IDENTIFICATION OF POTENT BIOMARKERS FOR PROSTATE CANCER THROUGH AR, MAPK AND m-TOR SIGNALING PATHWAYS MINING

Enrolment No. **- 131503, 131511**

Name of Student(s) **- Abhinav Mishra, Nimisha Asati**

Name of Supervisor **- Dr. Tiratha Raj Singh**

May – 2017

Submitted in partial fulfillment of the Degree of

Bachelor of Technology

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION AND TECHNOLOGY,

WAKNAGHAT

Certificate

This is to certify that the project report entitled "**IDENTIFICATION OF POTENT BIOMARKERS FOR PROSTATE CANCER THROUGH AR, MAPK AND m-TOR SIGNALING PATHWAYS MINING**", submitted by **Abhinav Mishra** and **Nimisha Asati** in partial completion for the award of degree of Bachelor of Technology in Bioinformatics to Jaypee University of Information Technology, Waknaghat has been carried out under my guidance.

This work has not been submitted partly or completely to any other University or Institute for the award of this or any other degree or diploma.

……..

Dr. Tiratha Raj Singh

Assistant Professor (Senior Grade)

Date:……..

Acknowledgement

Known words become inadequate to express our gratitude for our mentor and supervisor Dr. Tiratha Raj Singh, Assistant professor in Department of Biotechnology and Bioinformatics, who initiated us into the realm of research and supervised us with finite patience and without whose invaluable suggestion and unstinted co-operation, the present desertion would not have been possible. We are also thankful to Ms. Ankita Shukla (PhD Scholar) for her incessant upkeep and encouragement. It was her who delivered us the technical guidance and direction to this project and continuously impelled us to work harder on this endeavour. Also, We would like to thank the officials of Jaypee University of Information Technology, Waknaghat for their help and cooperation.

……..

Abhinav Mishra (131503)

Date: …….

……..

Nimisha Asati (131511)

Date: …….

Contents

Introduction

Prostate Cancer (PC) is the second most widespread disease among men in the world (Figure 1 & 2) and most often identified cancer in developing countries. The prostate is an exocrine gland, its secretions are discharged outside the body of the male reproductive system. It is directly beneath the bladder, a walnut sized gland in front of the rectum. The liquidity of semen in prostate gland is maintained by the epithelial cells which produce a protein named PSA (prostate-specific antigen). Since, certain PSA seepages into the bloodstream, PSA levels of a patient is measured by checking blood. Any prostate condition is indicated if PSA levels are high i.e. above 4ng/ml. If the levels are more than 10ng/ml, the likelihood of occurrence is above 50%.

In most cases, the cancer starts in the gland cells called *adenocarcinoma*.

The probability of development of PC is more in older men. In men aged 65 or older out of 10 cases in 6 are diagnosed, and it is infrequent before age 40. At the time of diagnosis, the average age is around 66.

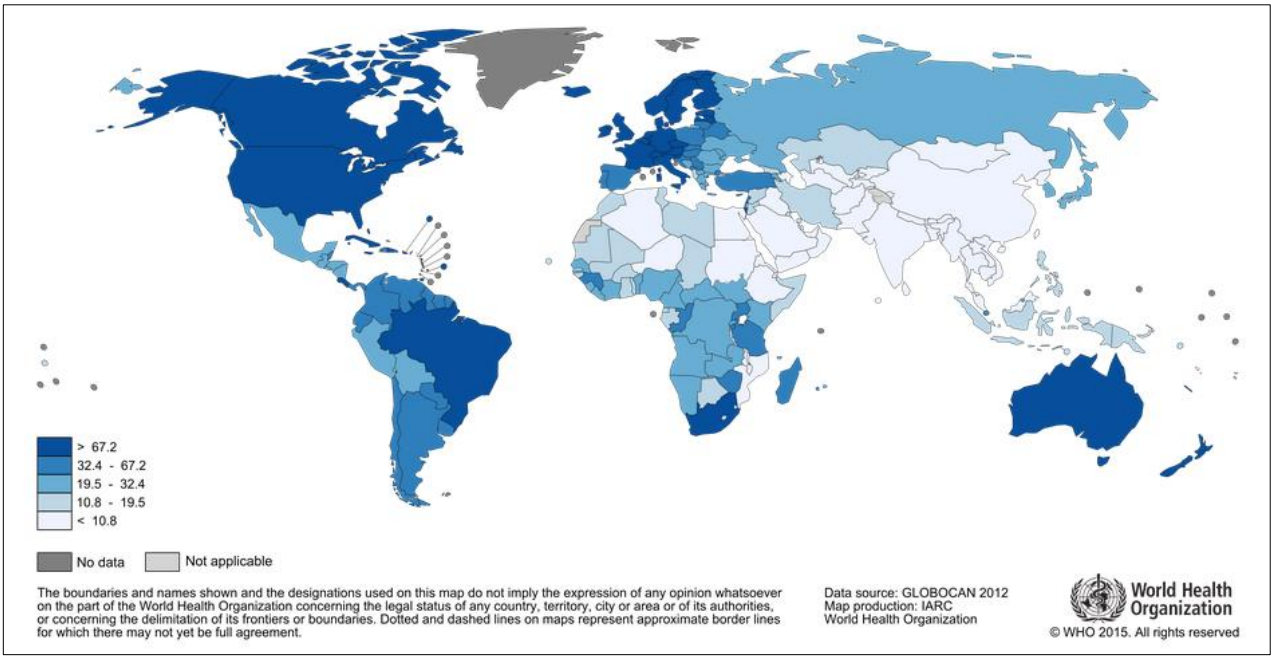


Figure 1. Estimated Prostate Cancer Incidence Worldwide in 2012. ©GLOBOCAN 2012

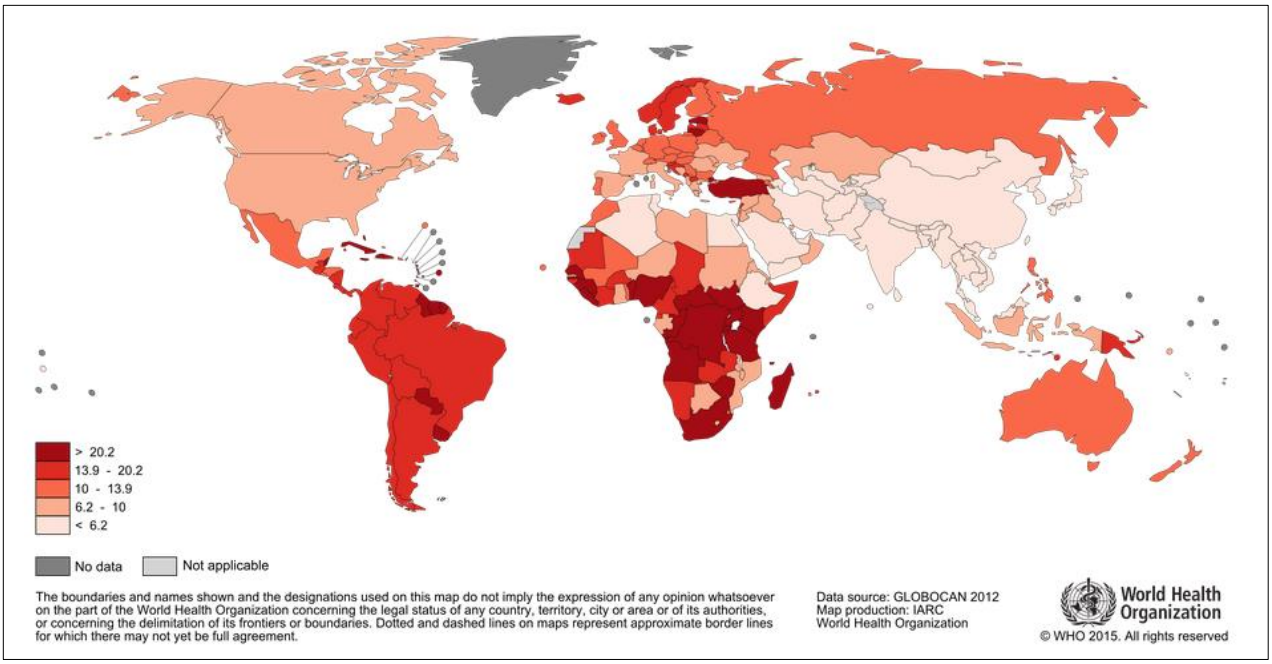


Figure 2. Estimated Prostate Cancer Mortality Worldwide in 2012. ©GLOBOCAN 2012

It starts out as a pre-cancerous condition in which several tests assist in the diagnosis of PC. Since, it is typically a very slow advancing disease that instigate with insignificant changes in the shape and size of the gland cells named as Prostatic intraepithelial neoplasia (PIN). There is huger risk of having cancerous cells in the prostate of any patient who has a high-grade PIN after a prostate biopsy.

The other test is called as Proliferative inflammatory atrophy (PIA) in which with the signs of inflammation, the prostate cells appear smaller than normal. It may lead to high-grade PIN.The most common system for determining the stage of cancer is the TNM (Tumour/Nodes/Metastases). It is essential to distinguish between cancers that are classified to the gland, and others which disperse to other parts of the body. Clinical T1 and T2 cancers originate only in the gland, while T3 and T4 have dispersed outside the gland.

The Gleason score identifies cancerous tissue whose grading range lies from 2 to 10. More abnormality in the tissues is identified in the patients with a high score. There are several causes including age, genetics, diet, medication, obesity, sexually transmitted diseases (STDs), obesity, and chemical exposures.

Initial PC typically triggers no symptoms but advanced ones causes problems urinating, blood in the urine or semen, erectile dysfunction, pain in the bones, weakness or numbness in the legs or feet.

In existence, two types of staging for the disease has been developed. Firstly, clinical stage, which is the best approximation of the coverage of the disease, centered on physical exam, prostate biopsy, lab tests and any imaging tests. After surgery, the pathologic stage occurs.

Possible complications that can occur during progression of disease are:

Metastasis accounts for the condition when, the cancer can disperse to other body sections, organs or bones, through the blood or the lymphatic coordination. There is a risk of severe kidney problems, if the cancer reaches to the ureters, which are the tubes that transfer urine from kidneys to bladder. They cause pain and fractures if it spreads to the bones.

Incontinence occurs when the prostate cancer or treatments can cause lack of voluntary control over urination which is known as incontinence.

Metabolic factors are those which increase the risk of dying due to various causes like elevated blood sugar levels, high blood pressure, high blood lipid levels and a high BMI.

Survival rates are as follows (as of 03/11/2016, retrieved from cancer.org American Cancer Society):

- 100% - 5-year relative survival rate
- 98% - 10-year relative survival rate
- 95% - 15-year relative survival rate

There can be visible risks during or after the prostate surgery which includes bleeding leading to clotting in the legs/lungs, responses to anesthesia, damage to other organs and surgical infections.

Hormone therapy is often very effective procedure to slow down the growth of already spread PC, but it is often less effective when incorporated for a longer time. Various terms are used to define cancers having no reprisal to hormones.

When the hormone therapy is maintaining the testosterone level in the body at low levels, but the cancer is growing, termed as **Castrate-resistant prostate cancer (CRPC).** The level ranges are called as castrate levels, those which would be expected if the testicles were removed. Nevertheless, the disease may react to various forms of hormonal therapy.

When the cancer that is no longer relieved or cured by any form of hormone therapy. It is called as **Hormone-refractory prostate cancer (HRPC)**.

Scientists can better understand how PC develops, through new evidences based on genetic variations related to the disease. The chromosomes that are inherited from both parents act as the factors based on which most of the genetic mutations are occurred and are studied as they may increase the risk of occurrence. Also, a certain variant of mitochondrial DNA, which is maternally inherited, might also raise the risk of occurrence of PC. Some hormonal medicines called *5-alpha reductase inhibitors* have also been tested by scientists as a way of reducing risk of the disease.

There is another latest technique, which enhances the colour Doppler. In this method, patient is first injected with a distinct agent that contains micro bubbles. It aids in improving the ultrasound images. Further studies are required before its use becomes common. Favourable results have been reported which are to be supported by more studies. More Doctors are studying if MRI procedure can be incorporated with TRUS (Trans rectal ultrasound). Patients who earlier had negative TRUS-guided biopsies will guide the biopsies when the doctor is suspecting cancer.

Role of AR, MAPK and m-TOR signalling pathways

Androgen Receptor Pathway

Prostate cancer cells, like normal prostate cells, necessitate androgens to grow and survive. The ratio of the cell proliferation rate to the cell death rate defines the growth of prostate cancer. To identify the fundamental molecular alterations in prostate-cancer cells, the pedantic determinants of the phenotype of PC cells includes carcinogen defences (GSTP1), signalling pathways containing growth factors (NKX3.1, PTEN, and p27), and AR. Detoxifying enzymes includes Glutathione S-transferases (GSTP1). Genomic damage is mediated by carcinogens in the cells of PIN conditions without GSTP1. Growth and survival of prostate cells in the normal prostate is regulated by NKX3.1, PTEN, and p27.When the levels of PTEN and NKX3.1 are low, there is a reduction in the level of p27& apoptosis and an increase in proliferation. To develop successful future therapies, it is necessary to understand the critical events and complexities of AR signalling in the progression to CRPC. Demonstrations through several studies suggest that tissues in CRPC have an increase in expression of androgens compared to both benevolent prostate tissue and PC that have not been treated before with hormonal ablation. Significant evidence supports the fact that the occurrence and development of castrate resistant prostate cancer (CRPC) is casually related to continued transactivation of AR. Beneath the discerning pressures of the AR-reduced microenvironment, the castration-resistant state develops during the progression of prostate cancer. The AR protein, being a member of a steroid receptor family which fits to a superior family of nuclear receptors, plays a significant role in regulation of prostate cancer growth as already suggested more than 50 years ago, by Huggins and Hodges [13] [14]. Androgen receptor is a transcription factor that is normally activated by its androgen ligand. The gene is composed of eight exons (Figure 3), of which A encodes the transactivation domain, B and C encodes one each of the zinc-finger elements which are responsible for the binding of DNA and exons D to H encodes the ligand-binding domain.

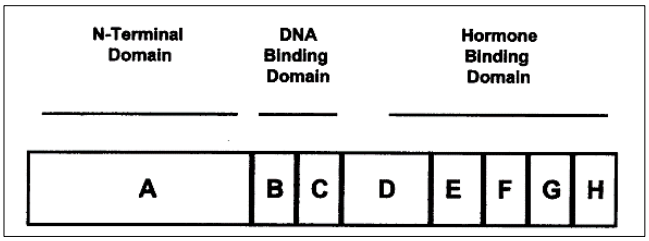


Figure 3. Schematic view of human AR.

During the androgen withdrawal therapy by AR gene amplification, mutations or by changed activity of its coactivators, the pathway also could be activated. These mechanisms lead to the inception of androgen-independent prostate cancer through the tumour cells. Recent studies show that AR activity is intimately linked to PC, depicting the hormone refractory and androgen independent systems. When the concentration of androgens that are circulating are low, numerous mechanisms such as gene amplification and upregulation of its protein intensities, mutations that let response of AR to other as well as its activation without ligand binding, consenting its central role and acting as a direct aim for therapy. Many attempts have been made to search for gland-specific molecules those might help as potent biomarkers or as therapeutic agents due to the limitations in the standard treatment procedures of prostate cancer. Although, AR is an essential player that controls different elements in all phases of prostate carcinogenesis, but many other signalling pathways along with their interactions with AR signalling, are also critically implicated especially in advanced stages of prostate cancer. In our research, (1) MAPK signalling, (2) m-TOR signalling pathways are targeted along with MAPK (Figure 4). Recently, Ras/MAPK signalling has been linked to androgen sensitivity in both AR-sensitive and AR-insensitive PC cell lines. Longstanding transfection of mutants in Ras that activate Ras/MAPK caused in over sensitivity of LNCaP cells to AR, which resulted in reduction of the level of androgen needed for cell growth, PSA expression and upkeep of cell tumorigenicity. To promote progression to metastatic CRPC, AR signalling may collaborate with the other oncogenic pathways associated with anoikis i.e. programmed cell death and cell survival. AR itself acts as a possible target of MAPK induction. The increase in activation of AR transcription and ability to recruit ARA70 is closely connected to phosphorylation of AR by ERK-2. So, it is feasible that through the MAPK pathway, DHT can provoke an autocrine stimulation of AR transcriptional activity which will cause AR phosphorylation and an enhanced recruitment of coactivators. The presence of rapid and non-genomic AR signalling is undisputable. The non-genomic regulation functions in AR through the activation of interwove complex signalling

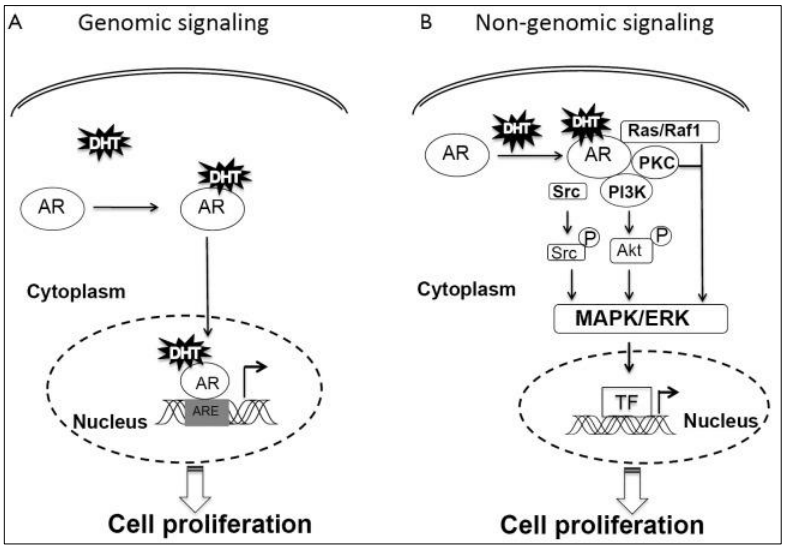


Figure 4. Genomic and non-genomic AR signaling in prostate cancer cells [16].

cascades causing the manifestation of proliferative genes and responses. It may also adapt genomic signalling of AR, allowing a synchronized, continuous and energetic response to AR stimuli leading to a potential mechanism of resistance to anti-androgens.

MAPK Pathway

The mitogen-activated protein kinase (MAPK) pathway is involved in various cellular functions like cell proliferation, differentiation and migration. It is also a very conserved module.

In mammals, at least four distinctly regulated groups of MAPKs are peculiar which are - extracellular signal related kinases (ERK)-1/2, Jun amino-terminal kinases (JNK1/2/3), p38 proteins (p38alpha/beta/gamma/delta) and ERK5, that are activated by specific MAPKKs. Every MAPKK can be triggered by one or more than one MAPKKK. This increases the complexity and diversity of MAPK signalling. Evidently, each MAPKKK confers a quick reaction to distinct stimuli.

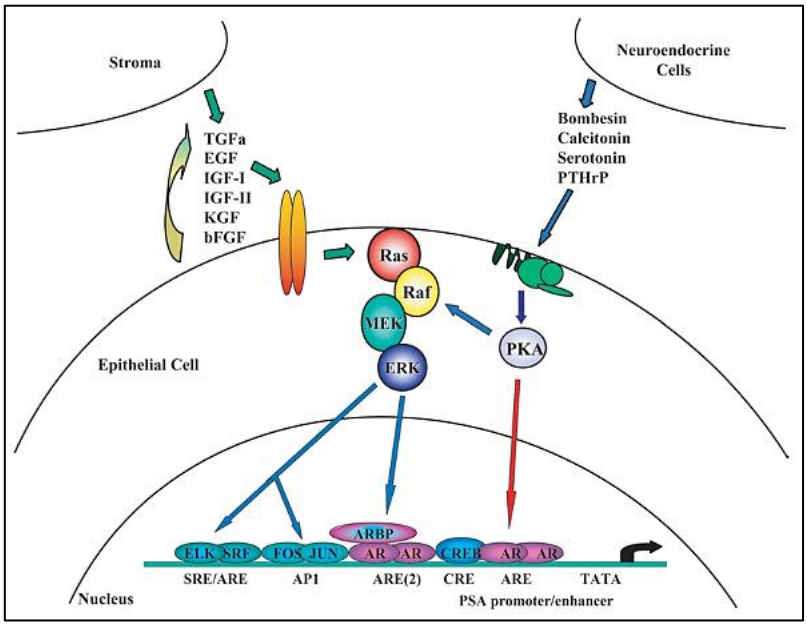


Figure 5. Gioeli D. [9]. **Growth factor signaling in prostate cancer progression**. As prostate cancer progresses to a hormone-independent and metastatic disease, there is an increase in the production of growth factors and their cognate receptors. Additionally, tumor cells become capable of autocrine production of growth factors such as TGFα. Signal transduction pathways involving growth factors have been exposed to rouse AR activation, signifying that the increase in growth factor as well as receptor expression could be fundamental in PC progression to androgen independence. ARBP, AR-binding protein; PTHrP, parathyroid hormone-related peptide; SRE, steroid response element; SRF, serum response factor.

In our current interpretation of monitoring of cellular events by growth factors and stresses, MAPKs gave great importance for eccentric fast developments. More than a dozen MAP kinase families are identified which are genetically conserved proteins, due to their initial discovery in yeast. Although, there are over one hundred publications in the literature depicting the effects of numerous manipulations of MAPK, which includes growth factors, chemical modifiers and androgens on prostatic cells, but MAPK pathways have not been thought in the PC.

Every MAPK cascade contains a core MAPK module, which has three or greater than three enzymes activated in sequence:

1. a MAPK
2. an immediate upstream kinase (MAPKK)
3. an additional kinase upstream of the MAPKK (MAPKKK)

The interaction of IL-6 with the MAPK pathways is supposed to be a major autocrine agent in the development of HRPC and therefore is a topic of interest [6]. ERK is the enzyme responsible for increased invasive and meta-static ability in PC. The investigations in PC3 cells prove that ERK is essential for clonogenicity, cell migration and invasion but plays only a slight role in growth and proliferation of the cells [24]. p38 might be a target for PC therapy given its proved influence to some disease characteristics, such as androgen dependence and metastatic phenotype acquisition [34]. The ability of ERK, JNK and p38 to behave either as prostate cancer suppressors or promoters varies on the kind of cells, evolving stage, and specific stimuli. Yet, the roles of these proteins at the molecular level are not recognized at all. There has been an increase in AR levels in 65% of CRPC cases which alone was not overly significant; however, combined with stabilization through c-Jun and phosphorylation from the MAPK cascade, the small increase becomes highly significant with increased activity [27]. This may also explain the correlation that exists between levels of MAPK in the nucleus and levels of AR found in tumours (Figure 5). In the reports of Raf-1 and MAPK protein expression using immunohistochemistry, the Her2/Raf-1/MAPK/AP-1 flow might indorse the growth of tumour, leading to initial deterioration as well as reduction in disease-specific survival. In addition, pathway associates may act as new therapeutic agents for PC.

There is adequate proof to propose that the pathway plays a significant role in the modulation of AR commotion in reaction to Ras. Studies probing MAPK activity in PC material suggested MAPK activity relates to development of a progressively complex and hormone independent PC [9]. Moreover, the Ras effector loop mutants that had the utmost natal result on LNCaP cells in vitro were the same that activated the pathway. Also, all AR independent xenografts showed elevated phosphor MAPK, regardless of whether their androgen independence was selected by sequential track, or generated by expressing Ras.

However, the mechanism(s) by which MAP kinase signalling modulates expression of these reporter constructs remains to be fully elucidated.

m-TOR Pathway

mTOR pathway is deregulated in most advanced PCs and acts as an important hub pathway for the growth signals to integrate with processes like protein synthesis, proliferation, survival, metabolism and differentiation which are the downstream cellular processes. So, it provides mechanisms for cancer cells to surmount the strain associated with androgen deprivation [6]. mTOR (Mechanistic Target of Rapamycin) is a highly-conserved serine/threonine protein kinase, which is found in two complexes which are mTOR complex 1 (mTORC1) and mTOR complex 2 (mTORC2). mTORC1 gets activated when growth factors, energy, amino acids, oxygen levels and stress, all these components are present. These components regulate many biological processes, which includes autophagy, lipid metabolism, ribosome biogenesis and protein synthesis. Whereas, mTORC2 consist of mTOR, mSin1, Rictor, Protor, Deptor, mLST8, Tel2 and Tti1. It reacts to growth factors and controls cytoskeletal organization, metabolism and survival. A direct connection is suggested between the PI3K-AKT-mTOR and androgen receptor (AR) signalling cascade in preclinical studies. This shows a robust interaction between these pathways during the development of androgen insensitivity and ADT resistance. So, there is a clear basis for the ceaseless clinical development of several novel inhibitors of the PI3K pathway, which can hinder the growth and survival of CRPC. PI3K-AKT-mTORsignalling pathway is undoubtedly reveals a very crucial node that directs ADT resistance and stimulates tumour growth in the adjusting the castrate levels of testosterone. In addition, this pathway is remodelled at genomic as well as the transcriptional level in almost all advanced PCs [37].

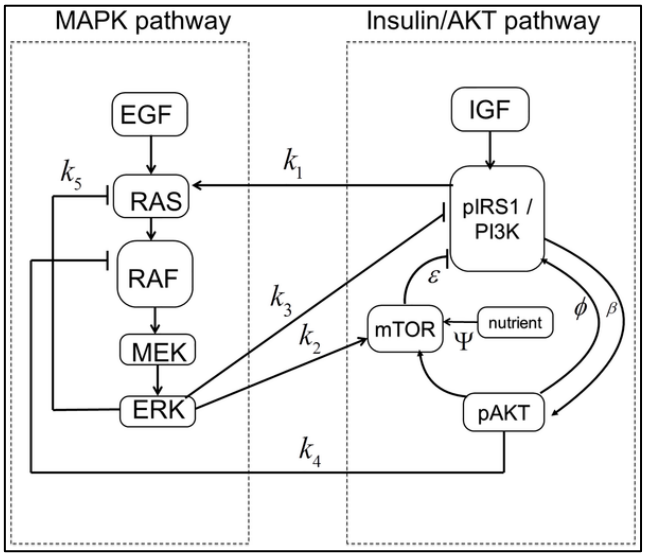


Figure 6. Arkun Y. [2] Dynamic Modeling and Analysis of the Cross-Talk between Insulin/AKT and MAPK/ERK Signaling Pathways.

It is fascinating to conjecture the reason for deregulation of PI3K-AKT-mTOR signalling pathway in prostate Cancer. There is a considerable strain that a prostate epithelial cell bears throughout the process of invasion, transformation, hormone deprivation and tumour growth. A possibility is that tumour cells necessitate hyper activation of the pathway and its downstream networks to surmount the important cellular stresses that saddle a cell during a period in cancer. So, it is unclear as PI3K-AKT-mTOR signalling controls which normal cellular processes that can be appropriated to steer

pathogenesis in cancer? The above described cross-section of pathway components, along with the cellular processes they influence, exemplifies the fundamental purpose of PI3K-AKT-mTOR signalling in cellular homeostasis and in establishment of a possible superiority for its deregulation in PC. During the diseased condition, the signalling pathway happens to be non-functional in 42% of localized disease and all of the advanced-stage disease, pinpointing the changes in the pathway which may act as precondition for progression of CRPC [37]. The hyper activation of the pathway is adequate to induce PC development and mTORC1 and mTORC2 together are essential to ease this progression in vivo [6]. The deregulation of the signalling pathway redoes the AR signalling constraints and hence, there is a reduction in the intrinsic needs for the androgens that cause castration resistance to stimulate PC growth. [6] [19] [21]. A decrease in the levels of AR activity reciprocally enriches AKT signalling through down regulation of PHLPP. On the contrary, in the environment of PTEN loss, mTOR inhibition fuels the AR levels through upregulation of HER3. This intensifies the stability of AR. [28]. Arkun Y. et. Al. advanced a mathematical model which examines the steady-state and dynamic characteristics of major feedback loops. These loops synchronize the cross-talk between insulin-AKT and MAPK/ERK signalling pathways (Figure 6) [2].

In addition to the need of biomarkers, these discoveries exhibit multiple opportunities which in reaction to inhibition of pathway, enables the interaction between mTOR signalling pathway and androgen signalling axis. AR positively regulated mTOR activity at several testosterone levels. The two negative regulators of mTOR, TSC1 and TSC2, might be encompassed due to the upregulation by AR knockdown. Though, this outcome only followed when testosterone levels were low. The experiment is meant only to provide a baseline for other studies. The inhibition of function of AR during the process of glucose deprivation, suppresses the activity of mTOR but results in apoptosis of more cells. Therefore, the complementary increase in the function of AR due to a suppressed mTOR signal is beneficial for being. When this loop is distributed at the time of Androgen deprivation therapy initiation, it may postpone, or even stop, the reappearance of PC [40].

Methods

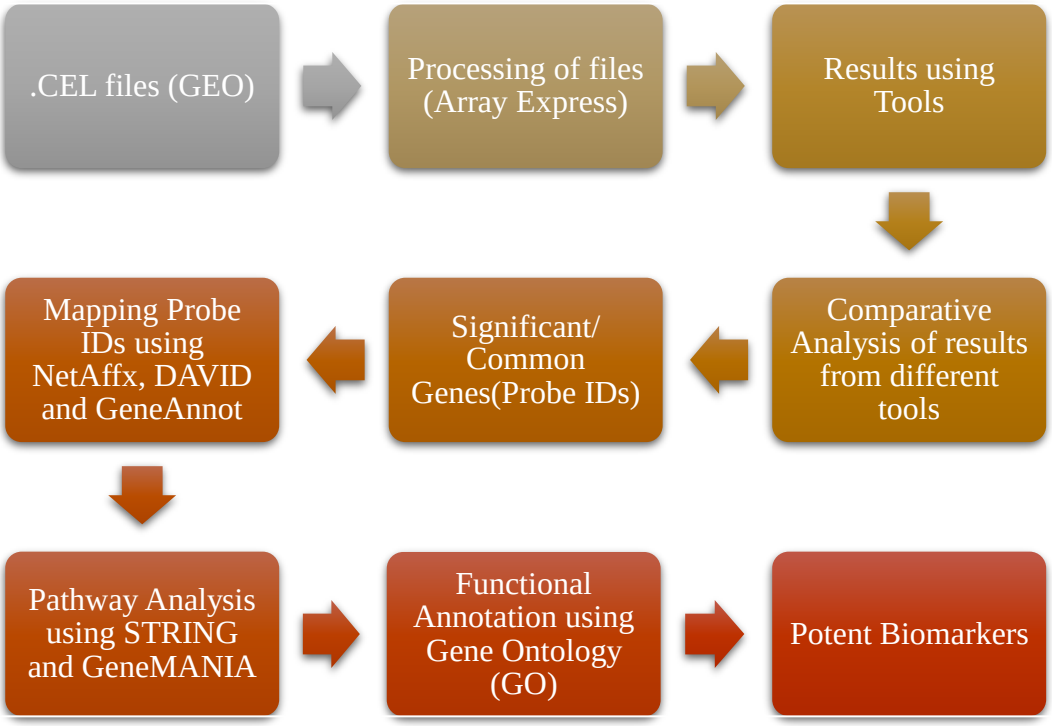


Figure 7. The proposed methodology for each series GSExxxx|x from GEO Datasets (GDS) for AR, MAPK and mTOR pathways.

Numerous biomarkers for occurrence and treatment of prostate cancer in different forms like genes, proteins, cell lines, miRNA have been identified with respect to the three pathways i.e., AR, mTOR and MAPK. Numerous of them have been examined in context of PC proneness. We classified studies pertinent to PC from the open literature repository PubMed. The annexation standards for this work were epidemiological studies of the connotation between polymorphisms in DNA repair system and PC risk. Research articles taken from the search phrase *"Prostate Cancer"* AND *"AR Pathway"*, likely for other targeted pathways, were collected for biomarker identification after review of the articles. Then, the biomarkers were grouped based on their type i.e. a gene, protein, miRNA, cell line etc. The corresponding effect shown by the biomarker such as over-expression, mutation, and inhibition was defined and the description of the process involving other proteins or at specific site of occurrence is described. This study gave us the information about the factors involved in the resistance and amplification of prostate cancer. Following is the table which consist the information regarding the data sets that we incorporated in our microarray data analysis for all the three pathways.

Pathway	Series	Significance	Samples (Test\Control)
AR	GSE2443	Yes	10 \ 10
	GSE8702	Yes	5 \ 5
	GSE21887	Yes	4 \ 4
	GSE33316	Yes	5 \ 5
	GSE67537	No[1]	9 \ 12
MAPK	GSE23038	Yes	6 \ 6
	GSE39735	Yes	3 \ 3
	GSE20906	Yes	2 \ 2
	GSE29438	Yes	8 \ 4
m-TOR	GSE26332	Yes	1 \ 1
	GSE32875	No[1]	4 \ 4
	GSE49232	Yes	2 \ 2

Table 1. Series Datasets for all three pathways.

[1] No correlation at any p – value/log FC

We collected the gene expression data using the GEO series ID's obtained from PubMed using the similar search phrase as *"Prostate Cancer" AND "AR Pathway" AND "Homo Sapiens"* which resulted in 27 entries. Then, we selected 5, 4, 3 series datasets for AR, MAPK, m-TOR pathways respectively (Figure 8). The processed gene expression data corresponding to these ID's was downloaded from ArrayExpress which has the probe ID in multiple rows and the samples in columns to identify the differential expressed genes within DNA repair pathways. Also, the same data was downloaded from GEO (Gene Expression Omnibus) database in CEL format for further analysis but there were no CEL files provided for series GSE20906 and GSE29438 (Table 1). The ArrayExpress data was then uploaded in the online analysis tool WebMeV (Multiple Experiment Viewer) which is a free and open-source cloud service platform that supports analysis, visualization, and stratification of large genomic data, particularly for RNASeq and microarray data. It takes the sample data as input and performs various analysis on the gene expression data. The tab delimited file is uploaded with only first row and column as header. It consists of options for Statistics, Clustering, Meta-analysis and Normalization. We performed GeneSD (standard deviation) in the range 0-0.998 showing different colours in the heat map for top 20 probes/genes. The genes are classified based on the standard deviation of their expression values for all samples. Then, GeneMAD (median of the absolute deviation) which is better at removing random clusters of multiple outliers of the expression values, was performed in the range 0-0.994, depicted by different colours in the heat map. Both the results were stored as tab delimited file with the probe ID and its SD and MAD values and the heat map in PNG format. Principal component analysis (PCA) which is used for clustering large number of genes in complex biological networks was carried out on the same data in which each dot represents a prostate cancer sample plotted against its expression levels for the probes/genes. The plots were saved as PNG format. Different clustering method like k-means clustering was also performed using Euclidean distance.

GEO2R was used for the expression analysis to compare two or more groups of samples, to distinguish genes that are expressed differentially throughout experimental states. Series ID is given as the input, displaying the list of the samples and their characteristics like source, cell type, genotype or variation. Based on literature of that series ID, the samples are grouped into test and control samples for diseased and non-diseased samples respectively (Table 1). GEO2Runcover skewed intensity distributions and by design usage of log transformation giving 'Top 250' probes imparted as the table of genes well-ordered by significance. The Probe ID, its *t-value* which is a measure of the difference in means taking into consideration the data spread and sample sizes, its p-*value* which when smaller than a predefined significance level (0.01), defines the gene as

differentially expressed. The *B value* i.e. the log-odd function for each gene which is differentially expressed between the two groups. It is often used for ranking genes. Then the fold change on a logarithmic scale (logFC) value is shown and the gene symbol and its title is displayed. The *F-statistic* calculates overall test of significance for that gene by combining the t-statistics for all pair-wise comparisons. So, the result is exported in tsv (tab separated value) file and analysed in Microsoft Excel. The top 20 and bottom 20 genes i.e. the over-expressed and under-expressed genes respectively are identified by sorting the *p-value* column in descending order and discarding the value except <0.01 and <0.05 (as defined earlier as the significance level of 99% and 95% respectively). Then, the corresponding annotations are performed using these probe ID's, first in NetAffx™, which enumerates the probe sequences and the consensus sequence interrupted by the probesIDs. For the chip sets, which are larger, sequence data in genomic context is displayed by the interactive maps. Then in, DAVID i.e. The Gene ID Conversion Tool is a server-based application which converts user's input gene or identifiers from one kind to another required kind in complete form with a mapping database of interconnected gene IDs. Also, GeneAnnot, which links between Affymetrix arrays and the rich human gene annotations available in GeneCards.

The WB-DEGS (Within and Between Group Comparisons for Differentially Expressed Gene Selection) tool, performs pre-processing, visualization, and genes selection with an accuracy to minimize the false positive rates using some classical methods of gene selection [43] (unpublished work from Dr. Tiratha Raj's Lab) (Figure 8). We uploaded the selected CEL files for series GSE2443 "**Prostate cancer - comparison of androgen-dependent and -independent micro dissected primary tumour**", and performed the pre-processing of data. Then we performed statistical analysis and mapped the overexpressed and under expressed genes.

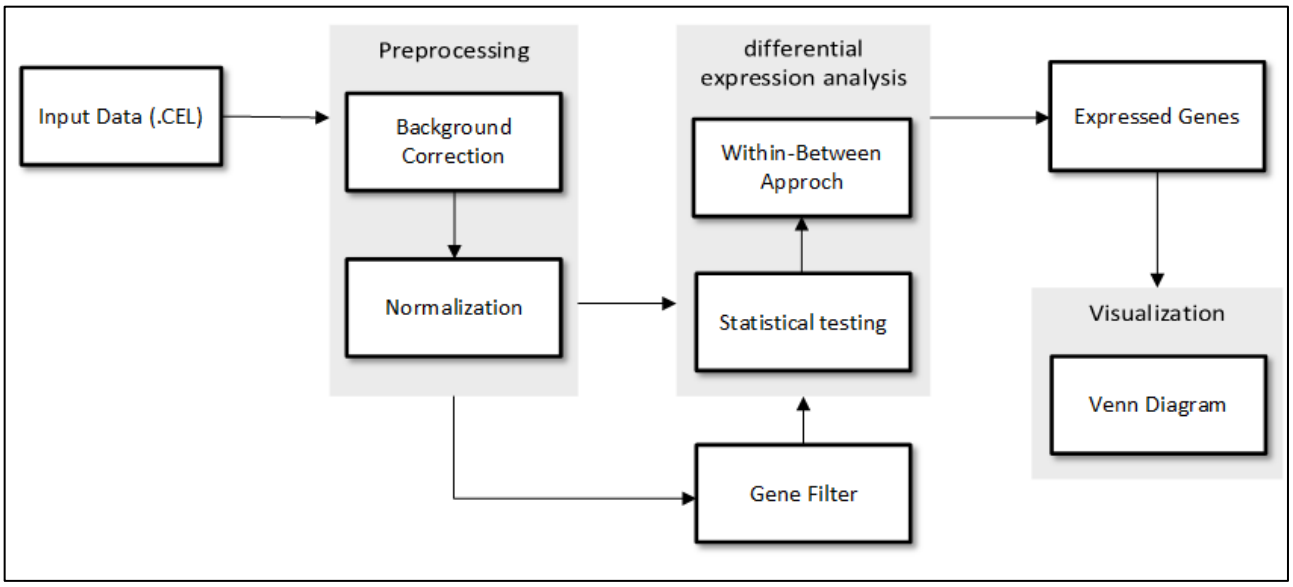


Figure 8. Workflow of WB-DEGS

After installation of latest version of R [31], we installed the necessary packages from CRAN and Bioconductor [1] [7] [8] [10] [12] [32] [33] [35]:

from CRAN:

```
install.packages("shiny")
install.packages("VennDiagram")
```

from Bioconductor:

```
source("http://bioconductor.org/biocLite.R")
biocLite("affy")
biocLite("affyPLM")
biocLite("limma")
biocLite("siggenes")
biocLite("twilight")
biocLite("genefilter")
```

Shiny app always has the similar arrangement [39]. At a minimum, the following two files:

1. ui.R
2. server.R

We can run WB-DEGS opening ui.R or server.R or by following code (Figure 9 & 10) :

```
library(shiny)
runApp("WB_Degs")
```

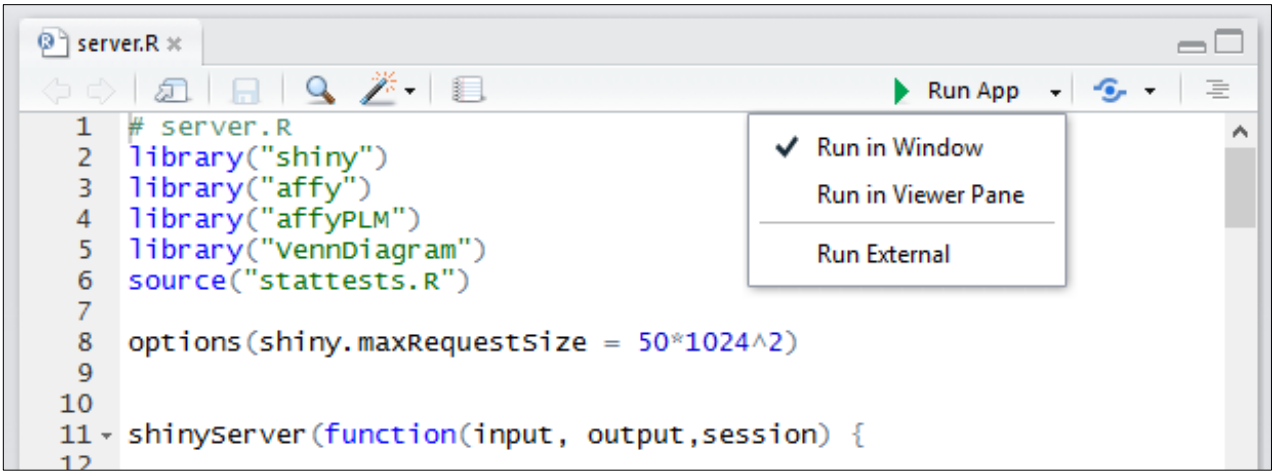


Figure 9. Running WB-DEGS in RStudio 0.99.903 (R v3.3.1)

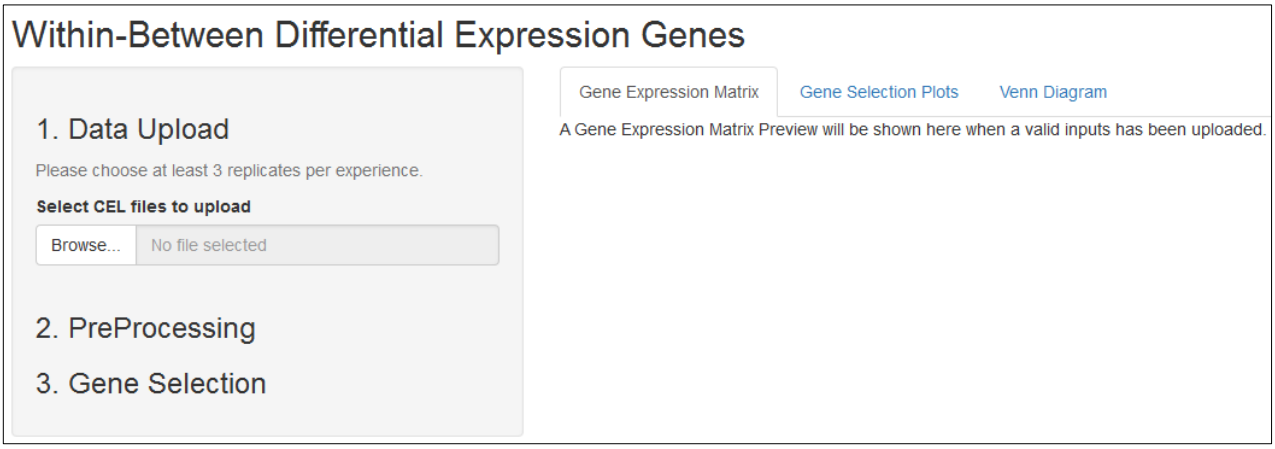


Figure 10. WB-DEGS GUI window (Running in Window menu).

We uploaded the selected .CEL files of GSE2443 in Data Upload Option, then, we used RMA as background correction method with Quantiles as normalization method. After the pre-processing of data, we divided the samples into two groups: (1) Group 1: test group, (2) Group 2: control group based on the curated sample data from the cited paper from GEO [5]. In the final step, we applied statistical analysis for estimation of local and global false discovery rate (FDR) and mapped the overexpressed and under expressed genes (Figure 11).

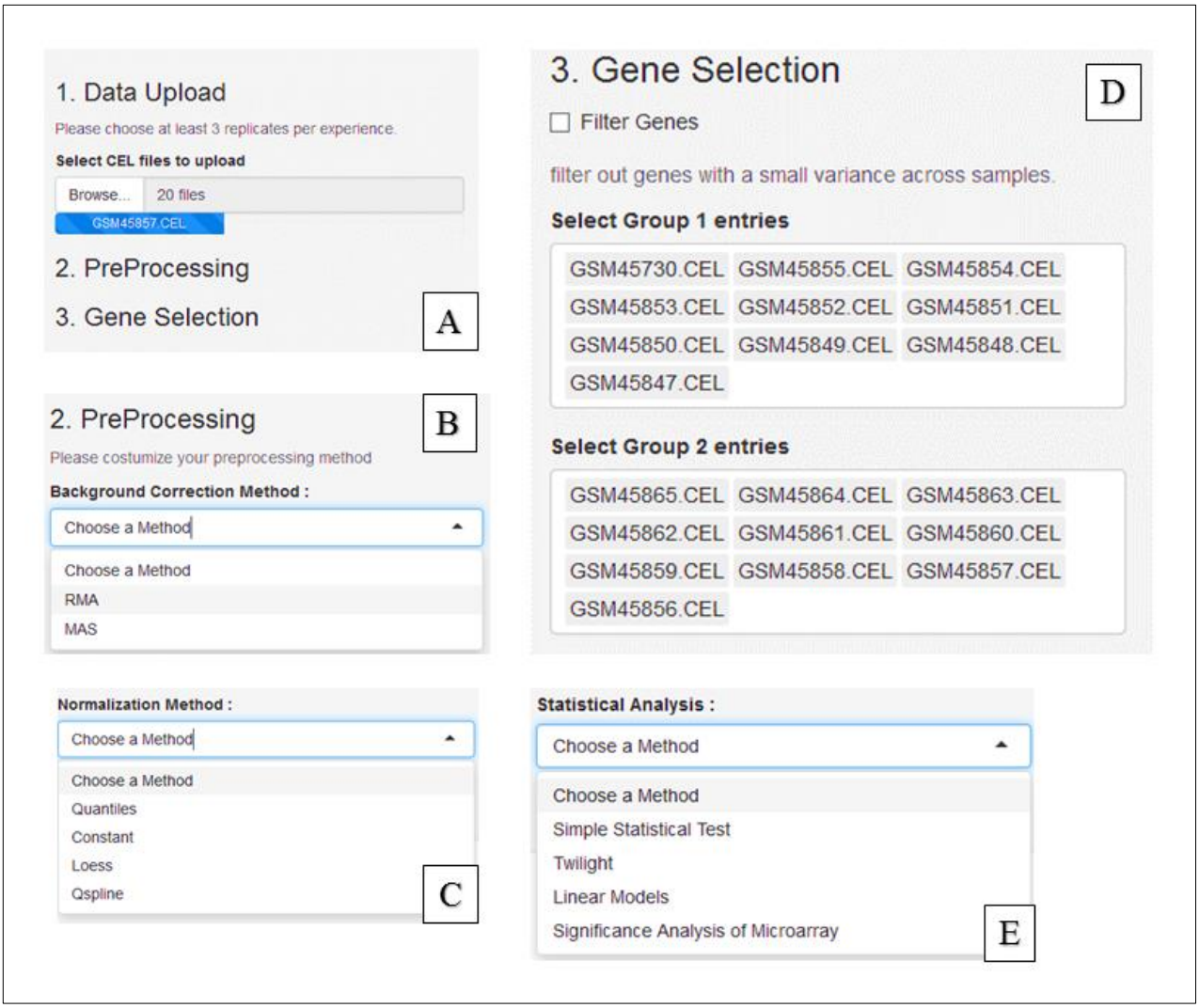


Figure 11. Detailed Workflow of WB-DEGS for series GSE2443.

A. Uploading .CEL files (20 samples).

B. Choosing background correction method (RMA/MAS).

C. Choosing normalizing method (Quantiles/Constant/Loess/Qspline).

D. Gene Selection: Group 1: Control group | Group 2: Test group.

E. Statistical Analysis (t-test, twilight, linear, SAM).

After mapping all the probe IDs obtained from different tools and analysis, we performed the comparative analysis of all the results from GEO2R, MeV and WB-DEGS for finding the common genes (probe IDs) (Table 6).

Results

The GEO2R [44] analysis for GSE2443 gave a great number of gene but after applying p-value cut-off and selecting genes from top and bottom of the list, we were left with only a few significant entities. The overlapping genes found may have some significance essential to PC after analysing them through three different methods. The genes identified as CEP57 (HGNC: 30794) and PDLIM5 (HGNC:17468) (Figure 12) in GSE2443 "**Prostate cancer - comparison of androgen-dependent and -independent micro dissected primary tumour**".

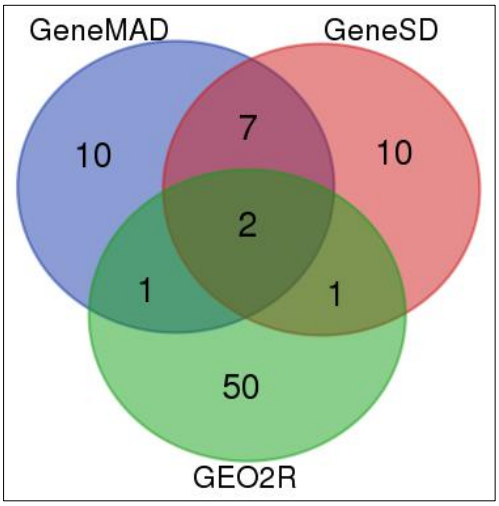
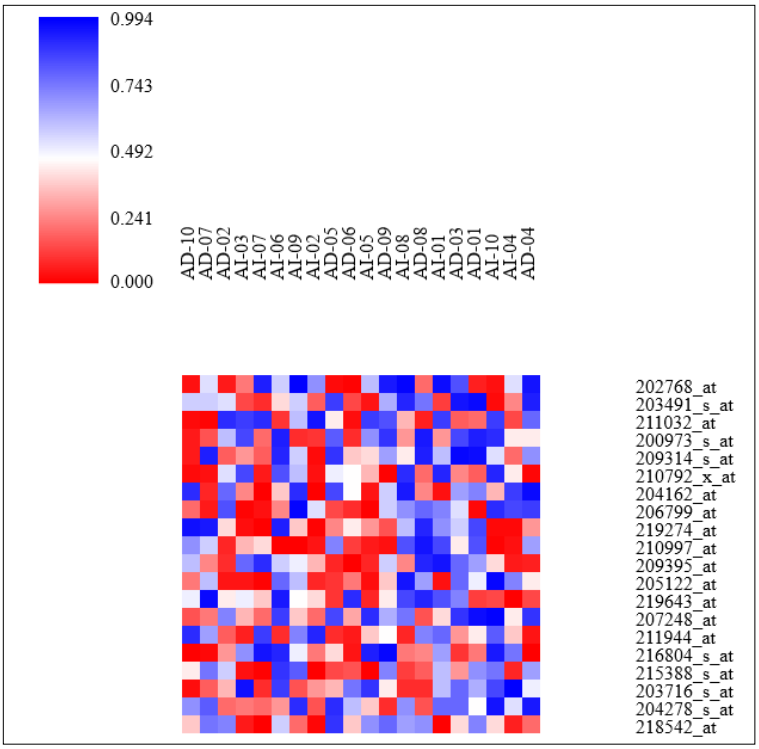


Figure 12. Common Genes in series GSE2443.

GeneMAD and GeneSD revealed the expression of probe IDs from 20 samples of series GSE2443 (Figure 13 & 14). The range is from 0 to 0.994 for the expression values for each sample in the data set.

The genes which were found common between the microarray data analysis using GEO2R, WB-DEGS and TM4 (WebMeV) were tabulated and grouped differently for all the three pathways. Then, input was provided in STRING database for each pathway and the network was constructed. The interaction value between the two genes was considered greater than or equal to 0.9 which is considered significant and others are discarded. Then, using Conditional formatting in excel, the genes were categorized into AR, MAPK and mTOR based on three different colours to visually identify their interactions (Table 2).



Red: Under expressed

Blue: Overexpressed

Figure 13. GeneMAD analysis for GSE2443 (20 samples).

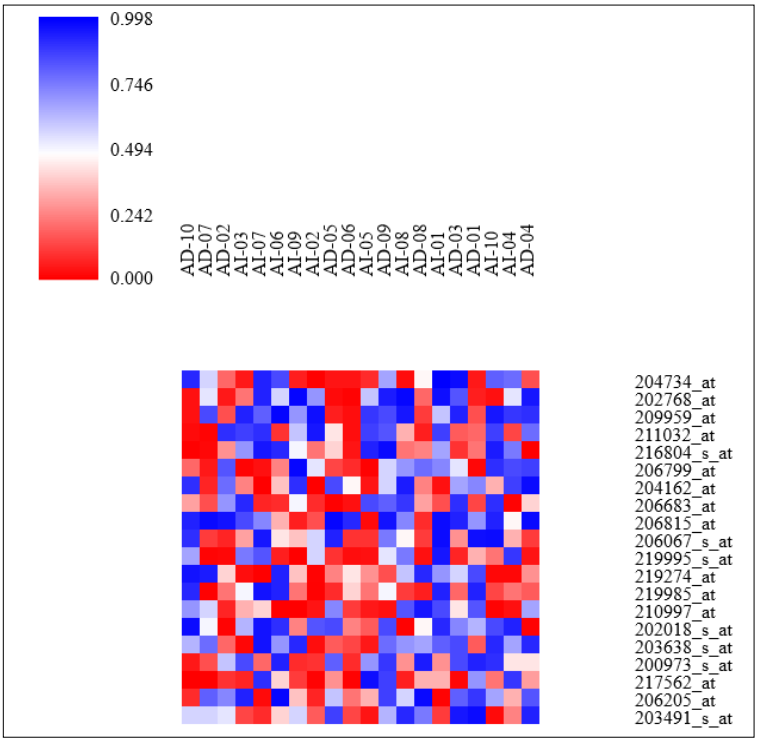


Figure 14. GeneSD analysis for GSE2443 (20 samples)

The mapping of probeIDs to their respective genes was done using the supplementary file, containing the unabridged list of genes provided in the cited article [6]. The SAM analysis results at the delta value 1.5 gave 15 probes out of which 14 were mapped to the genes in NetAffx™ (Figure 15).

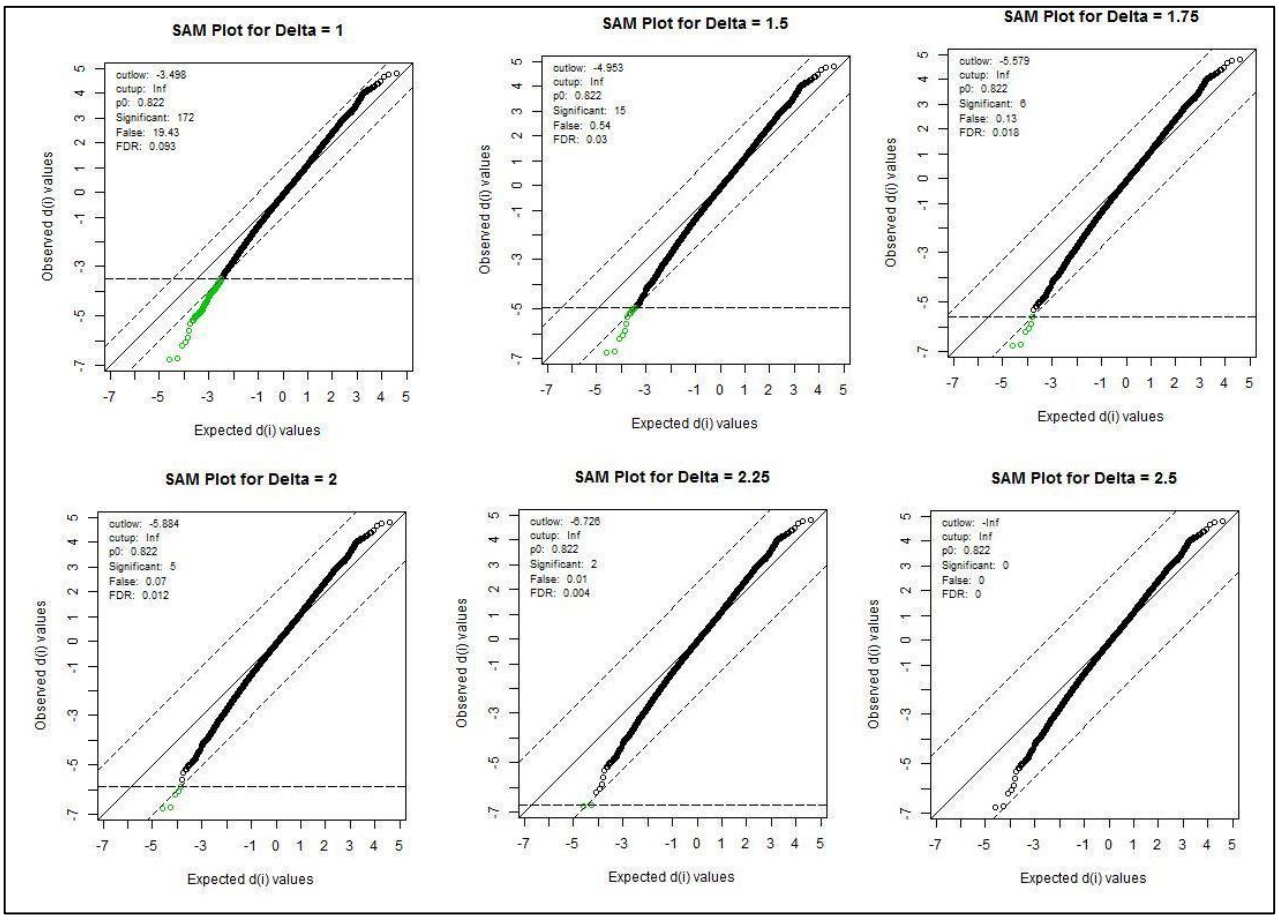


Figure 15. Between Group Comparisons in SAM analysis (GSE2443).

The Simple Statistical Test (*paired t-test*) gave 11 probes out of which 10 were successfully mapped to their respective genes (Figure 16). There was no significance in within group comparisons.

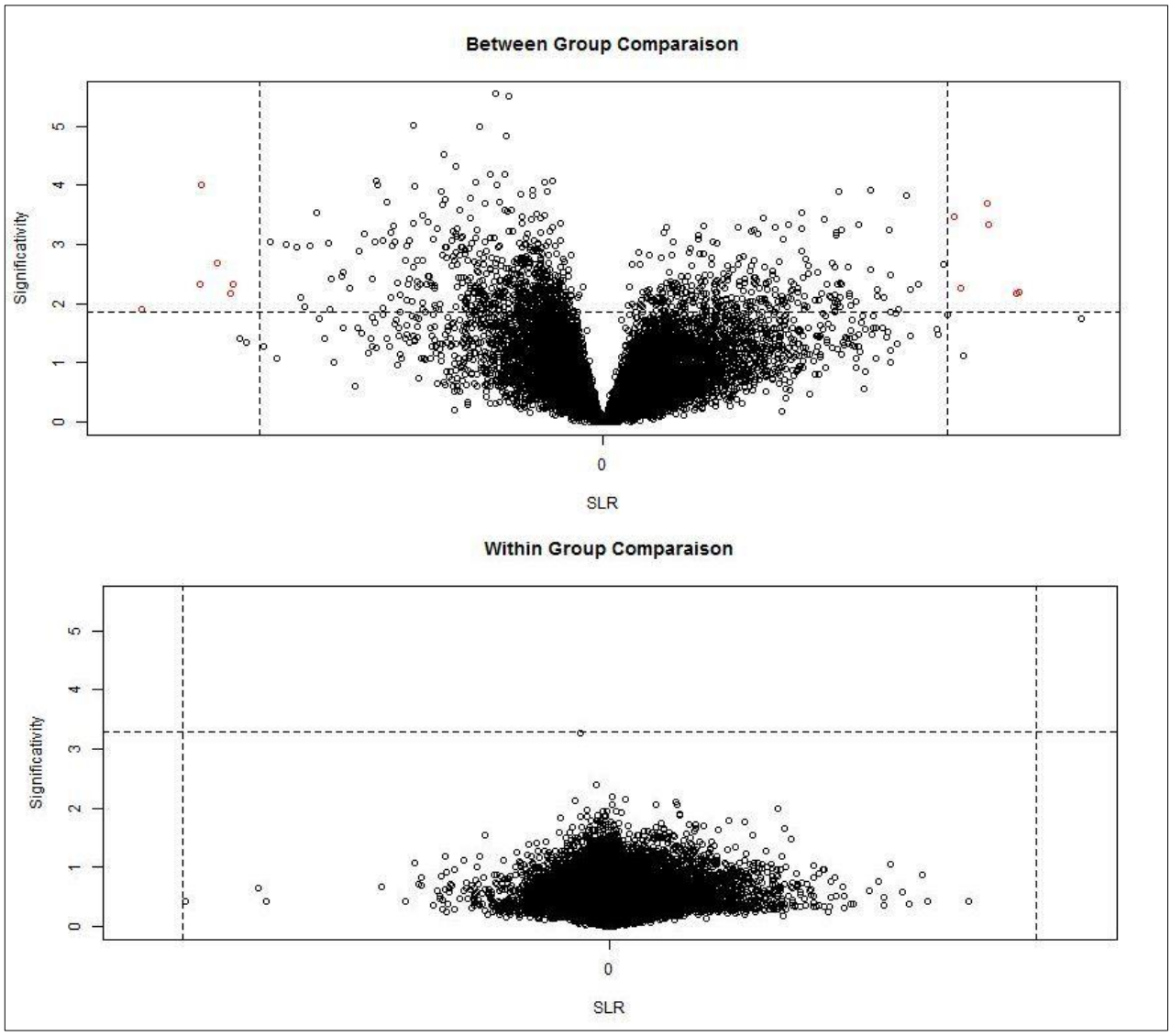


Figure 16. Simple Statistical Test (GSE2443).

The Twilight statistical analysis at FC 1.5 gave 29 probes out of which 28 probes were mapped to genes (Figure 17) [35]. There was no significance in within group comparisons.

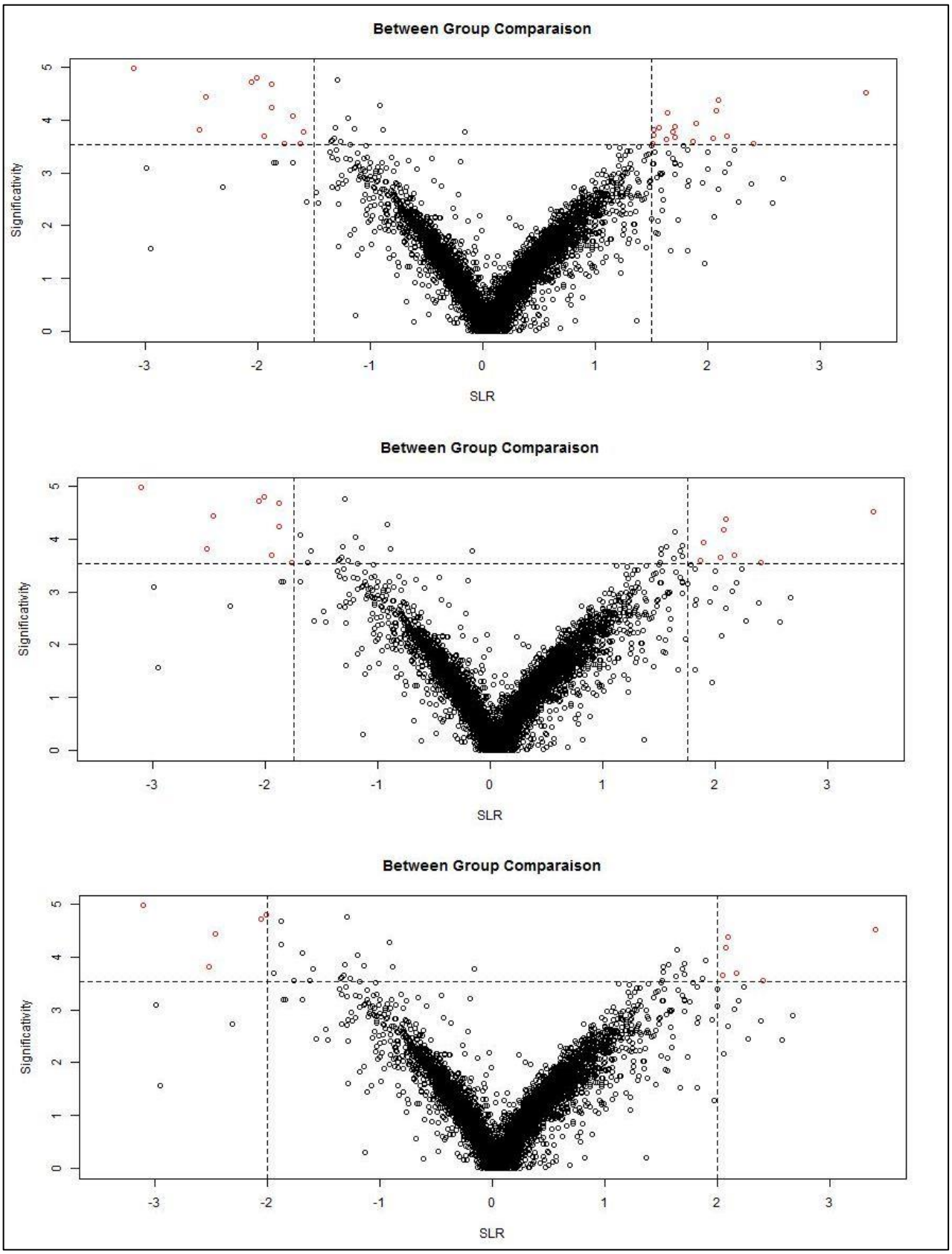


Figure 17.Twilight at logFC cutoff value of 1.5, 1.75 and 2 (Top to Bottom) for GSE2443.

The Linear model analysis gave a significant number of genes only in comparison between groups (Figure 18) but, the expressed genes showed the *p-values* greater than 0.01, which were **not significant** and thus were discarded.

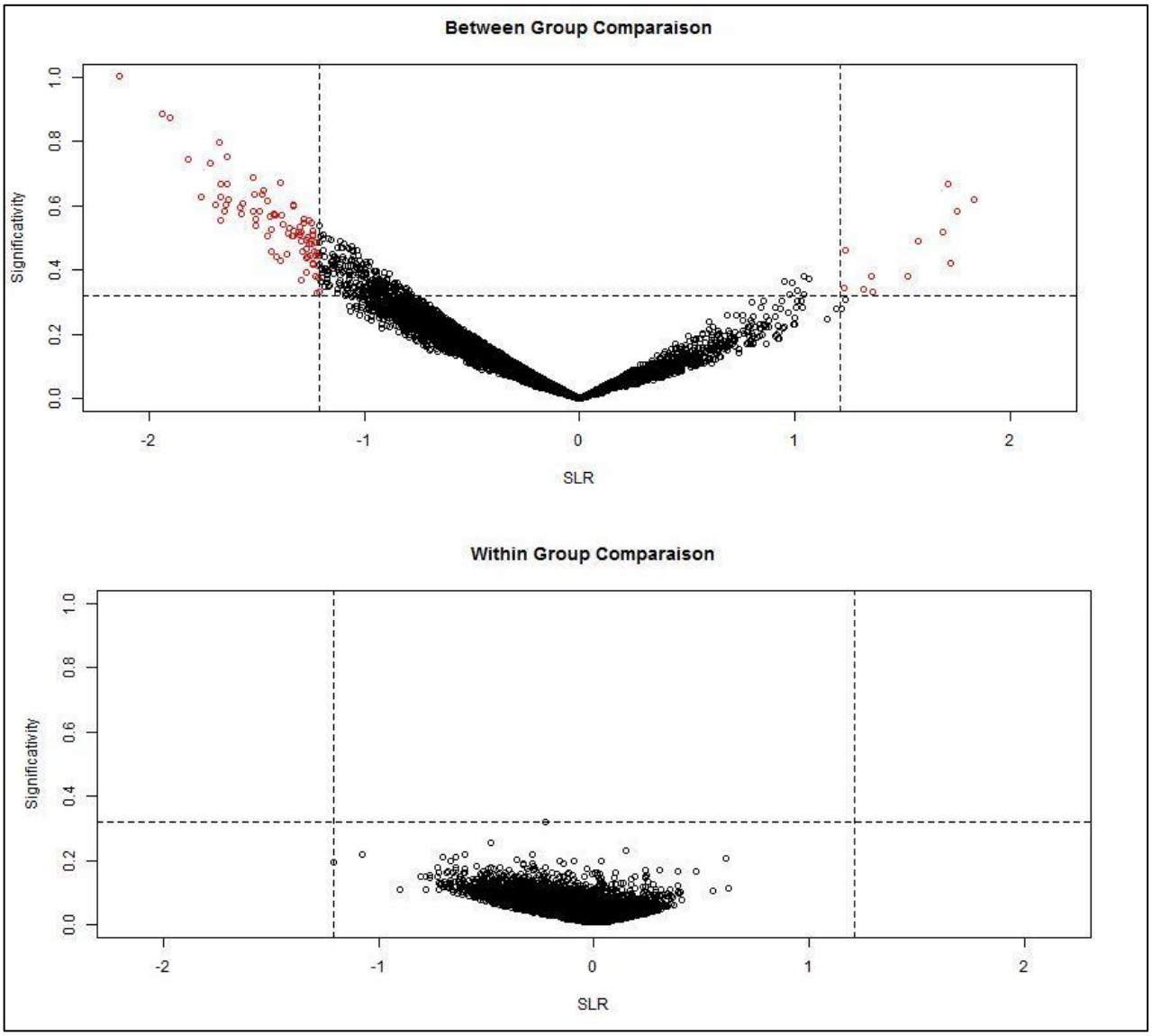


Figure 18. Linear Models for GSE2443.

We identified 8 genes in GSE8702 "**Longitudinal Analysis of Progression to Androgen Independence**" as RPL23 (HGNC:10316), RPS9 (HGNC:10442), DDC (HGNC:2719), KLK3 (HGNC:6364), PLA2G2A (HGNC:9031), KLK2 (HGNC:6363), CCL20 (HGNC:10619), FAM198B (HGNC:25312) (Figure 19, 20).

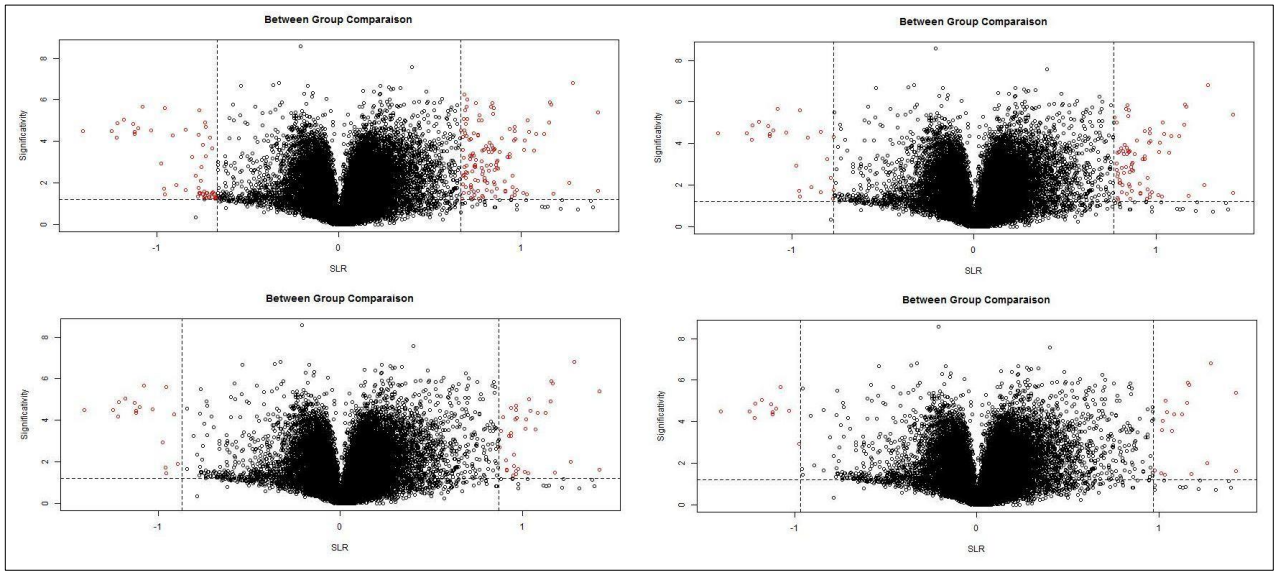


Figure 19. Simple Statistical Test at logFC cutoff values 0.67 (top left), 0.77 (top right), 0.87 (bottom left), 0.97 (bottom right) for GSE8702.

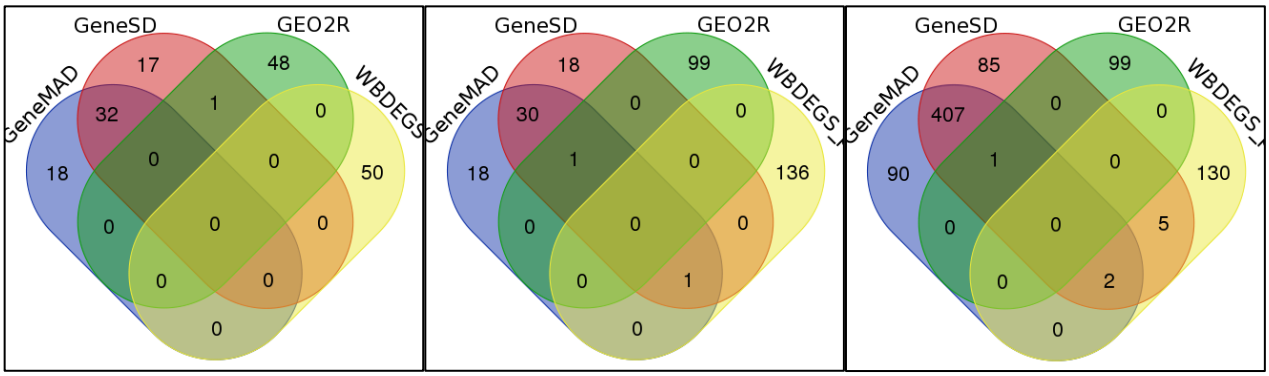


Figure 20. (Left to Right)1. Based on logFC value 2. Top 50 genes from GEO2R, GeneMAD and GeneSD since p - value is less than 0.01 3. Top 500 genes from GeneMAD and GeneSD since p – value is less than 0.01. (GSE8702)

In GSE21887 "**Identification of EP4 as a Potential Target for the Treatment of Castration-Resistant Prostate Cancer Using a Novel Xenograft Model**", We found 27 genes based on the following results (Figure 21, 22).

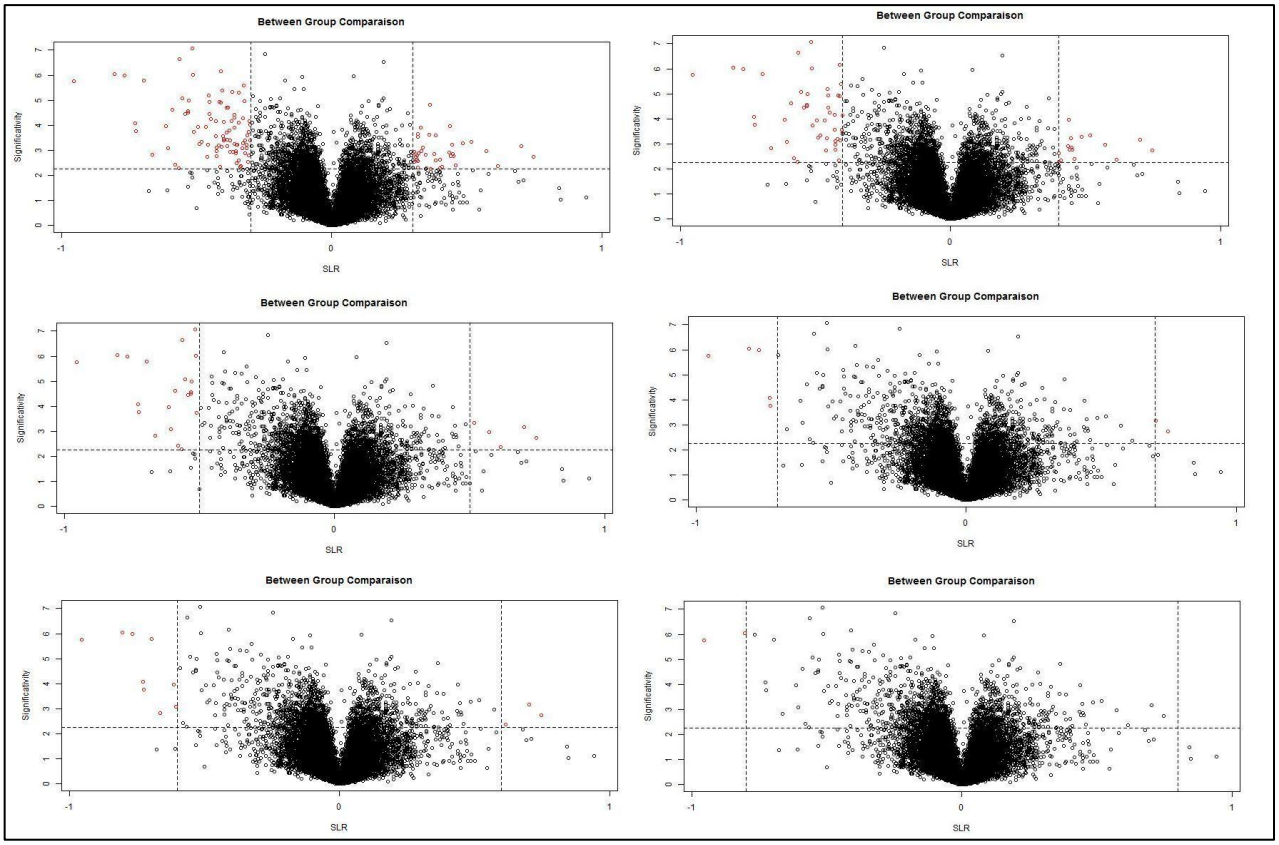


Figure 21. Simple Statistical Test at logFC cutoff values 0.3 (top left), 0.4 (top right), 0.5 (middle left), 0.6 (middle right), 0.7 (bottom left) and 0.8 (bottom right) for GSE21887.

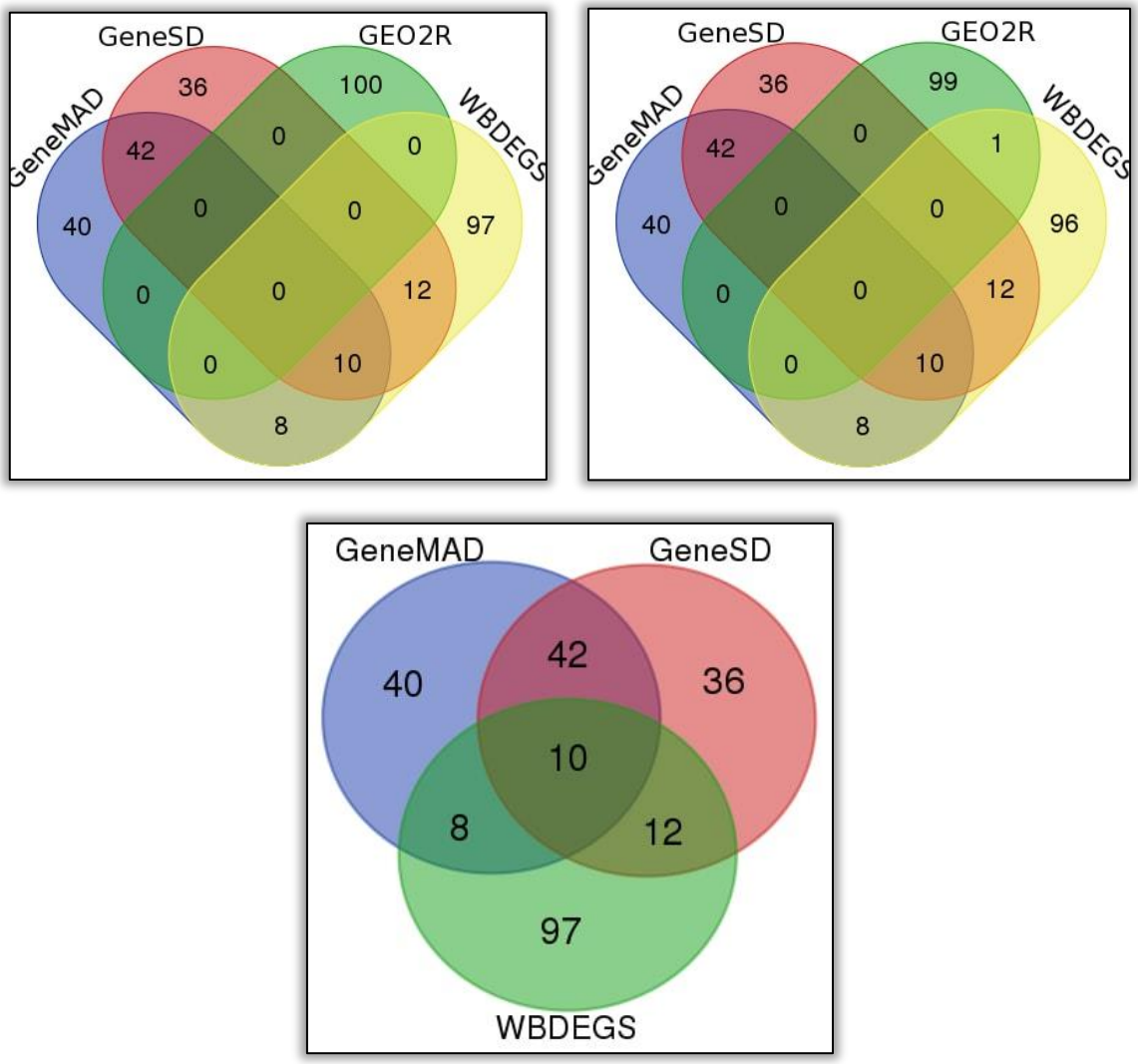


Figure 22. 1. Input of top 100 genes ranked since p - value is led than 0.01 (top left) 2. Input of top 100 genes since p - value is less than 0.05 (top right) 3. Input of top 100 genes since p - value is less than 0.01 (bottom) for GSE21887.

Again, In GSE33316 "**The effect of androgen deprivation on human prostate xenograft tumour LuCaP35**", 13 genes were found based on the results (Figure 23, 24, 25 and 26).

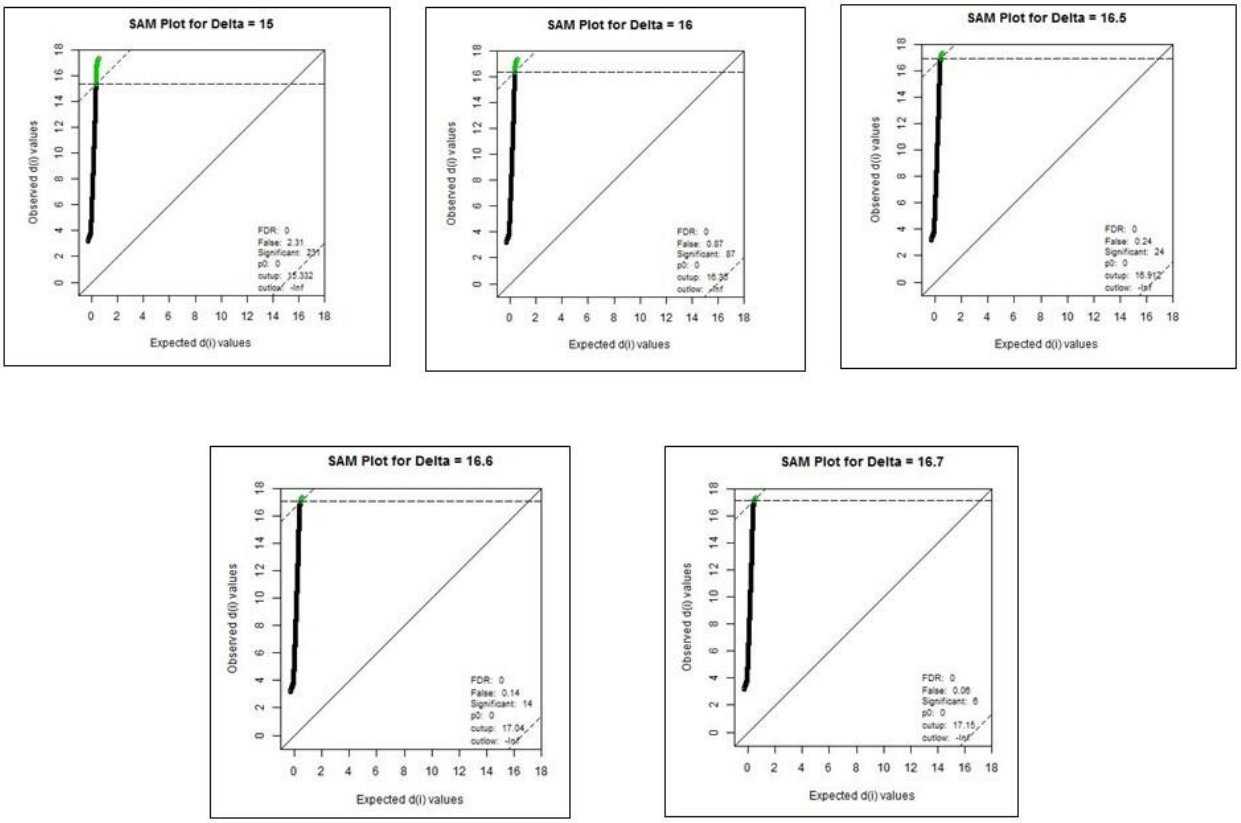


Figure 23. SAM Analysis for GSE33316.

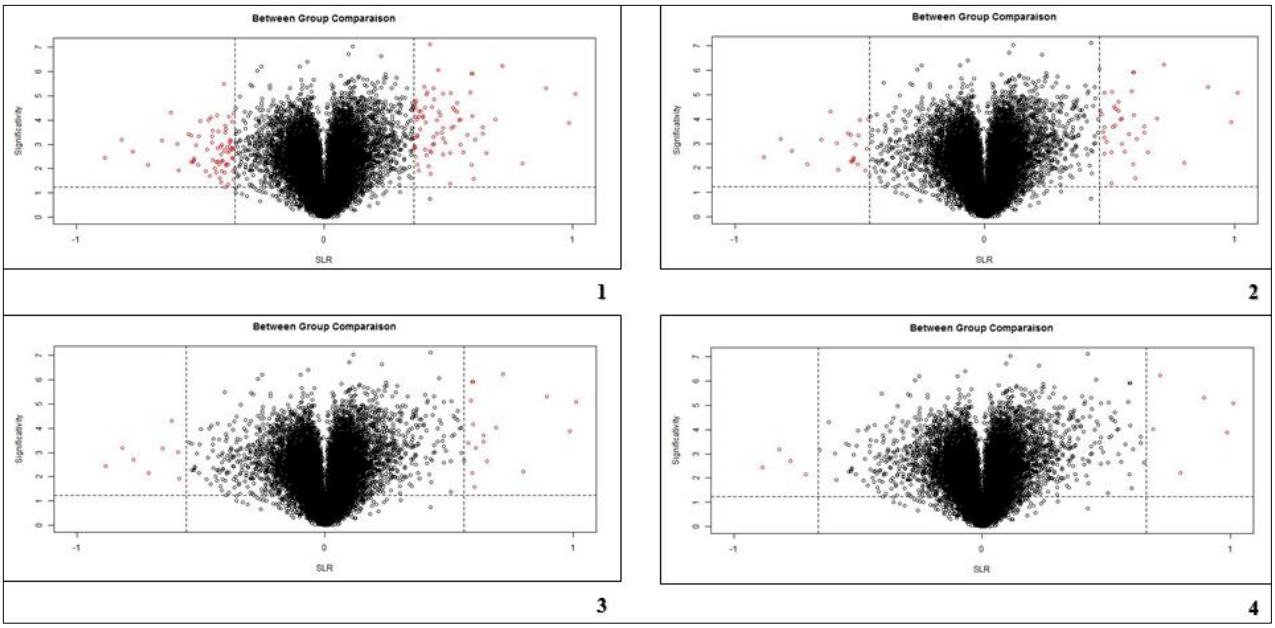


Figure 24. Simple Statistical Test at log FC values 0.36 (1), 0.46 (2), 0.56 (3), 0.66 (4) for GSE33316

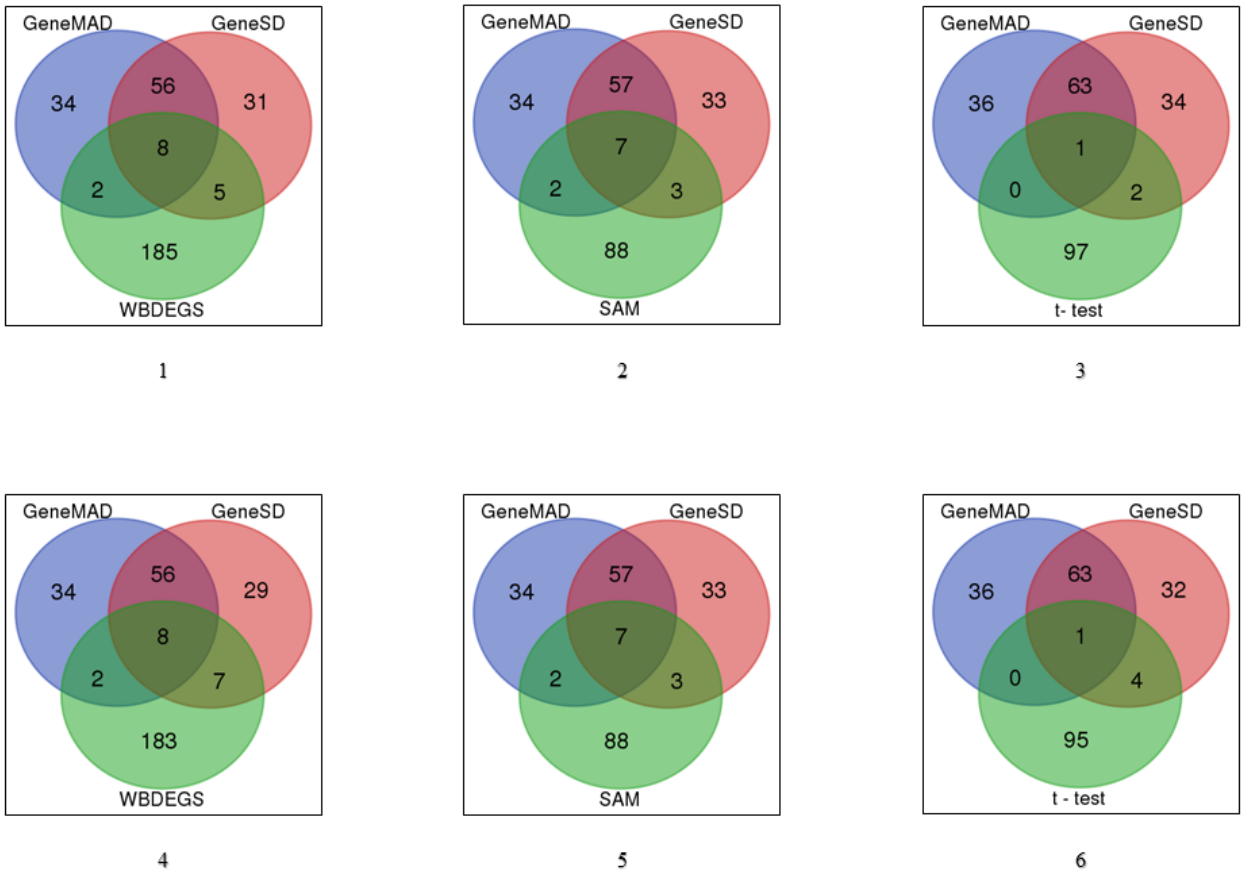


Figure 25. For GSE33316, 1. p - value less than 0.01 (Fig. 1-3).
2. p – value less than 0.05 (Fig. 4-6).

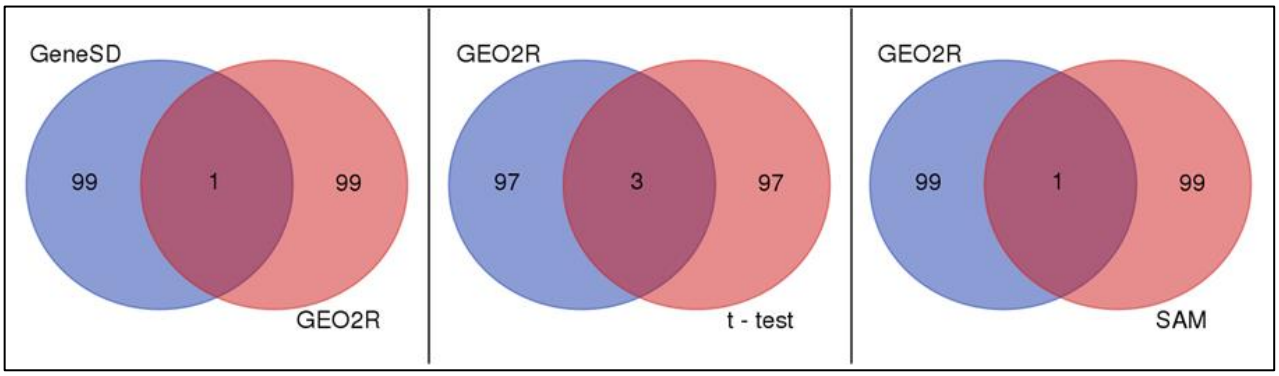


Figure 26. Pair wise Comparisons of genes showing p - value less than 0.05 for GSE33316.

In GSE67537 "**Pathway-based integrative analysis reveals a key role for the hexosamine biosynthetic pathway in castrate resistant prostate cancer with therapeutic implications**", there were no common genes at any level.

So, our analysis for Androgen receptor pathway (AR) ended up with 50 genes in total and only one genes in common across all the five series datasets. Now, for Mitogen activated protein kinases pathway (MAPK), all four series datasets have given some common/significant genes. In GSE20906 "**Gene expression profiles of PNT1a prostate cells expressing FGFR-4 Arg388 or Gly388"**, We found 11 genomic entities based on following results (Figure 27, 28 and 29).

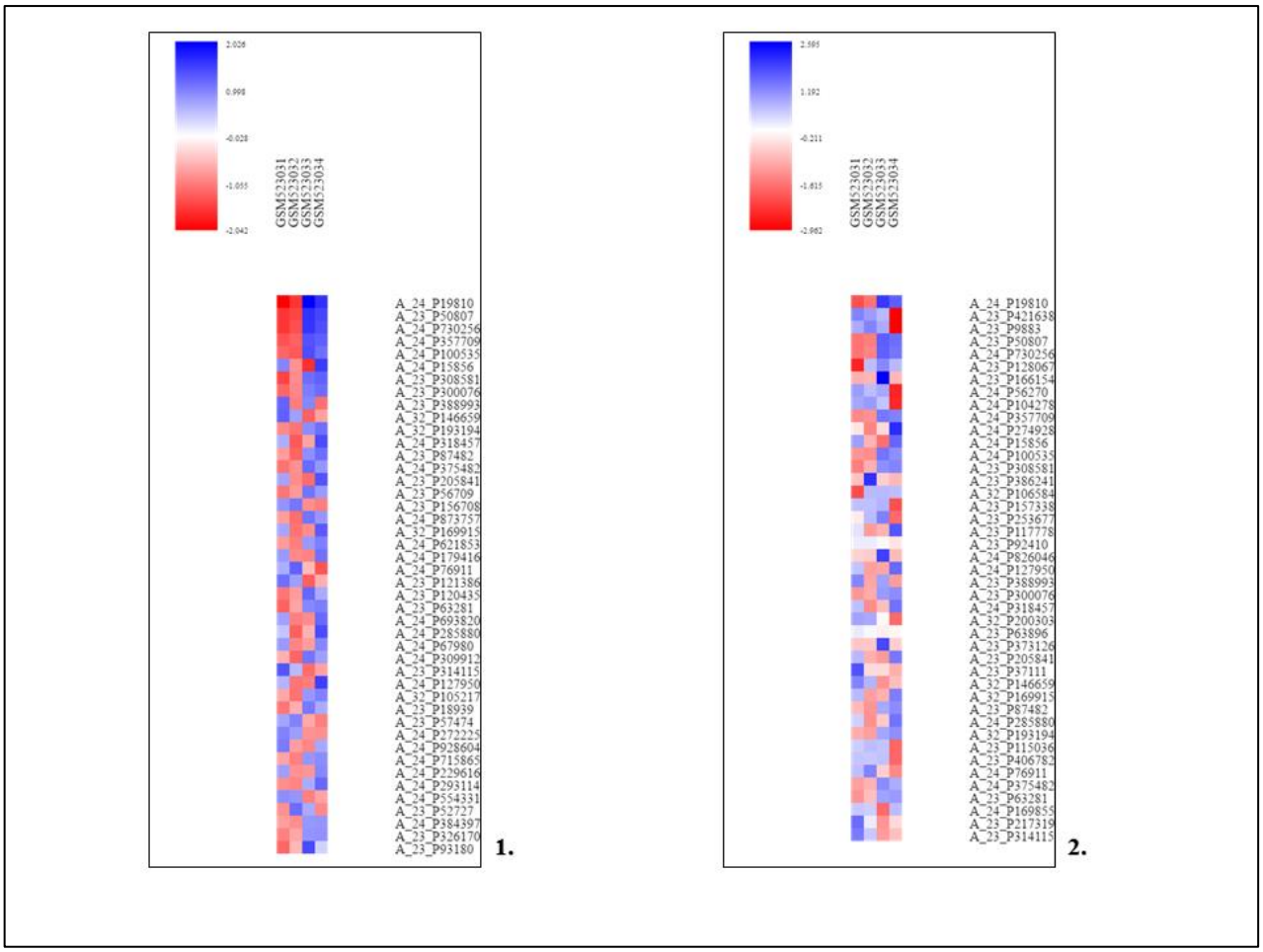


Figure 27. Expression Levels using 1. GeneMAD Top 50 2. GeneSD Top 50 for GSE20906.

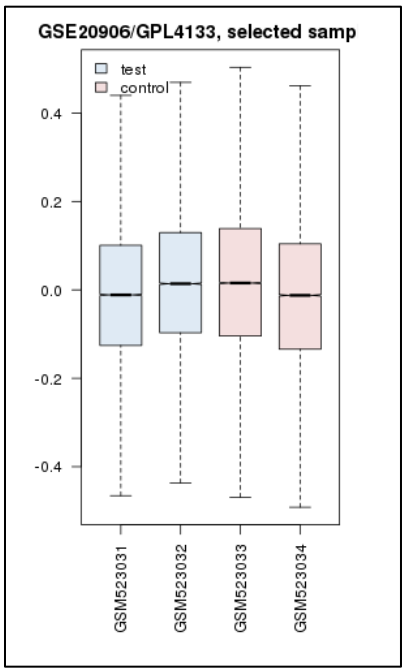


Figure 28. Box and Whiskers Plot generated using GEO2R for GSE20906.

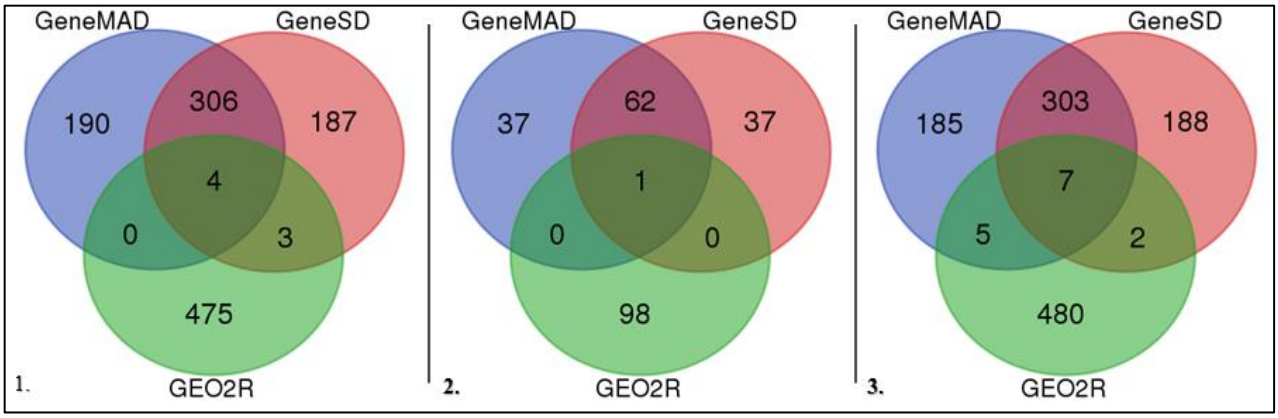


Figure 29. 1. Input of top 500 genes since p - value is less than 0.01. 2. Input of top 100 genes since p - value is less than 0.05. 3. Input of top 500 genes since p - value is less than 0.05. (GSE20906)

In GSE23038 "**Normal prostate cells were immortalized and cultured for 650 days till several transformation hallmarks were observed**", We found 53 genes based on the following results (Figure 30 and 31).

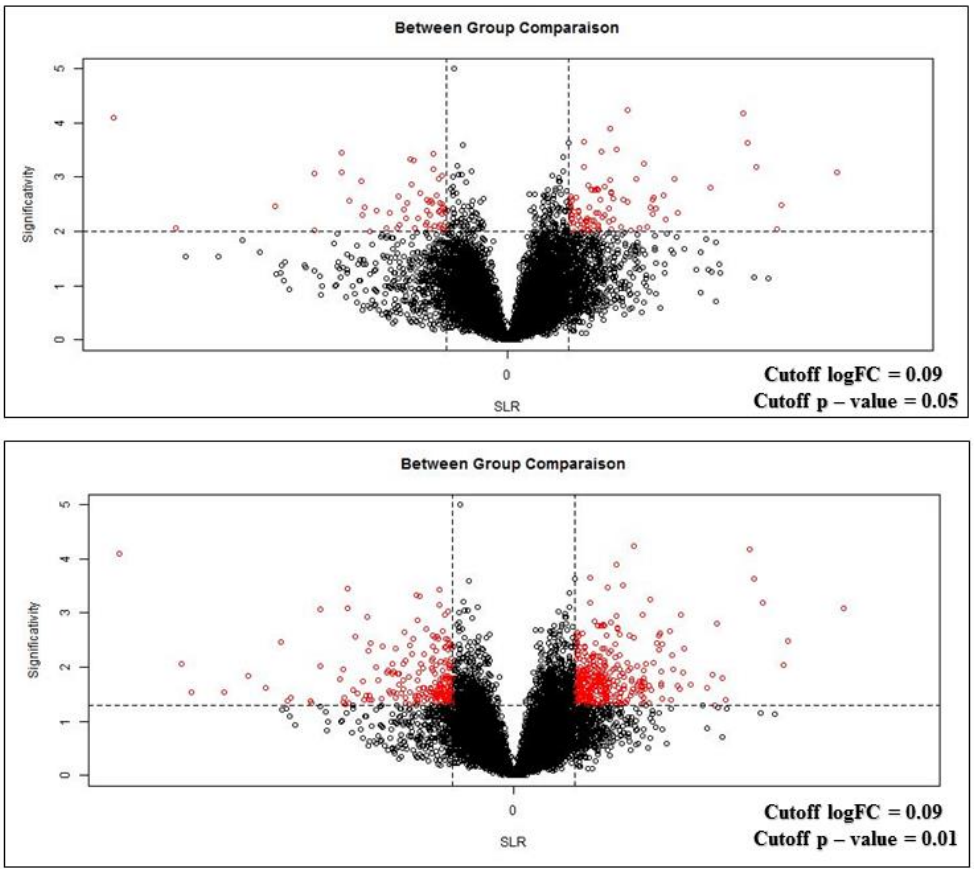


Figure 30. Simple Statistical Test for GSE23038.

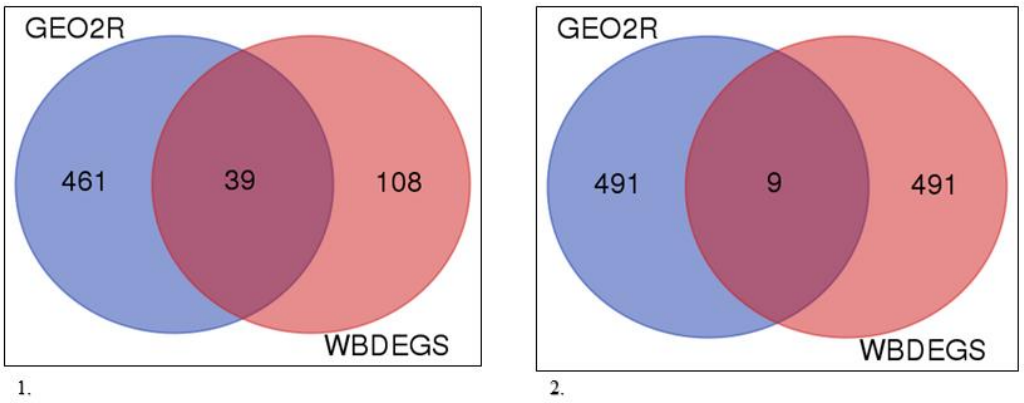


Figure 31. Input of top 500 genes for 1. p -value less than 0.05 2. p – value less than 0.01 for GSE23038.

In GSE29438 "**Oncogenic ETS proteins regulate a Ras/MAPK gene expression program in the absence of MAPK signalling**", We found 4 significant genes based on the following results (Figure 32 and 33).

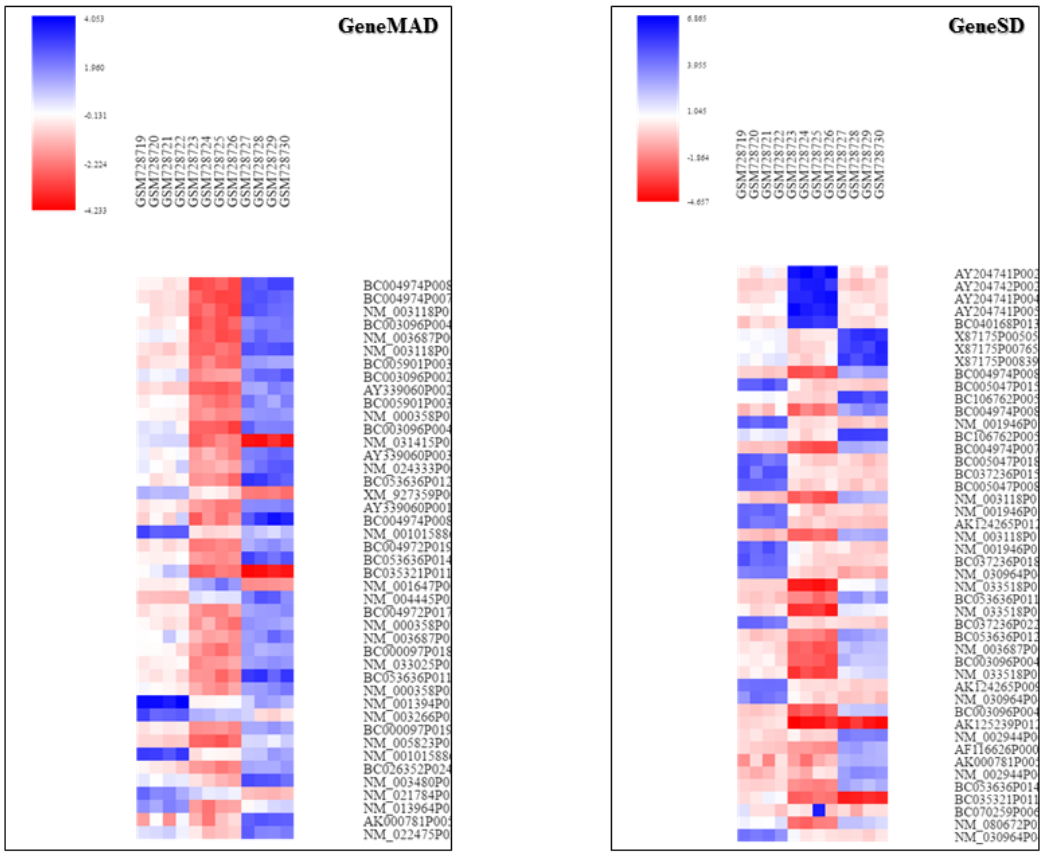


Figure 32. WebMeV (TM4) Analysis for GSE29438.

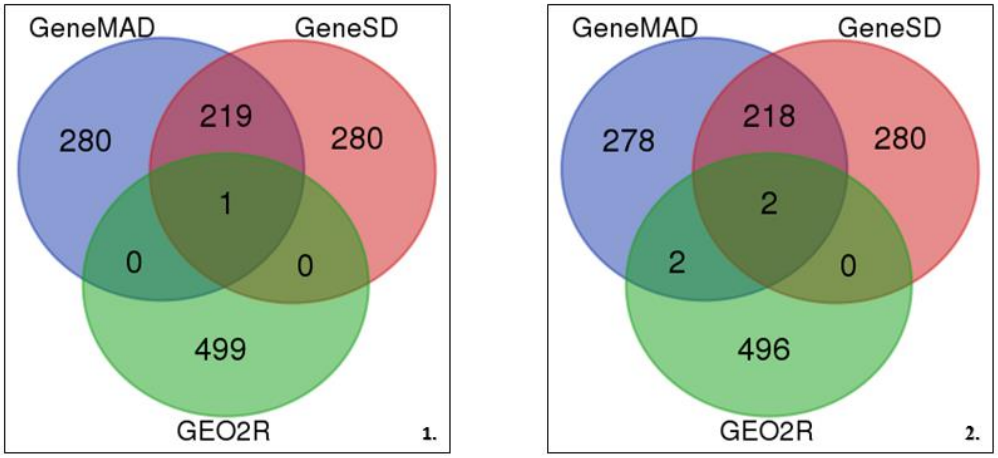


Figure 33. Input of top 500 genes for 1. p – value less than 0.01. 2. p – value less than 0.05 for GSE29438.

In GSE39735 "**Identification of miR-205 targets using an RIP-Chip assay with AGO2 antibody**", we found 156 significant genes based on the following results (Figure 34, 35, 36, 37 and 38).

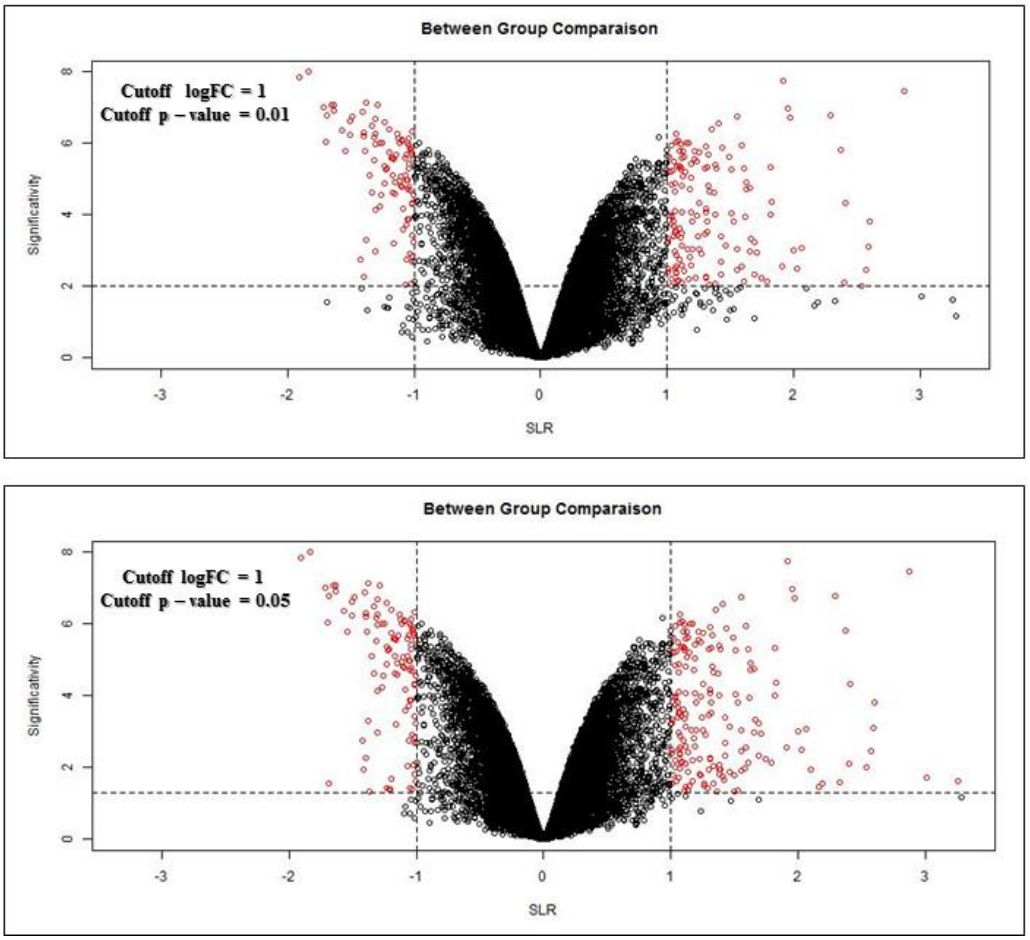
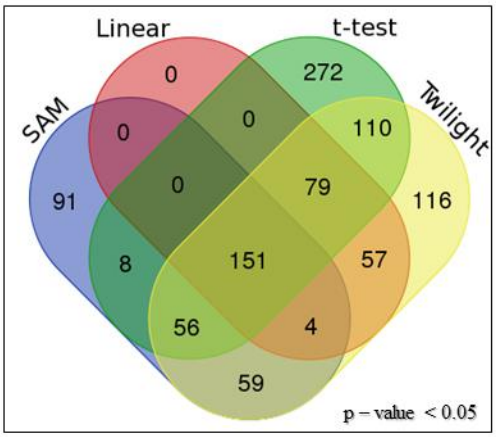


Figure 34. Linear Models for GSE39735.



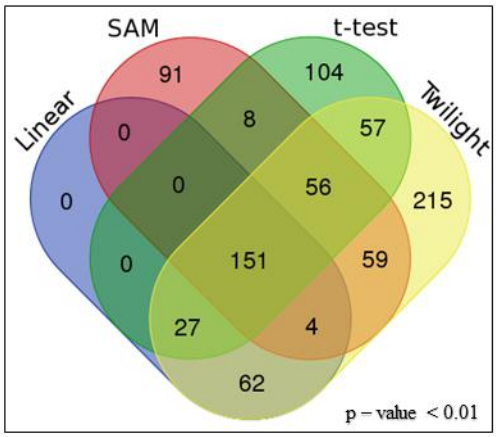


Figure 35. Venn diagram based on p – values for GSE39735.

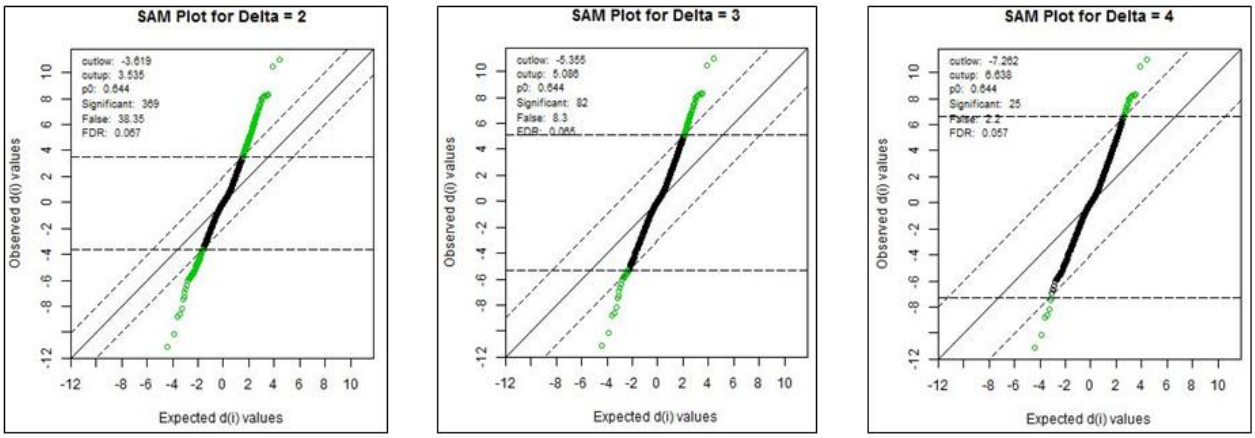


Figure 36. SAM Analysis for GSE39735.

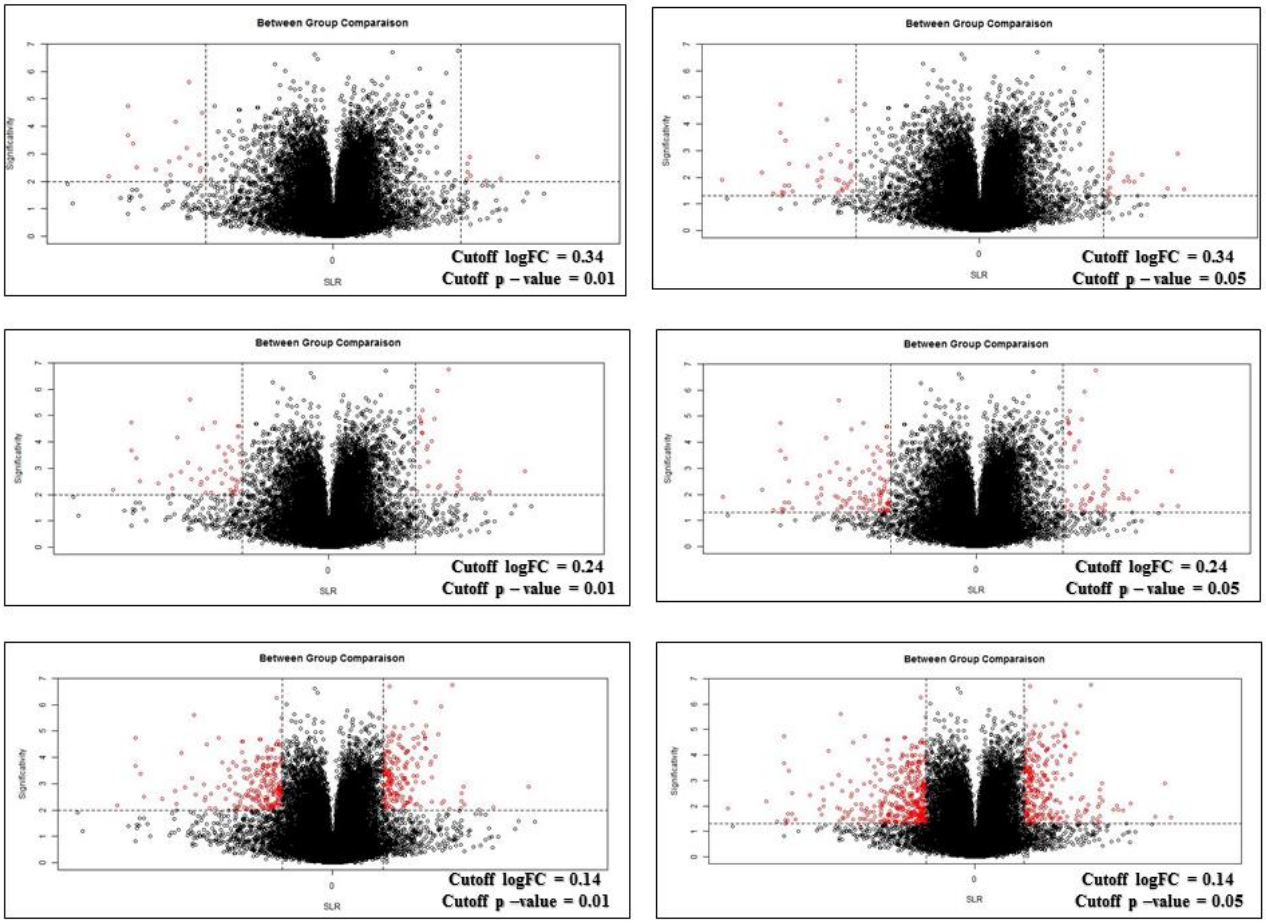
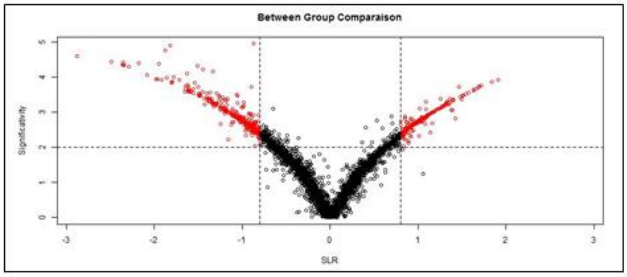
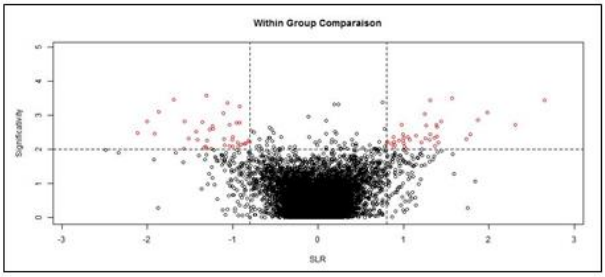


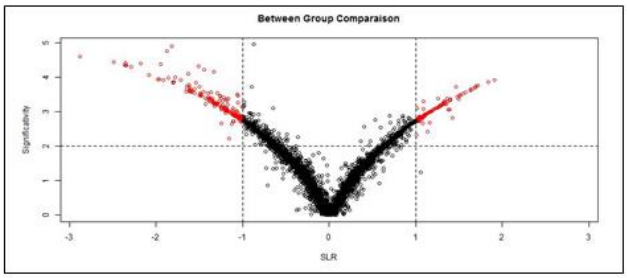
Figure 37. Simple Statistical Test for GSE39735.



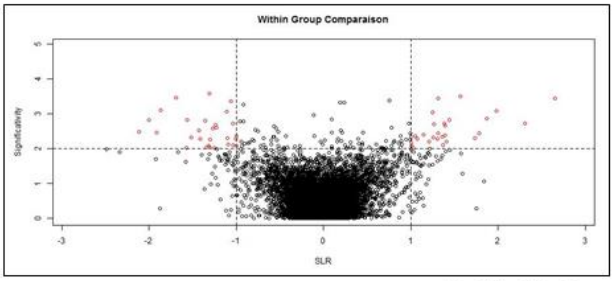
Cutoff logFC = 0.8
Cutoff p – value = 0.01



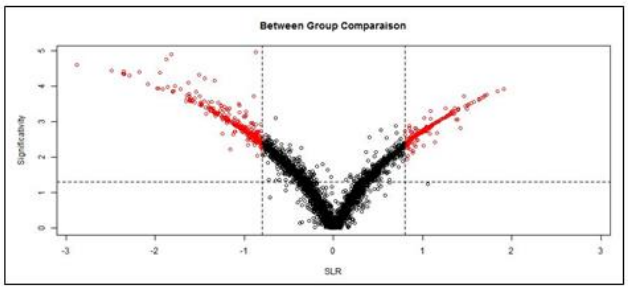
Cutoff logFC = 0.8
Cutoff p – value = 0.01



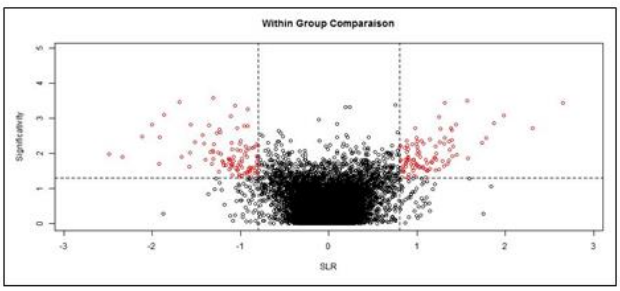
Cutoff logFC = 1
Cutoff p – value = 0.01



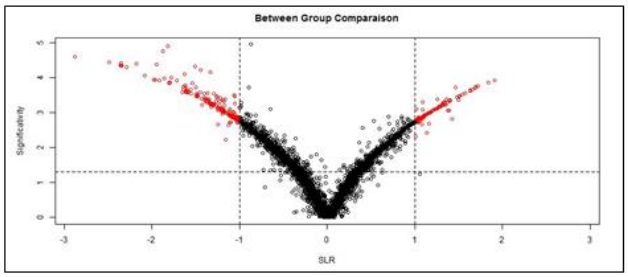
Cutoff logFC = 1
Cutoff p – value = 0.01



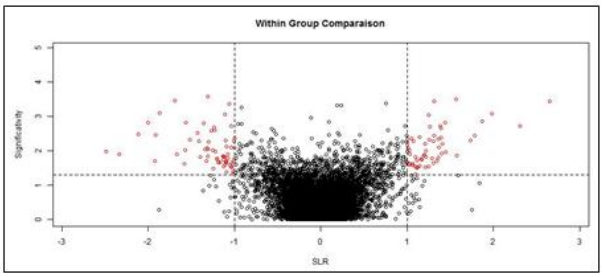
Cutoff logFC = 0.8
Cutoff p – value = 0.05



Cutoff logFC = 0.8
Cutoff p – value = 0.05



Cutoff logFC = 1
Cutoff p – value = 0.05



Cutoff logFC = 1
Cutoff p – value = 0.05

Figure 38. Twilight Models for GSE39735.

So, our analysis for Mitogen activated protein kinases pathway (MAPK) ended up with 224 genes in total with 9 genes in common across all the four series datasets. Now, for Mechanistic target of rapamycin (m-TOR), all three series datasets have given a few more common/significant genes.

In GSE26332 "**Expression data after miR-99a transfection in C4-2 prostate cancer cells**", We have found 52 genes based on the following results (Figure 39, 40 and 41). There was an error in GEO2R due to only two samples available for grouping them into test and control subjects.

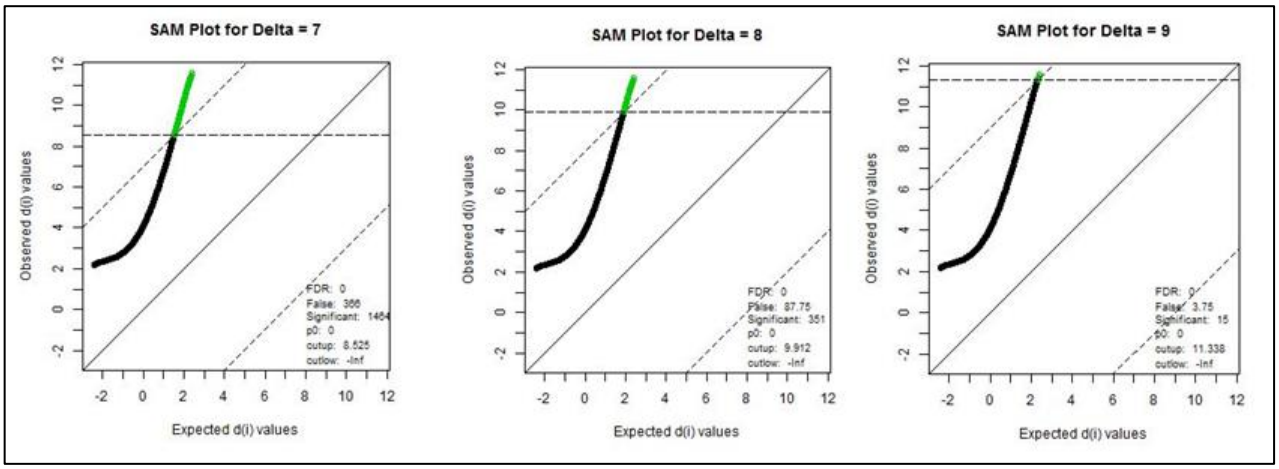


Figure 39. SAM Analysis for GSE26332.

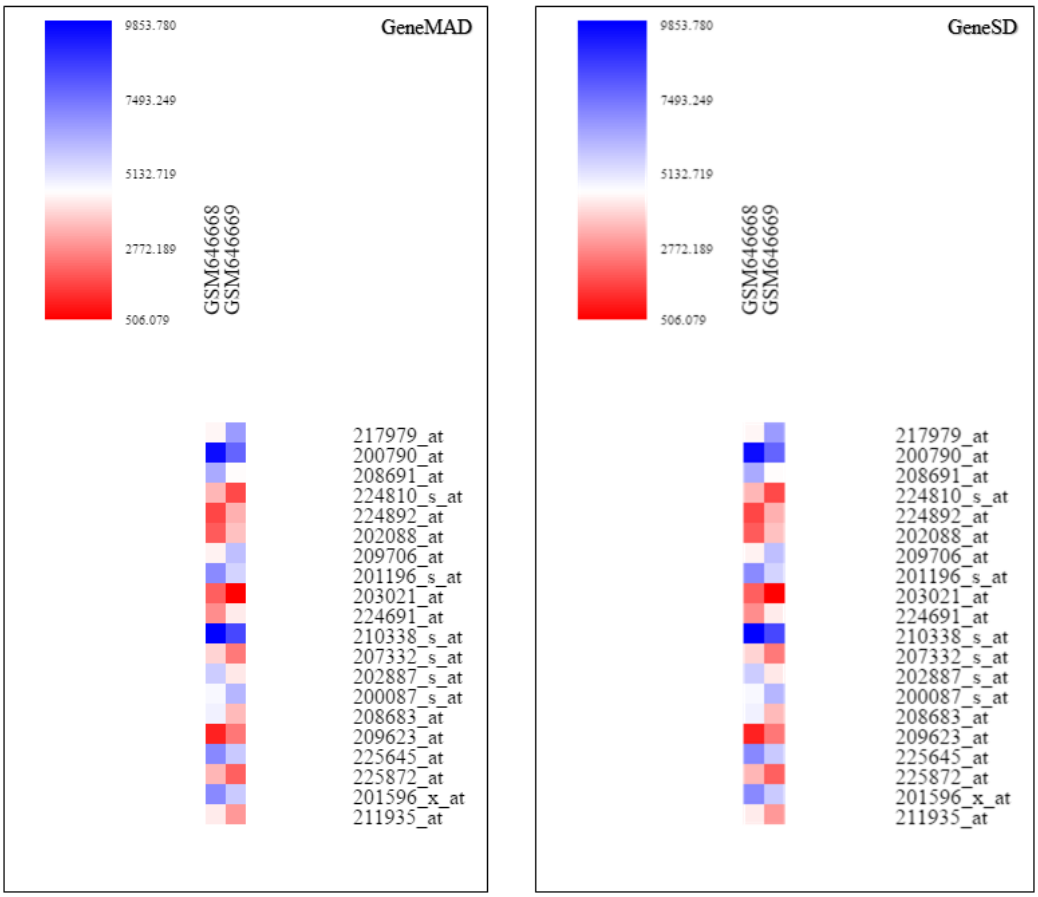


Figure 40. WebMeV (TM4) Analysis for Top20 genes. The results for both GeneMAD and GeneSD are same due to single test and control samples in the dataset.

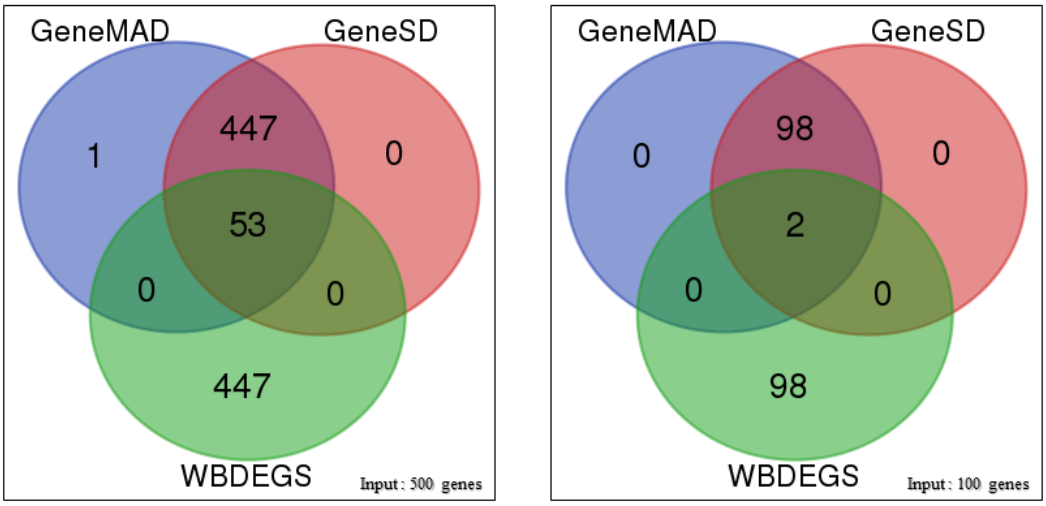


Figure 41. Venn Diagram based on p - value < 0.01 for GSE26332.

In GSE49232 "**Gene expression analysis of in vivo-grown tumours treated with compounds that either de-bulk the tumour or target cancer stem cells**", We found 7 genes based on the following results (Figure 42, 43 and 44).

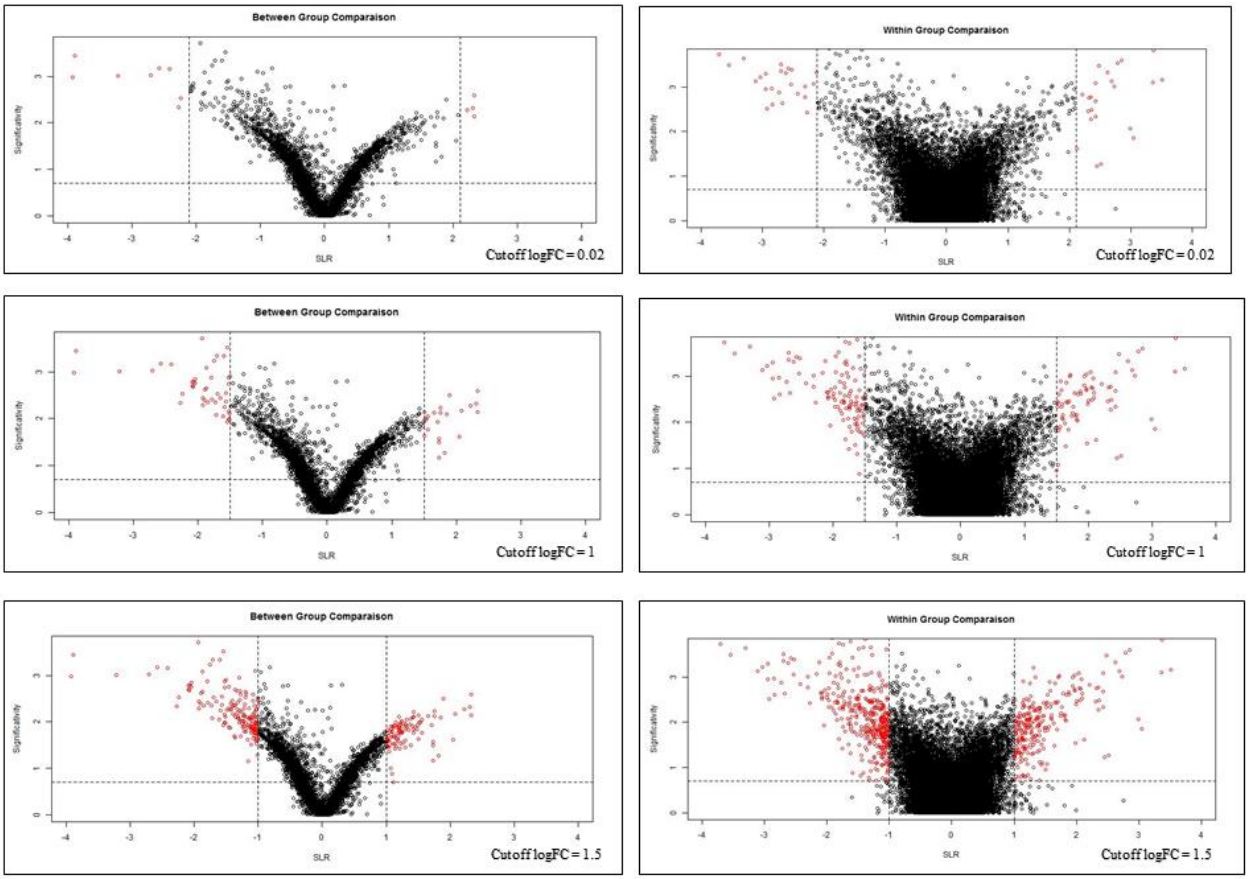


Figure 42. Twilight Models for GSE49232.

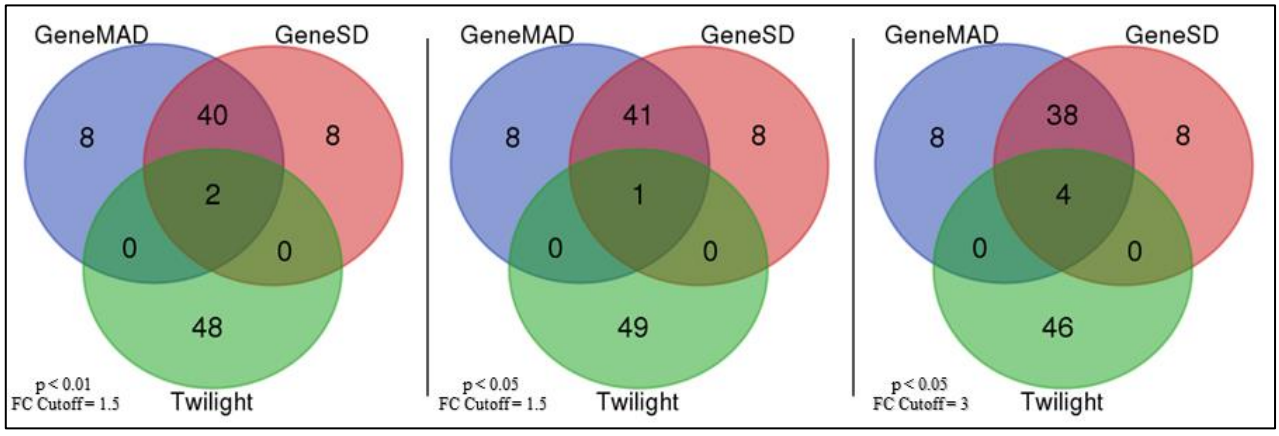


Figure 43. Venn Diagram for top 50 genes.

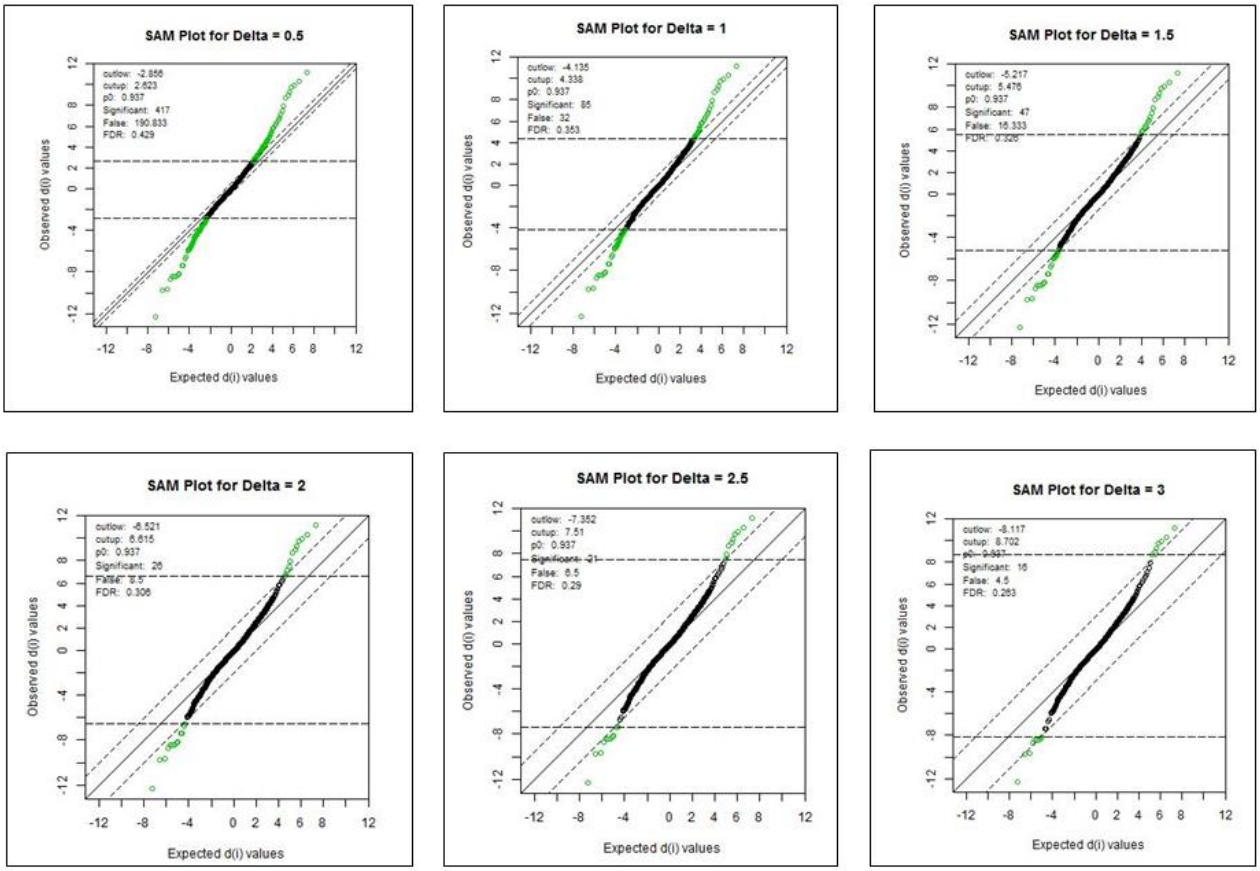


Figure 44. SAM Analysis for GSE49232.

In GSE32875 "**Identification of novel androgen-regulated pathways and mRNA isoforms through genome-wide exon-specific profiling of the LNCaP transcriptome**", there were no genes in common between all the methods.

So, our analysis for Mechanistic target of rapamycin (m-TOR) ended up with 59 genes in total with 3 genes in common across all the three series datasets.

Now, the genes obtained through all these three analysis were identified and were processed for functional annotation so that we can obtain few relevant genes. A total of **349** genes were identified in all pathways with **229 genes in MAPK, 63 genes in mTOR and 51 genes in AR pathway.**

	AR
	mTOR
	MAPK

Table 2. The colour codes for AR, mTOR and MAPK pathway used in the research.

The genes based on these colours were categorized into the three pathways to visualize the interactions between them. The input of STRING was provided as these genes and the networks created for each pathway and multiple- pathway (in the pair of two). This determined the interactions between the genes belonging to same or different pathway.

From the network results obtained from STRING database, only, significant interactions having combined score (calculated as average of scores of neighbourhood on chromosome, gene fusion, phylogenetic co-occurrence, homology, coexpression, experimentally determined interaction, database annotated and automated text mining) greater than or equal to 0.9 was taken into consideration.

The **inter-pathway interactions** include:

AR-mTOR **mTOR-MAPK** **AR-MAPK**

And **intra-pathway interactions** include:

AR-AR **mTOR-mTOR** **MAPK-MAPK**

Results of the Inter-pathway and intra-pathway interactions between genes from STRING:

INTRA-PATHWAY INTERACTIONS

AR-AR	
RPL35A	RPL23
RPS20	RPL29
RPS29	RPS6
RPS29	RPS10
RPS8	RPS6
RPS8	RPS29
RPS10	RPS9
RPS29	EEF2
RPS6	RPL5
RPS8	RPS10
RPS8	EEF2
RPS8	RPL23
RPS6	EEF2
RPL35A	RPL29
RPS20	RPS10
RPS20	RPS29
RPS20	RPL23
RPS20	RPL5
RPL23	RPL29
RPS20	RPS8
RPS8	RPL5
RPL23	RPL5
RPS29	RPS9
RPS6	RPS9
RPL35A	RPL5
RPS20	RPS9
RPS6	RPS10
RPS29	RPL5
RPS20	RPS6
RPS20	EEF2
RPL5	RPL29
RPL5	EEF2
RPL23	RPS6
RPS8	RPS9
RPS29	RPL23
RPS8	RPL29
RPL5	RPS10
RPS29	RPL29
RPL35A	RPS8
RPL23	EEF2

mTOR-mTOR	
SEC11A	SPCS1
NDUFA12	NDUFB5
EDN1	HIF1A
HIF1A	HSP90AB1
BNIP3	HIF1A
PRKACB	YWHAZ
BRK1	ACTR3
HSP90AB1	HSPA9
CDK1	CALD1
PRKACB	GNB4
CDK1	PRKACB
PRKACB	RAB11A
TOMM5	HSPA9
SRSF7	POLR2L
EDN1	PRKACB

MAPK-MAPK	
FAU	RPL31
IL6R	IL6
HSP90B1	LRP1
UCHL5	KIAA0368
STX2	VAMP8
NCOR1	KAT2B
CDH3	CDH1
C1D	NCOR1
HSPA4	HSP90B1
FAU	EIF5B
DHFR	MTR
ST3GAL6	B4GALT1
CDH11	CDH3
RCOR1	HSPA4
RPL31	EIF5B
CDH11	CDH1
IRF6	EGR1
IRF6	KAT2B
SMN2	SMN1
CDK19	NCOR1
PDE3A	NT5E
QPRT	NT5E
HIST1H2BK	HIST1H2BD

RPL5	RPS9
EEF2	RPS9
RPS20	RPL35A
RPL35A	RPS6
RPL35A	RPS29
RPS6	RPL29
RPS10	RPL29
RPS10	EEF2
RPL35A	RPS10
RPL23	RPS10
RPL35A	EEF2
RPL23	RPS9
EEF2	RPL29
RPL35A	RPS9
RPS9	RPL29
CCL20	OPRK1

INTER-PATHWAY INTERACTIONS

AR-mTOR	
IGFBP3	KLK3
RPL29	EIF5
RPS6	SERBP1
SERBP1	RPS10
RPS29	SERBP1
RPS20	SERBP1
DDC	MAOA
SERBP1	EEF2
NCAPD3	CDK1
RPS8	SERBP1
SERBP1	RPS9
HIF1A	CP
RPS6	EIF5
RPL29	SPCS1
RPS29	SPCS1
RPL23	SPCS1
RPL35A	EIF5
RPL5	EIF5
RPS20	EIF5
RPS8	EIF5
RPS10	EIF5
RPS20	SPCS1
RPL35A	SPCS1
RPL23	EIF5
RPS10	SEC11A
RPS29	SEC11A
RPS6	SPCS1
CDK1	CEP57
RPS8	SPCS1
RPL5	SPCS1
RPL5	SEC11A
RPS6	SEC11A
NCAPD3	SET
RPS10	SPCS1
RPL29	SEC11A
RPS20	SEC11A
RPL35A	SEC11A
RPL23	SEC11A
PLA2G2A	CHPT1
RPS29	EIF5
RPS8	SEC11A

mTOR-MAPK	
EDN1	EDNRA
CDK1	MNAT1
CDK1	CDK6
CDK1	NCOR1
DHFR	CDK1
EDN1	IL6
EIF5B	EIF5
FAU	SERBP1
FAU	EIF5
FAU	SEC11A
FAU	HSP90AB1
FAU	SPCS1
GPX2	GGT1
HIF1A	HSPA4
HIF1A	NT5E
HIF1A	NCOR1
HSP90AB1	HSPA4
HSP90AB1	AKT3
HSP90AB1	HSPA13
HSP90B1	HSPA9
IGFBP3	ADAM12
POLR2L	MNAT1
PRKACB	MNAT1
RPL31	EIF5
RPL31	SPCS1
RPL31	SEC11A
TKT	RPIA
TTC32	HSP90AB1

AR-MAPK	
RPL31	RPS8
FAU	RPL23
FAU	RPL29
FAU	EEF2
FAU	RPS29
FAU	RPS20
RPL31	RPL23
RPL31	RPL5
RPL35A	RPL31
RPL31	RPS6
FAU	RPL5
RPL31	RPL29
FAU	RPS8
FAU	RPS9
FAU	RPS10
FAU	RPS6
RPL31	RPS10
RPS20	RPL31
FAU	RPL35A
RPL31	RPS29
RPL31	EEF2
RPS8	EIF5B
RPL5	EIF5B
RPL31	RPS9
RPL23	EIF5B
RPS20	EIF5B
RPS6	EIF5B
RPS29	EIF5B
LIFR	IL6
RPL29	EIF5B
GDF15	EGR1
RPS10	EIF5B
SARNP	RBMS1
KLK3	CDK6
KLK3	EGR1
RPL35A	EIF5B

The same analysis was obtained from GeneMANIA which helps to generate genetic interactions and predict the function of the genes. Same type of input was provided with genes belonging to different pathways and the results were generated in the same format but with the weight value between the two genes. The weight represented the strength of interactions between them and this was processed by discarding the weights greater than 0.1 which is a significant interaction.

Results of the inter-pathway and intra-pathway interactions between genes from GeneMANIA:

INTRA-PATHWAY INTERACTIONS

AR-AR	
RPS20	RPS6
RPS8	RPS6
RPL29	RPL23

mTOR-mTOR	
SEC11A	EIF5
CDK1	ACTR3
YWHAZ	ACTR3
YWHAZ	SERBP1
HIF1A	BNIP3
SEC11A	SPCS1
HIF1A	MUC1

MAPK-MAPK	
VAMP8	STX2
UCHL5	KRT80
MNAT1	TRIM5
NR4A2	PRKAA2
GGT1	GGT2
GGTLC2	GGT2
GGTLC2	GGT1
GGTLC1	GGT2
GGTLC1	GGT1
GGTLC1	GGTLC2
PRKAA2	HDLBP
GCH1	DHFR
B4GALT1	B4GALT6
TGFBI	CPA4
NR4A2	RCOR1
IL6	IL6R
KCNJ2	TRAK2
SMN1	SMN2

INTRA-PATHWAY INTERACTIONS

AR-mTOR	
IGFBP3	KLK3
RPL29	EIF5
PRKACB	RPS6
ACTR3	RPL35A

mTOR-MAPK	
HIF1A	TUFT1
HIF1A	B4GALT1
YWHAZ	RASSF2
IGFBP3	ADAM12
IARS	METTL18
SRSF7	CDK19
ELF3	EHF
ACTR3	EPCAM
KAT2B	SERBP1
KAT2B	RAB11A
PGK1	TNC
EDNRA	EDN1
MTR	DAZAP2

AR-MAPK	
EGR1	GDF15
AFF3	AFF2

Discussion

The final genes (one gene in AR, nine genes in MAPK and three genes in mTOR) that we found from the analysis, which are suspected to play a major role in PC in all the three pathways namely AR, mTOR and MAPK are following:

EDN1	**RPS6**	**SERBP1**	**UCHL5**
EIF5	**RPS8**	**NDUFA12**	
RPL23	**RPS20**	**CDK1**	
RPL29	**PRKACB**	**EIF5B**	

These final genes were then identified in the pathways in STRING database which creates the networks for association of functional proteins. The hubs were identified in the networks which are the genes which are connected to multiple genes. So, these genes influence the functions of each other. We identified the interactions between the genes based on the three pathways taken into considerations – AR, MAPK and mTOR. The inter-pathway and intra-pathway interactions here also are depicted using the three colours.

The analysis led to the reduction in number of genes from expression data to pathway data and finally to the common 13 genes obtained. These genes are then further analysed based on their functional annotation to retrieve their importance in the regulation of prostate cancer in the AR, mTOR and MAPK pathway. These genes were identified based upon the results of Venn diagram which found the common genes between the 3 tools used for analysis of the gene expression data. Based on these diagrams, the probe ID's were gathered and their corresponding gene was identified from DAVID tool by gene ID conversion. The genes obtained thus were then identified on the network and their connectivity is studied and the non-significant values are discarded.

The functional significance of the genes and their regulations are identified by finding the connectivity to different genes. The hubs are also identified which are the genes which are connected to various other genes in the network (Table 5).

The common genes inter-connections obtained from STRING and GeneMANIA are the putative genes which play an important role in mediating the progression of the prostate cancer between the AR, mTOR and MAPK pathways (Table 3 & 4).

COMMON INTER-CONNECTIONS:

AR-AR		mTOR-mTOR		MAPK-MAPK	
RPS20	RPS6	SEC11A	SPCS1	VAMP8	STX2
RPS8	RPS6	BNIP3	HIF1A	IL6	IL6R
RPL29	RPL23			SMN1	SMN2

Table 3. Common Intra - pathway connections between STRING and GeneMANIA for all three pathways.

AR-mTOR		mTOR-MAPK	
IGFBP3	KLK3	IGFBP3	ADAM12
RPL29	EIF5	EDNRA	EDN1

Table 4. Common Inter - pathway connections between STRING and GeneMANIA for all three pathways.

Hub Gene	Nodes
PRKACB	CALD1
	CDK1
	CDK6
	CEP57
	DHFR
	EDN1
SERBP1	EEF2
	EIF5
	FAU
	FAU
	GNB4
	KIAA0368
	MNAT1
	MNAT1
CDK1	NCOR1
	NDUFB5
	PRKACB
	RAB11A
	RPL23
	RPL29
	RPL31
	RPL35A
EIF5B	RPL5
	RPS10
	RPS10
	RPS20
	RPS20
	RPS29
	RPS29
	RPS6
	RPS6
	RPS8
	RPS8
	RPS9

Table 5. List of Hub genes with their respective nodes across AR, MAPK and mTOR pathway.

Table 6. Gene Description of the common genes found in all three pathways

Gene Symbol	Gene Name	HGNC Symbol	PANTHER Family/Subfamily	PANTHER Protein Class	UniProtKB
CDK1	cyclin dependent kinase 1	1722	CYCLIN-DEPENDENT KINASE 1 (PTHR24056)	non-receptor serine/threonine protein kinase(PC00167); non-receptor tyrosine protein kinase(PC00168)	P06493
EDN1	endothelin 1	3176	ENDOTHELIN-1 (PTHR13874)		P05305
EIF5	eukaryotic translation initiation factor 5	3299	EUKARYOTIC TRANSLATION INITIATION FACTOR 5 (PTHR23001)	G-protein modulator (PC00022); translation initiation factor(PC00224)	P55010
EIF5B	eukaryotic translation initiation factor 5B	30793	EUKARYOTIC TRANSLATION INITIATION FACTOR 5B (PTHR43381)	G-protein(PC00020); hydrolase(PC0012); translation elongation factor(PC00222);translation initiation factor(PC00224)	O60841
NDUFA12	NADH:ubiquinone oxidoreductase subunit A12	23987	NADH DEHYDROGENASE [UBIQUINONE] 1 ALPHA SUBCOMPLEX		Q9UI09

			SUBUNIT 12 (PTHR12910)		
PRKACB	protein kinase cAMP-activated catalytic subunit beta	9381	CAMP-DEPENDENT PROTEIN KINASE CATALYTIC SUBUNIT BETA (PTHR24353)		P22694
RPL23	ribosomal protein L23	10316	60S RIBOSOMAL PROTEIN L23 (PTHR11761)	ribosomal protein(PC00202)	P62829
RPL29	ribosomal protein L29	10331	60S RIBOSOMAL PROTEIN L29 (PTHR12884)	ribosomal protein(PC00202)	P47914
RPS20	ribosomal protein S20	10405	40S RIBOSOMAL PROTEIN S20 (PTHR11700)	ribosomal protein(PC00202)	P60866
RPS6	ribosomal protein S6	10429	40S RIBOSOMAL PROTEIN S6 (PTHR11502)		P62753
RPS8	ribosomal protein S8	10441	40S RIBOSOMAL PROTEIN S8 (PTHR10394		P62241
SERBP1	SERPINE1 mRNA binding protein 1	17860	PLASMINOGEN ACTIVATOR INHIBITOR 1 RNA-BINDING PROTEIN (PTHR12299)	RNA binding protein(PC00031)	Q8NC51
UCHL5	ubiquitin C-terminal	19678	UBIQUITIN CARBOXYL-	cysteine protease(PC00081)	Q9Y5K5

	hydrolase L5		TERMINAL HYDROLASE ISOZYME L5 (PTHR10589)		

The information about the gene, its full name, HGNC and UniProtKB ID, Panther protein class and family is resulted from the Panther database which allows the analysis of protein and its functions (Table 6).

The functional significance of these genes is obtained through the Gene ontology terms i.e. molecular function, biological process, cellular component and protein class (Table 7, 8 ,9 and 10). Panther database which provides information regarding protein functional analysis through evolutionary relationships was used (Figure 45).

FUNCTIONAL ANNOTATION:

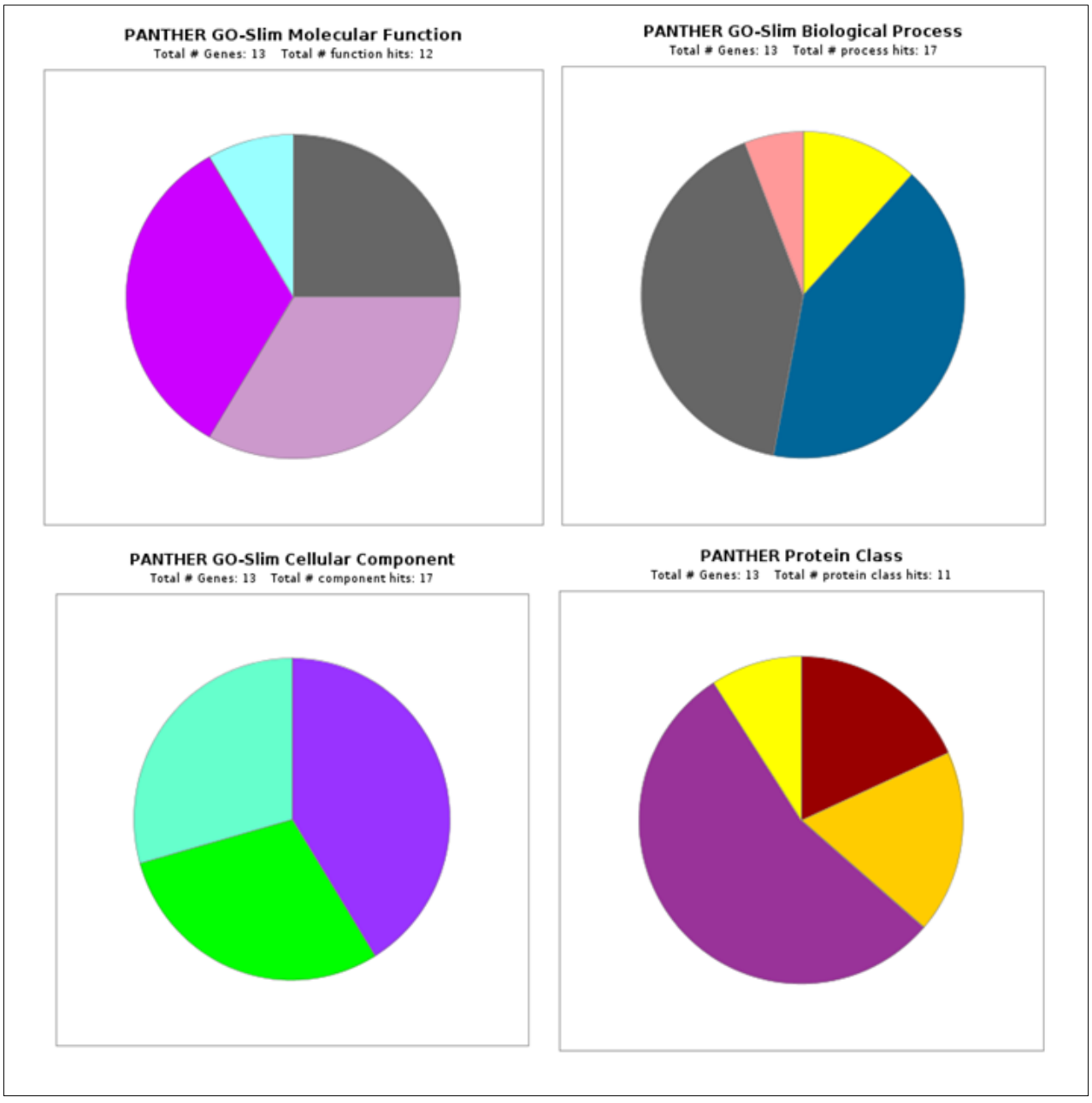


Figure 45. Enrichment Analysis of genes using PANTHER– GO [1] [3] [47] (Table 7-10).

-	-	-	-
Binding (GO:0005488)	**Catalytic Activity (GO:0003824)**	**Structural Molecule Activity (GO:0005198)**	**Translation Regulator Activity (GO:0045182)**
RPL23	EIF5	RPL23	EIF5
EIF5	EIF5B	RPS20	
SERBP1	UCHL5	RPL29	
	CDK1	RPS8	

Table 7. PANTHER GO – slim **Molecular Function**.

-	-	-	-
Cellular Component Organization or Biogenesis (GO:0071840)	**Cellular Process (GO:0009987)**	**Metabolic Process (GO:0008152)**	**Multicellular Organismal Process (GO:0032501)**
RPS6	PRKACB	SERBP1	PRKACB
RPS8	RPS6	RPS6	
	RPL29	RPL29	
	EIF5B	EIF5B	
	UCHL5	UCHL5	
	RPS8	RPS8	
	CDK1	CDK1	

Table 8. PANTHER GO – slim **Biological Process**.

-	-	-
Cell Part (GO:0044464)	**Macromolecular Complex (GO:0032991)**	**Organelle (GO:0043226)**
RPL23	RPL23	RPL23
RPS6	RPS6	RPS6
RPS20	RPS20	RPS20
RPL29	RPL29	RPL29
EIF5B	RPS8	RPS8
UCHL5		
RPS8		

Table 9. PANTHER GO – slim **Cellular Component**.

-	-	-
Hydrolase (PC00121)	**Nucleic Acid Binding (PC00171)**	**Transferase (PC00220)**
EIF5B	RPL23	CDK1
UCHL5	EIF5	
	RPS20	
	RPL29	
	SERBP1	
	EIF5B	

Table 10. PANTHER GO **Protein Class**.

Conclusion

We identified 13 candidate genes through the comparative analysis of the results obtained from the three microarray data analysis tools (MeV, GEO2R and WB-DEGS). The results that we presented substantiates our approach of meta-analysis of expression data and pathway analysis of the genes obtained to find the putative targets which have biological and functional significance in the progression of prostate cancer. We also performed the comparative analysis of the tools that we used to find the overlapping genes in all the three targeted DNA repair pathways. We propose few entity(s) based on quantitative data compilation, after investigating their association with the disease as well their overall association with the three pathways by connecting and visualizing them at pathway level. Identification of the connections of genes, gene hubs and their functions can be novel therapeutic targets after experimental verification. They are therefore, proposed as potent biomarkers for the Prostate Cancer and their involvement in AR, MAPK and mTOR DNA repair pathways.

References

1. Gene Ontology Consortium: going forward. (2015). *Nucleic Acids Research, 43*(D1), D1049-D1056. doi:10.1093/nar/gku1179
2. Arkun, Y. (2016). Dynamic Modeling and Analysis of the Cross-Talk between Insulin/AKT and MAPK/ERK Signaling Pathways. *PLoS One, 11*(3), e0149684. doi:10.1371/journal.pone.0149684
3. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nat Genet, 25*(1), 25-29. doi:10.1038/75556
4. Azevedo, A., Cunha, V., Teixeira, A. L., & Medeiros, R. (2011). IL-6/IL-6R as a potential key signaling pathway in prostate cancer development. *World J Clin Oncol, 2*(12), 384-396. doi:10.5306/wjco.v2.i12.384
5. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., . . . Soboleva, A. (2013). NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res, 41*(Database issue), D991-995. doi:10.1093/nar/gks1193
6. Best, C. J., Gillespie, J. W., Yi, Y., Chandramouli, G. V., Perlmutter, M. A., Gathright, Y., . . . Chuaqui, R. F. (2005). Molecular alterations in primary prostate cancer after androgen ablation therapy. *Clin Cancer Res, 11*(19 Pt 1), 6823-6834. doi:10.1158/1078-0432.ccr-05-0585
7. Bolstad, B. M. (2004). *Low-level analysis of high-density oligonucleotide array data: background, normalization and summarization.* University of California, Berkeley.
8. Bolstad, B. M., Collin, F., Brettschneider, J., Simpson, K., Cope, L., Irizarry, R. A., & Speed, T. P. (2005). Quality Assessment of Affymetrix GeneChip Data. In R. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry, & S. Dudoit (Eds.), *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (pp. 33-47). New York, NY: Springer New York.
9. Bourgon, R., Gentleman, R., & Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences, 107*(21), 9546-9551.
10. Brettschneider, J., Collin, F., Bolstad, B. M., & Speed, T. P. (2008). Quality assessment for short oligonucleotide microarray data. *Technometrics, 50*(3), 241-264.
11. Carver, Brett S., Chapinski, C., Wongvipat, J., Hieronymus, H., Chen, Y., Chandarlapaty,

S., . . . Sawyers, Charles L. Reciprocal Feedback Regulation of PI3K and Androgen Receptor Signaling in PTEN-Deficient Prostate Cancer. *Cancer Cell, 19*(5), 575-586. doi:10.1016/j.ccr.2011.04.008

12. Chalifa-Caspi, V., Shmueli, O., Benjamin-Rodrig, H., Rosen, N., Shmoish, M., Yanai, I., . . . Lancet, D. (2003). GeneAnnot: interfacing GeneCards with high-throughput gene expression compendia. *Brief Bioinform, 4*(4), 349-360.
13. Chen, H., & Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics, 12*(1), 35. doi:10.1186/1471-2105-12-35
14. Cojoc, M., Peitzsch, C., Kurth, I., Trautmann, F., Kunz-Schughart, L. A., Telegeev, G. D., . . . Dubrovska, A. (2015). Aldehyde Dehydrogenase Is Regulated by beta-Catenin/TCF and Promotes Radioresistance in Prostate Cancer Progenitor Cells. *Cancer Res, 75*(7), 1482-1494. doi:10.1158/0008-5472.can-14-1924
15. D'Antonio, J. M., Ma, C., Monzon, F. A., & Pflug, B. R. (2008). Longitudinal analysis of androgen deprivation of prostate cancer cells identifies pathways to androgen independence. *Prostate, 68*(7), 698-714. doi:10.1002/pros.20677
16. Davis, S., & Meltzer, P. S. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics, 23*(14), 1846-1847. doi:10.1093/bioinformatics/btm254
17. Edlind, M. P., & Hsieh, A. C. (2014). PI3K-AKT-mTOR signaling in prostate cancer progression and androgen deprivation therapy resistance. *Asian J Androl, 16*(3), 378-386. doi:10.4103/1008-682x.122876
18. Gautier, L., Cope, L., Bolstad, B. M., & Irizarry, R. A. (2004). affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics, 20*(3), 307-315. doi:10.1093/bioinformatics/btg405
19. Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., . . . Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol, 5*(10), R80. doi:10.1186/gb-2004-5-10-r80
20. Gioeli, D. (2005). Signal transduction in prostate cancer progression. *Clinical Science, 108*(4), 293. Retrieved from http://www.clinsci.org/content/108/4/293.abstract
21. Hagman, Z., Haflidadottir, B. S., Ceder, J. A., Larne, O., Bjartell, A., Lilja, H., . . . Ceder, Y. (2013). miR-205 negatively regulates the androgen receptor and is associated with adverse outcome of prostate cancer patients. *Br J Cancer, 108*(8), 1668-1676. doi:10.1038/bjc.2013.131

22. Heinlein, C. A., & Chang, C. (2004). Androgen receptor in prostate cancer. *Endocr Rev, 25*(2), 276-308. doi:10.1210/er.2002-0032
23. Hollenhorst, P. C., Ferris, M. W., Hull, M. A., Chae, H., Kim, S., & Graves, B. J. (2011). Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes Dev, 25*(20), 2147-2157. doi:10.1101/gad.17546311
24. Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res, 37*(1), 1-13. doi:10.1093/nar/gkn923
25. Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc, 4*(1), 44-57. doi:10.1038/nprot.2008.211
26. Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., . . . Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Meth, 12*(2), 115-121. doi:10.1038/nmeth.3252
27. Huggins, C., & Hodges, C. V. (2002). Studies on prostatic cancer. I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. 1941. *J Urol, 167*(2 Pt 2), 948-951; discussion 952.
28. Huggins, C., Stevens, R. E., Jr, & Hodges, C. V. (1941). Studies on prostatic cancer: Ii. the effects of castration on advanced carcinoma of the prostate gland. *Archives of Surgery, 43*(2), 209-223. doi:10.1001/archsurg.1941.01210140043004
29. Ihaka, R., & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of computational and graphical statistics, 5*(3), 299-314.
30. Junttila, M. R., Li, S. P., & Westermarck, J. (2008). Phosphatase-mediated crosstalk between MAPK signaling pathways in the regulation of cell survival. *Faseb j, 22*(4), 954-965. doi:10.1096/fj.06-7859rev
31. Kaarbo, M., Klokk, T. I., & Saatcioglu, F. (2007). Androgen signaling and its interactions with other signaling pathways in prostate cancer. *Bioessays, 29*(12), 1227-1238. doi:10.1002/bies.20676
32. Kahn, B., Collazo, J., & Kyprianou, N. (2014). Androgen receptor as a driver of therapeutic resistance in advanced prostate cancer. *Int J Biol Sci, 10*(6), 588-595. doi:10.7150/ijbs.8671
33. Kaissi, O., Nimpaye, E., Singh, T. R., Vannier, B., Ibrahimi, A., Ghacham, A. A., & Moussa, A. (2013). Genes selection comparative study in microarray data analysis. *Bioinformation, 9*(20), 1019-1022. doi:10.6026/97320630091019
34. Kaissi, O., Nimpaye, E., Singh, T. R., Vannier, B., Ibrahimi, A., Ghacham, A. A., &

Moussa, A. (2013). Genes selection comparative study in microarray data analysis. *Bioinformation, 9*(20), 1019-1022. doi:10.6026/97320630091019

35. Kaushik, A. K., Shojaie, A., Panzitt, K., Sonavane, R., Venghatakrishnan, H., Manikkam, M., . . . Sreekumar, A. (2016). Inhibition of the hexosamine biosynthetic pathway promotes castration-resistant prostate cancer. *Nature Communications, 7*, 11612. doi:10.1038/ncomms11612
36. Kogan-Sakin, I., Tabach, Y., Buganim, Y., Molchadsky, A., Solomon, H., Madar, S., . . . Rotter, V. (2011). Mutant p53(R175H) upregulates Twist1 expression and promotes epithelial-mesenchymal transition in immortalized prostate cells. *Cell Death Differ, 18*(2), 271-281. doi:10.1038/cdd.2010.94
37. Koivisto, P., Kolmer, M., Visakorpi, T., & Kallioniemi, O. P. (1998). Androgen receptor gene and hormonal therapy failure of prostate cancer. *Am J Pathol, 152*(1), 1-9.
38. Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y. A., Williams, E., . . . Brazma, A. (2015). ArrayExpress update—simplifying data submissions. *Nucleic Acids Research, 43*(D1), D1113-D1116. doi:10.1093/nar/gku1057
39. Laplante, M., & Sabatini, D. M. (2009). mTOR signaling at a glance. *J Cell Sci, 122*(Pt 20), 3589-3594. doi:10.1242/jcs.051011
40. Li, L., Ittmann, M. M., Ayala, G., Tsai, M. J., Amato, R. J., Wheeler, T. M., . . . Thompson, T. C. (2005). The emerging role of the PI3-K-Akt pathway in prostate cancer progression. *Prostate Cancer Prostatic Dis, 8*(2), 108-118. Retrieved from http://dx.doi.org/10.1038/sj.pcan.4500776
41. Liao, R. S., Ma, S., Miao, L., Li, R., Yin, Y., & Raj, G. V. (2013). Androgen receptor-mediated non-genomic regulation of prostate cancer cell proliferation. *Transl Androl Urol, 2*(3), 187-196. doi:10.3978/j.issn.2223-4683.2013.09.07
42. Liu, G., Loraine, A. E., Shigeta, R., Cline, M., Cheng, J., Valmeekam, V., . . . Siani-Rose, M. A. (2003). NetAffx: Affymetrix probesets and annotations. *Nucleic Acids Research, 31*(1), 82-86. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC165568/
43. Lonergan, P. E., & Tindall, D. J. (2011). Androgen receptor signaling in prostate cancer development and progression. *J Carcinog, 10*, 20. doi:10.4103/1477-3163.83937
44. Majumder, P. K., & Sellers, W. R. (2005). Akt-regulated pathways in prostate cancer. *Oncogene, 24*(50), 7465-7474. doi:10.1038/sj.onc.1209096
45. Maroni, P. D., Koul, S., Meacham, R. B., & Koul, H. K. (2004). Mitogen Activated Protein kinase signal transduction pathways in the prostate. *Cell Commun Signal, 2*(1), 5. doi:10.1186/1478-811x-2-5

46. Mendoza, M. C., Er, E. E., & Blenis, J. (2011). The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem Sci, 36*(6), 320-328. doi:10.1016/j.tibs.2011.03.006
47. Mi, H., Dong, Q., Muruganujan, A., Gaudet, P., Lewis, S., & Thomas, P. D. (2010). PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Research, 38*(suppl_1), D204-D210. doi:10.1093/nar/gkp1019
48. Morgan, T. M., Koreckij, T. D., & Corey, E. (2009). Targeted therapy for advanced prostate cancer: inhibition of the PI3K/Akt/mTOR pathway. *Curr Cancer Drug Targets, 9*(2), 237-249.
49. Mukherjee, R., McGuinness, D. H., McCall, P., Underwood, M. A., Seywright, M., Orange, C., & Edwards, J. (2011). Upregulation of MAPK pathway is associated with survival in castrate-resistant prostate cancer. *Br J Cancer, 104*(12), 1920-1928. doi:10.1038/bjc.2011.163
50. Mulholland, D. J., Tran, L. M., Li, Y., Cai, H., Morim, A., Wang, S., . . . Wu, H. (2011). Cell autonomous role of PTEN in regulating castration-resistant prostate cancer growth. *Cancer Cell, 19*(6), 792-804. doi:10.1016/j.ccr.2011.05.006
51. Pienta, K. J., & Bradley, D. (2006). Mechanisms underlying the development of androgen-independent prostate cancer. *Clin Cancer Res, 12*(6), 1665-1671. doi:10.1158/1078-0432.ccr-06-0067
52. Rajan, P., Dalgliesh, C., Carling, P. J., Buist, T., Zhang, C., Grellscheid, S. N., . . . Elliott, D. J. (2011). Identification of novel androgen-regulated pathways and mRNA isoforms through genome-wide exon-specific profiling of the LNCaP transcriptome. *PLoS One, 6*(12), e29088. doi:10.1371/journal.pone.0029088
53. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, gkv007.
54. Rodriguez-Berriguete, G., Fraile, B., Martinez-Onsurbe, P., Olmedilla, G., Paniagua, R., & Royuela, M. (2012). MAP Kinases and Prostate Cancer. *J Signal Transduct, 2012*, 169170. doi:10.1155/2012/169170
55. Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., . . . Quackenbush, J. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques, 34*(2), 374-378.
56. Scheid, S., & Spang, R. (2005). twilight; a Bioconductor package for estimating the local

false discovery rate. *Bioinformatics, 21*(12), 2921-2922. doi:10.1093/bioinformatics/bti436

57. Schwender, H., & Ickstadt, K. (2008). Empirical Bayes analysis of single nucleotide polymorphisms. *BMC Bioinformatics, 9*(1), 144. doi:10.1186/1471-2105-9-144
58. Sun, D., Lee, Y. S., Malhotra, A., Kim, H. K., Matecic, M., Evans, C., . . . Dutta, A. (2011). miR-99 family of MicroRNAs suppresses the expression of prostate-specific antigen and prostate cancer cell proliferation. *Cancer Res, 71*(4), 1313-1324. doi:10.1158/0008-5472.can-10-1031
59. Sun, Y., Wang, B. E., Leong, K. G., Yue, P., Li, L., Jhunjhunwala, S., . . . Johnson, L. (2012). Androgen deprivation causes epithelial-mesenchymal transition in the prostate: implications for androgen-deprivation therapy. *Cancer Res, 72*(2), 527-536. doi:10.1158/0008-5472.can-11-3004
60. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., . . . von Mering, C. (2015). STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Research, 43*(Database issue), D447-D452. doi:10.1093/nar/gku1003
61. Tabach, Y., Kogan-Sakin, I., Buganim, Y., Solomon, H., Goldfinger, N., Hovland, R., . . . Domany, E. (2011). Amplification of the 20q chromosomal arm occurs early in tumorigenic transformation and may initiate cancer. *PLoS One, 6*(1), e14632. doi:10.1371/journal.pone.0014632
62. Tan, M. H., Li, J., Xu, H. E., Melcher, K., & Yong, E. L. (2015). Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacol Sin, 36*(1), 3-23. doi:10.1038/aps.2014.18
63. Taylor, B. S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B. S., . . . Gerald, W. L. (2010). Integrative genomic profiling of human prostate cancer. *Cancer Cell, 18*(1), 11-22. doi:10.1016/j.ccr.2010.05.026
64. Terada, N., Shimizu, Y., Kamba, T., Inoue, T., Maeno, A., Kobayashi, T., . . . Ogawa, O. (2010). Identification of EP4 as a potential target for the treatment of castration-resistant prostate cancer using a novel xenograft model. *Cancer Res, 70*(4), 1606-1615. doi:10.1158/0008-5472.can-09-2984
65. Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., . . . Narechania, A. (2003). PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Research, 13*(9), 2129-2141. doi:10.1101/gr.772403
66. Thomas, P. D., Kejariwal, A., Guo, N., Mi, H., Campbell, M. J., Muruganujan, A., & Lazareva-Ulitsky, B. (2006). Applications for protein sequence–function evolution data:

mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Research, 34*(suppl_2), W645-W650. doi:10.1093/nar/gkl229

67. Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., . . . Morris, Q. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research, 38*(Web Server issue), W214-W220. doi:10.1093/nar/gkq537
68. Wu, Y., Chhipa, R. R., Cheng, J., Zhang, H., Mohler, J. L., & Ip, C. (2010). Androgen receptor-mTOR crosstalk is regulated by testosterone availability: implication for prostate cancer cell survival. *Anticancer Res, 30*(10), 3895-3901.
69. Yu, W., Feng, S., Dakhova, O., Creighton, C. J., Cai, Y., Wang, J., . . . Ittmann, M. (2011). FGFR-4 Arg(3)(8)(8) enhances prostate cancer progression via extracellular signal-related kinase and serum response factor signaling. *Clin Cancer Res, 17*(13), 4355-4366. doi:10.1158/1078-0432.ccr-10-2858