represents the average log-ratio (log fold-change) in the two group comparison. The 2-fold change method selects all genes above the line x=0.5 and below the line x=-0.5, as differentially expressed ones. Green color indicates highly expressed genes from the Prostate cancer

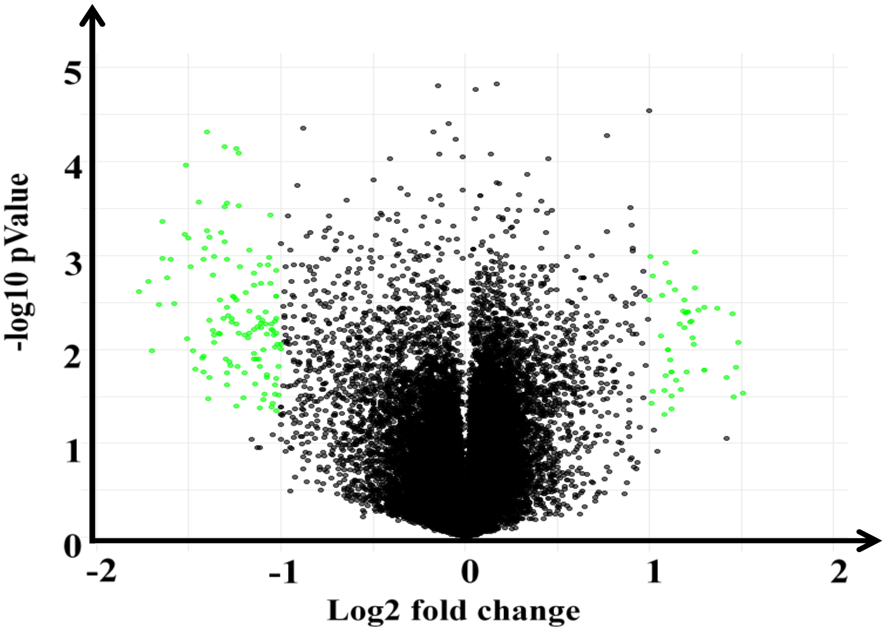
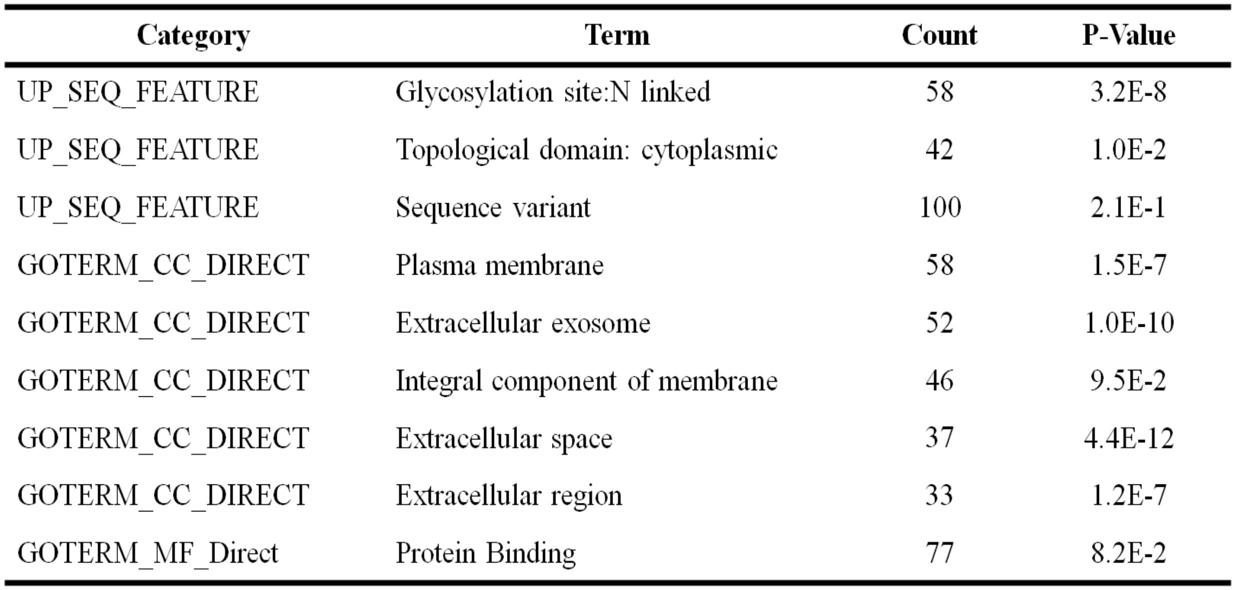


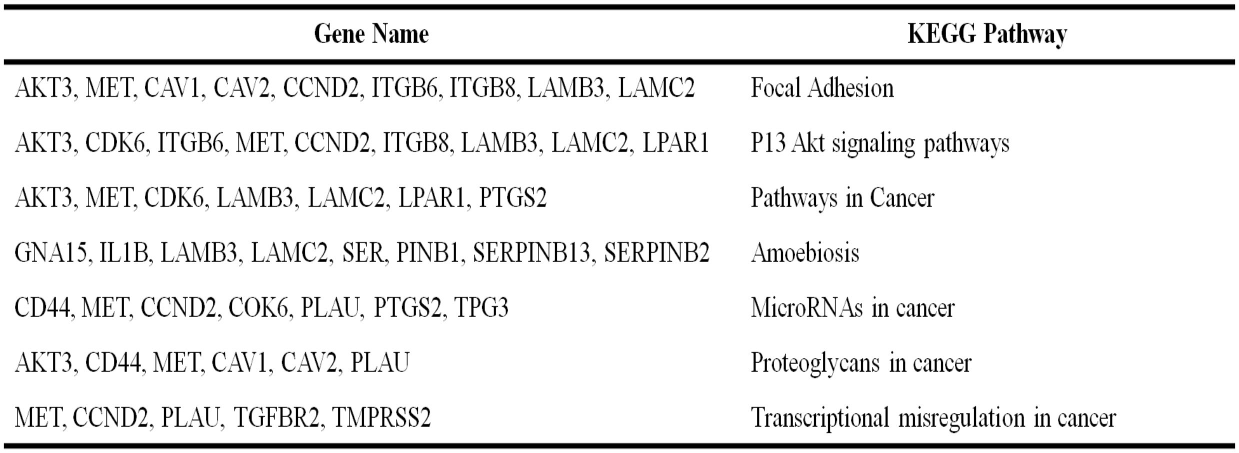
Table.1 GO categories with corresponding GO terms



Category	Term	Count	P-Value
UP_SEQ_FEATURE	Glycosylation site:N linked	58	3.2E-8
UP_SEQ_FEATURE	Topological domain: cytoplasmic	42	1.0E-2
UP_SEQ_FEATURE	Sequence variant	100	2.1E-1
GOTERM_CC_DIRECT	Plasma membrane	58	1.5E-7
GOTERM_CC_DIRECT	Extracellular exosome	52	1.0E-10
GOTERM_CC_DIRECT	Integral component of membrane	46	9.5E-2
GOTERM_CC_DIRECT	Extracellular space	37	4.4E-12
GOTERM_CC_DIRECT	Extracellular region	33	1.2E-7
GOTERM_MF_Direct	Protein Binding	77	8.2E-2

GO: Gene Ontology

Table.2 KEGG pathway associations of the genes



Gene Name	KEGG Pathway
AKT3, MET, CAV1, CAV2, CCND2, ITGB6, ITGB8, LAMB3, LAMC2	Focal Adhesion
AKT3, CDK6, ITGB6, MET, CCND2, ITGB8, LAMB3, LAMC2, LPAR1	P13 Akt signaling pathways
AKT3, MET, CDK6, LAMB3, LAMC2, LPAR1, PTGS2	Pathways in Cancer
GNA15, IL1B, LAMB3, LAMC2, SER, PINB1, SERPINB13, SERPINB2	Amoebiosis
CD44, MET, CCND2, COK6, PLAU, PTGS2, TPG3	MicroRNAs in cancer
AKT3, CD44, MET, CAV1, CAV2, PLAU	Proteoglycans in cancer
MET, CCND2, PLAU, TGFBR2, TMPRSS2	Transcriptional misregulation in cancer

The steady increase in the morbidity of Prostate cancer in the recent years indicates a need for additional research on this disease. Detection of Prostate cancer has increased substantially since the introduction of serum Prostate-specific antigen (PSA) screening. In spite of this, it is the second leading cause of cancer mortality for men in the U.S., mainly due to inadequate therapies. Microarrays utilize for large-scale experimental studies to generate expression of thousands of genes parallely within the set of samples make biological observations statistically significant. The current study focused on analyzing and understanding of 140 highly expressed genes study associated with Prostate cancer, providing us a humongous information on the genetic susceptibility of disease to take decisive steps for translating these findings into clinical care. Gene expression profiling of Prostate tumors enables us to have a better understanding of tumor type for the development of better drug and treatment procedures.

The gene expression affected by complex biochemical pathways and signaling events can be studied eventually. This study can be a fruitful foundation for the *in vitro* validation of the enriched genes so that effective biomarkers for treatment and prognosis of Prostate cancer can be identified for future research. Diagnosis and treatment options for organ-confined Prostate cancer are also complicated by the heterogeneity of the disease and the lack of prognostic tools for distinguishing between aggressive and non-aggressive cancers. The current research has focused on identifying markers that can be utilized for early on-set detection of cancer, differentiating clinically relevant disease from non-aggressive forms, and therapeutic targets for androgen-dependent or independent cancer. Molecules involved in several regulatory pathways are included among the potential prognostic and therapeutic markers. The differential gene expression of PrEC and LNCaP identified 140 genes with a significant P-value ≤ 0.05. These DEGs retrieved are important for investigating the mechanism of disease development from Prostate cancer. It is well known that Prostate Cancer treatment has severe side effects, which necessitate the need to find better chemotherapeutic agents (Chaussy *et al.,* 2001).

The results of GO functional annotation and pathway enrichment analysis of DEGs retrieved and their associated pathways were significantly enriched with GO terms, classifications and were associated with P13-Akt Signaling Pathways, Proteoglycans in cancer, microRNAs in cancer etc. The P13-Akt signaling pathway identified in our analysis is one of the principle targets for treating Prostate cancer (Shtivelman *et al.,* 2014). The aberrant activation of Akt is often associated with malignancy (Mahajan *et al.,* 2012) and up-regulation in terms of mRNA production in Prostate cancer. The pathways are involved in various cellular functions, cell proliferation, and growth. Hence, there is a need to carry out additional research for identifying potential targets as therapeutic agents that may directly or indirectly be involved in treating Prostate cancer.

Acknowledgement

We would like to thank Department of Biotechnology & Genetic Engineering and Department of Biochemistry, Bharathidasan University and Department of Computer Science, Faculty of Computing & Information Technology in Rabigh, King Abdulaziz University for all the support extended toward this research.

References

Carlsson, S., and Vickers, A. 2015. Spotlight on Prostate cancer: the latest evidence and current controversies. *BMC Medicine*, 13(1), 60.

Caubet, J. F., Tosteson, T. D., Dong, E. W., Naylon, E. M., Whiting, G. W., Ernstoff, M. S., and Ross, S. D. 1997. Maximum androgen blockade in advanced Prostate cancer: a meta-analysis of published randomized controlled trials using nonsteroidal antiandrogens. *Urology*, 49(1), 71-78.

Chaussy, C., and Thüroff, S. 2001. Results and side effects of high-intensity focused ultrasound in localized Prostate cancer. *Journal of Endourology*, 15(4), 437-440.

Crawford, E. D., Eisenberger, M. A., McLeod, D. G., Spaulding, J. T., Benson, R., Dorr, F. A., and Goodman, P. J. 1989. A controlled trial of leuprolide with and without flutamide in prostatic carcinoma. *New England Journal of Medicine*, 321(7), 419-424.

Denis, L. J., De Moura, J. C., Bono, A., Sylvester, R., Whelan, P., Newling, D., and Depauw, M. 1993. Goserelin acetate and flutamide versus bilateral orchiectomy: a phase III EORTC trial (30853). *Urology*, 42(2), 119-130.

Desai, A., Madar, I. H., Asangani, A. H., Al Ssadh, H., and Tayubi, I. A. 2017. Influence of PCOS in Obese vs. Non-Obese women from Mesenchymal Progenitors Stem Cells and Other Endometrial Cells: An in silico biomarker discovery. *Bioinformation*, 13(4), 111-115.

Dijkman, G. A., Janknegt, R. A., De Reijke, T. M., and Debruyne, F. M. 1997. Long-term efficacy and safety of nilutamide plus castration in advanced Prostate cancer, and the significance of early Prostate specific antigen normalization. *The Journal of Urology*, 158(1), 160-163.

Gann, P. H. 2002. Risk factors for Prostate cancer. *Reviews in urology*, 4(Suppl 5), S3-S10.

Gene Ontology Consortium. Gene ontology consortium: Going forward. *Nucleic Acids Res* 2015; 43:D1049-56.

Huang, D. W., Sherman, B. T., and Lempicki, R. A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44-57.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer- Barclay, Y. D., Antonellis, K. J., Scherf, U., and Speed, T. P. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2), 249-264.

Janknegt, R. A., Abbou, C. C., Bartoletti, R., Bernstein-Hahn, L., Bracken, B., Brisset, J. M., and Debruyne, F. M. 1993. Orchiectomy and nilutamide or placebo as treatment of metastatic prostatic cancer in a multinational double-blind randomized trial. The *Journal of urology*, 149(1), 77-82.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research*, 42(D1), D199-D205.

Kihara, D., Yang, Y. D., and Hawkins, T. 2006. Bioinformatics resources for cancer research with an emphasis on gene function and structure prediction tools. *Cancer Informatics*, 2, 25-35.

Labrie, F., Dupont, A., Belanger, A., Giguere, M., Lacoursiere, Y., Emond, J., and Bergeron, V. 1985. Combination therapy with flutamide and castration (LHRH agonist or orchiectomy) in advanced Prostate cancer: a marked improvement in response and survival. *Journal of Steroid Biochemistry*, 23(5), 833-841.

Mahajan, K., and Mahajan, N. P. 2012. PI3K- independent AKT activation in cancers: A treasure trove for novel therapeutics. *Journal of Cellular Physiology*, 227(9), 3178-3184.

Nevins, J. R., and Potti, A. 2007. Mining gene expression profiles: expression signatures as cancer phenotypes. *Nature Reviews Genetics*, 8(8), 601-609.

Nirusha, P. P., Thriveni, T. G., and Gurumurthy, H. P. Microarray data analysis using r programming. International *Journal of Scientific Engineering and Applied Science*, 1(3), 264-282

Shtivelman, E., Beer, T. M., and Evans, C. P. 2014. Molecular pathways and targets in Prostate cancer. *Oncotarget*, 5(17), 7217-7259.

Slonim, D. K., and Yanai, I. 2009. Getting started in gene expression microarray analysis. *PLoS Comput Biol*, 5(10), e1000543.

How to cite this article:

Mujeeb Rahiman Thayyil Kunhumuhammed, Ashvini Desai, Inamul Hasan Madar and Iftikhar Aslam Tayubi. 2017. Differential Gene Expression Analysis of Prostate Cancer for Biomarkers and Potential Drug Targets identification. *Int.J.Curr.Microbiol.App.Sci.* 6(9): 185-193.
doi: https://doi.org/10.20546/ijcmas.2017.608.026