Original Research Article

Exploring the genetic differences between Primary and Metastatic Prostate Cancer using Bioinformatic Approaches: A Preliminary Study

Faris Aizat Ahmad Fajri[1], Muhammad Amru Nazri[1], Muhamad Harith Zulkifli[1], Fazlin Mohd Fauzi[1,2*]

[1]Faculty of Pharmacy, Universiti Teknologi MARA Selangor, Puncak Alam Campus, 42300 Bandar Puncak Alam, Selangor, Malaysia
[2]Collaborative Drug Discovery Research, Faculty of Pharmacy, Universiti Teknologi MARA Selangor, Puncak Alam Campus, 42300 Bandar Puncak Alam, Selangor, Malaysia

Abstract

Screening of prostate cancer (PCa) by measuring prostate cancer antigen has proven beneficial in reducing the mortality and progression of prostate cancer. However, its level can be affected if patients are taking certain drugs and/or suffering from certain medical conditions, causing a false negative. This can lead to PCa being undetected, where when untreated can lead to metastatic prostate cancer (MPC). Hence, in this study, genetic differences between PCa and MPC were explored using bioinformatics approaches to predict potential biomarkers for MPC. The study was divided into two parts, where the first involves feature selection and principal component analysis to differentiate PCa and MPC based on mRNA gene expression. Additionally, top 20 mutated genes for MPC were determined using odds ratio (OR). In the second phase, a predictive model was built using outcome of the mRNA gene expression analysis. The results showed that the mRNA expression of 26 identified genes could differentiate between PCa and MPC. This was further corroborated by the predictive model, where a sensitivity and specificity of 0.616 and 0.017 respectively was achieved. While importance is placed on sensitivity over specificity, further improvements involving more data need to be made to increase the specificity rate. Additionally, genes such as PAG24, BOP1 and GRWD1 should be investigated further as both potential biomarkers as well as potential pathways in MPC progression, based on further protein-protein interaction analysis. OR and protein-protein interaction suggests that androgen signalling pathway may crosstalk with NF-κB signalling and breast cancer pathway. This preliminary study shows that bioinformatics approaches could aid in understanding MPC, which could lead to the discovery of novel targeted therapy and potential biomarkers.

Keywords: prostate cancer, metastatic prostate cancer, Random forest, principal component analysis, feature selection

Corresponding author
Fazlin Mohd Fauzi
Level 11, FF1 Building, Faculty of Pharmacy,
UiTM Puncak Alam, Bandar Puncak Alam, 42300,
Selangor, Malaysia.
fazlin5465@uitm.edu.my

Received 20 Dec 2021; accepted 17 May 2022
Available online: 5 July 2022
https://doi.org/10.24191/IJPNaCS.v5i1.04

1.0 INTRODUCTION

According to the GLOBOCAN 2020 report by the World Health Organisation (WHO), prostate cancer ranked third with 7.3% in the number of new cancer-related cases in 2020 (1). Additionally, 3.8% of cancer-related deaths were attributed to prostate cancer (1). According to the Prostate Cancer Foundation, prostate cancer is the most common non-skin cancer in America affecting 1 in 8 males. In Asian countries, the number of prostate cancer cases has been steadily increasing throughout the years where initially the incidence was low (2).

Androgens, which can either be testosterone or dihydrotestosterone (DHT) with the latter being the more abundant, are hormones that are responsible for the growth and function of the prostate gland[3]. In prostate cancer, the dysregulation of androgen signalling pathway leads to overproduction of androgens and/or overstimulation of androgen receptor, leading to the growth of prostate cancer cells (3). Hence, the treatment of prostate cancer involves androgen deprivation therapy (ADT), which suppresses the production of androgens or inhibiting androgen receptor (AR) (4). Prostate cancer is detected through blood test that measures the prostate-specific antigen (PSA) level where a high level leads to the diagnosis of prostate cancer. However, PSA test suffers from inaccuracies as its level can be affected by drugs and conditions such as prostatitis and benign prostate hyperplasia. This can lead to a missed diagnosis where if not addressed promptly could lead to metastatic prostate cancer (MPC), which has a low survival rate (4). Detection of MPC through imaging such as computed tomography (CT), positron-emission tomography (PET) and magnetic resonance imaging (MRI) is incomplete, further complicating the diagnosis of MPC[4]. Similar to primary prostate cancer, MPC is also treated with ADT, however, several studies have demonstrated that patients on long term ADT are at higher risk of stroke and vulnerable to cardiovascular adverse effects (5,6). These highlights the need to understand the genetics of MPC in the discovery of novel drugs as well as for diagnostic purposes.

The characteristics of MPC were largely unknown until 150 metastatic biopsies were analysed through an international, multi-institutional study. The study unveiled a defect in DNA repair mechanism in MPC where mutations in DNA repair genes e.g. BRCA2, ATM and BRCA1 were observed in 23% of the cases (7). These results were further corroborated by Pritchard et al., (8) with a larger cohort of 692 men. 11.8% of the cases exhibited germline mutation I DNA repair genes e.g. BRCA2, ATM, CHEK2, and BRCA. Furthermore, the mutations were not correlated with age or family history of prostate cancer (8). Several other studies have discovered potential genetic alterations in MPC e.g. TP53, PTEN and AR (7), however, therapy targeting those genes have not yet been shown to be clinically beneficial.

Several studies involving the use of computational approaches have been employed to identify genes that are altered in MPC. Li et al., (9) employed the maximum relevance minimum redundancy (mRMR) method to discover surrogate genes for MPC by analysing microarray data of normal, primary prostate cancer and MPC tumours. The study identified four genes that could differentiate the three different phases, which are TUBB6, MYEF2, PARM1 and SLC25A22 (9). These genes are involved in cell communication, hormone-receptor mediated signaling, and transcription regulation, which may be responsible for the development of prostate cancer. Xue et al., (10) performed an integrative analysis of transcription factor (TF) and microRNA expression profiles by employing Gaussian mixture modelling and network pruning. The study identified mutually exclusive transcriptional drivers, AR, HOXC6 and NKX2-2 (10). These gene together

dysregulate metastasis-related miRNAs in prostate cancer. Additionally, poor clinical outcome have been reported from the overexpression of TFs (10). Bello et al., (11) applied system-based modelling approach known as kinome regulatisation (KiR), which identified multitargeted kinase inhibitors that suppress castration-resistant prostate cancer (CRPC). The two inhibitors identified, PP121 and SC-1 were later found to suppress the growth of CRPC *in vitro* and *in vivo* (11). Hence, the aim of this study is to explore the genetic differences between MPC and PCa, and consequently predict potential biomarkers for MPC through bioinformatic approaches using data obtained from public databases.

2.0 Materials and Method

2.1 Design of the study

The design of the study is represented in Figure 1 below. The study is divided into two phases where the first involves the mining of mRNA expression of primary and metastatic prostate cancer (PCa and MPC respectively) data obtained from cBioPortal (12) through principal component analysis (PCA) and feature selection. Odds ratio was also conducted to analyse significantly mutated genes in MPC. The second phase of the study involves the building of prediction model based on the results of the PCA to validate whether mRNA gene expression profile can be used to differentiate between PCa and MPC.

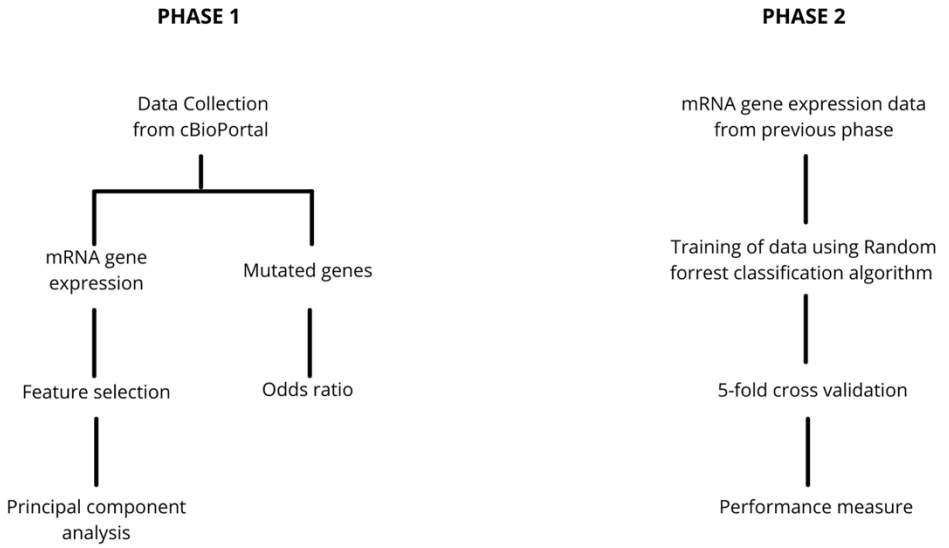

Figure 1. Design of the study where it is divided into two phases. The first phase involves the use of Principle Component Analysis on mRNA gene expression data, and Odds Ratio on mutated genes data. The second phase involves the building of a prediction model based on the result of mRNA gene expression from the first phase.

2.2 Dataset

In this study, the data of prostate cancer patients was obtained from the cBioPortal database (https://www.cbioportal.org/). cBioPortal is an international public database that store and distribute functional genomic data that was summited by research community. In this study, four datasets of prostate cancer patient were chosen, which are the DKFZ cancer cell 2018, SU2C/PCF Dream Team PNAS 2019, MSKCC/DFCI Nature Genetics 2018, and lastly from the MSKCC JCO Precis Oncol 2017. The breakdown of each datasets can be found in Table 1. Any duplicates were removed.

2.3 Feature selection

In this study, the Tree Based Feature Selection Method (TBSM) was employed as the feature selection method (17). To remove irrelevant or unimportant data, this method measures the impurity-based feature importance of each variable by using the concept of random forest algorithm. The degree of importance is based on how many samples are able to reach nodes against the total number of samples (17). The higher degree or percentage of feature importance, the higher the score. The nodes importance was calculated as such:

$$ni_1 = w_1 C_1 - w_{left(1)} C_{left(1)} - w_{right(1)} C_{right(1)} \quad \text{Eq. 1}$$

where:

ni_1 = the nodes importance of node 1
w_1 = the weighted sample reaching node 1
C_1 indicate the impurity value of node 1.
$left(1)$ and $right(1)$ = branches node in the left and the right respectively.

The importance of each feature on a decision tree is then calculated as:

$$fi_1 = \frac{\sum_{1:node\ 1\ splits\ on\ feature\ i} ni_1}{\sum_{k \in all\ nodes} ni_k} \quad \text{Eq. 2}$$

where:

fi_1 = importance of feature *i*
ni_1 = the importance of node 1.
Ni_k = the sum importance of all nodes

Next, the value of feature importance is normalized to a value between 0 and 1, calculated as such:

$$normfi_1 = \frac{fi_1}{\sum_{j \in all\ features} fi_1} \quad \text{Eq. 3}$$

$normfi_{i1}$ represents the normalized feature importance for *i* in tree 1. Next, the average of all the trees which is the final feature importance will be calculated. It is calculated by sum of the feature's importance value on each tree and divided by the total number of trees (17):

$$RFfi_i \frac{\sum_{j \in all\ tress} normfi_{i1}}{T} \quad \text{Eq. 4}$$

$RFfi_i$ is the importance of feature *i* calculated from all trees in the Random Forest model and *T* is the total number of trees.

Each feature will be assigned a value between 0 to 1 where a higher value indicates higher importance. The relative importance of a feature was calculated by comparing its value to the highest scoring feature as such (17):

$$Relative\ feature\ importance_{\ i} = \frac{RFfi_i}{RFfi_{max}} \times 100 \quad \text{Eq. 5}$$

Feature with the highest score will be assigned a value of 100%. Only features with a Relative Feature Importance score of 30 and above were retained for further analysis.

Table 1. Dataset of MPC and PCa used in the study which includes their origin, data type and amount of data.

Dataset	Number of samples	Type of data used	References
DKFZ cancer cell 2018	324	• Mutated genes • mRNA gene expression	13
SU2C/PCF Dream Team PNAS 2019	444	• Mutated genes • mRNA gene expression	14
MSKCC/DFCI Nature Genetics 2018	1013	• Mutated genes	15
MSKCC JCO Precis Oncol 2017	504	• Mutated gene	16

2.4 Principal Component Analysis

Principal component analysis (PCA) is a method of reducing the dimensionality of robust datasets, increasing its interpretability while preserving as much variability and minimizing information loss (18). This statistical technique creates new uncorrelated variables or principal components, that successively maximize variance. The PCA was performed using the scikit-learn package through the 'PCA' function in Python and plotted using the *ggplot* (19) package in RStudio (v1.4).

Given a data matrix, $\boldsymbol{X}$, of $n \times p$, where n is the number of rows of instances and p is the number of features, the principal component for each variable, $\boldsymbol{x}$, is calculated as the weighted average of the original variables. The matrix containing the principal components of the data is referred to as matrix Y and can thus be calculated as:

$$Y = W.\ X \qquad \text{Eq. 6}$$

where $\boldsymbol{W}$ is a matrix of coefficients that is obtained from the calculation of covariance, eigenvalues and eigenvector. Eigenvalues and eigenvectors are the linear algebra concepts that needed to be computed from the covariance matrix in order to determine the principal components of the data (20) :

$$y_{ij} = w_{1i}\, x_{1j} + w_{2i}\, x_{2j} + \ldots + w_{pi}\, x_{pj} \qquad \text{Eq. 7}$$

The covariance between two variables, x_i and x_j can be calculated as:

$$Cov(xi, xj) = \frac{1}{n-1} \sum_{i=1}^{n} (xi - \bar{xi})(xj - \bar{xj}) \qquad \text{Eq. 8}$$

The eigenvalues and eigenvectors are then determined from the covariance matrix. The eigenvectors (principal components) determine the directions of the new feature space, and the eigenvalues determine their magnitude.

2.5 Odds ratio

Odds ratio (OR) measures the association between exposure and the outcome by comparing the odd of the outcome occurring depending on the presence or absence of certain exposure

(21). In this study, the mutated genes were represented as the exposure. Meanwhile, the outcome was either primary or metastatic prostate cancer. The odds ratio will then measure the frequency of the mutated genes in metastatic prostate cancer and primary prostate cancer. Out of 2485 mutated genes, only the top 20 highest odds ratio of mutated genes with a *p*-value ≤ 0.05 were retained. OR was calculated as such:

$$odd\ ratios = \frac{a/c}{b/d} \qquad \text{Eq. 9}$$

where:

a = frequency of mutation in metastatic prostate cancer
c = total number of mutations in metastatic prostate cancer
b = frequency of mutation in primary prostate cancer
d = total number of mutations in primary prostate cancer

2.6 Protein-Protein Interaction prediction using STRING

Protein-Protein Interaction (PPI) prediction using STRING (https://string-db.org/) was employed to see whether two proteins may interact. STRING measures both direct (physical) and indirect (functional) interactions between two proteins, based on experimental data of protein-protein interactions (22).

A score is provided for each protein-protein association. The scores represent confidence scores, ranging from 0 to 1, indicating estimated likelihood that the association is biologically significant, given the supporting evidence (22). The supporting evidence is based on seven factors, which are neighbourhood in genome, gene fusions, co-occurrence across genomes, co-expression, experimental/biochemical data, association in curated databases and co-mentioned in PubMed abstracts (22). These factors are represented by colour coded edges. Based on the seven factors, a combined and final confidence score is computed. A good interaction should not only have a high combined score, but also have more than one factor contributing to the score.

Predictive model

2.7.1 Training set

The training set here contains the mRNA expression of MPC and PCa patients containing 26 genes identified in the previous phase.

2.7.2 Random Forest classification algorithm

Random forest is a technique for classification based on an ensemble, or forest, of decision tree. As the name suggests, a prediction will be made using tree-based algorithm method by constructing a forest from the production of several or large number of trees (known as decision trees) (23). The trees were built using training sets consisting of multiple features or variables for each of the instances in the training set. Then, output results were produced from the variables of the training set of interest. The result was obtained by aggregating all the outputs from different trees. There are two stages in Random Forest which are: (i) random forest creation and (ii) prediction from the random forest classifier created in the first stage (23).

Firstly, the algorithm will build *m* amount of decision trees. Each of the decision trees will be initiated with a single node where a number of randomly selected samples will serve as the data set. Then, a bootstrap sample of *n* number of variables of the training data was drawn and selected at random. From the random selected subset, the variable that provides the best split, measured using the Gini index, will split the node into two daughter nodes, specifying possible outcomes (23). The tree

was further split until a maximum size is reached without pruning. Gini index (S) is calculated as follows:

$$\text{Gini (S)} = 1 - \sum_j^2 P \qquad \text{Eq. 10}$$

Where *P* is the relative frequency of class *j* in S. Each time, the split then was divided into two subsets of S_1 and S_2 in which gini (S) data was divided into:

$$\text{Gini}_{\text{split}}\text{ (S)} = \frac{n1}{n}\text{ gini (S}_1\text{)} + \frac{n2}{n}\text{ gini (S}_2\text{)} \qquad \text{Eq. 11}$$

This process will repeat until the tree has reached a specified number of branches and is assigned a terminal leaf node. At the end of the tree, class probability will be calculated. In this study, *m* was set at 100, and *n* was set as the square root of total number of variables. The outcome was calculated as the mean of class probability from each decision trees. The algorithm was written in Python and using the *scikit-learn* package.

2.7.3 Internal validation

5-fold cross validation was used as internal validation. The data was separated into five different groups called fold. One of the folds will be chosen to represent the test set, while the rest were combined to serve as training set. Next, the predictive model will be fitted into the training set, tested on the test set and its performance will be calculated. This step will be repeated until all five folds have served as the test set.

2.7.4 Performance measure

The predictive model built by random forest algorithm was evaluated based on its specificity and sensitivity. Sensitivity evaluates the ability of the predictive model to predict true positive values. Meanwhile, specificity measures the ability of the predictive model to predict the true negative value. The formula to calculate both sensitivity and specificity as follows:

$$Sensitivity = \frac{true\ positves}{true\ positives + false\ negatives} \qquad \text{Eq. 12}$$

$$Specificity = \frac{true\ negatives}{true\ negatives + false\ positives} \qquad \text{Eq. 13}$$

3.0 Results

3.1 PCA profile of mRNA expression of PCa and MPC

The mRNA variables were reduced from 16,384 to 26 using feature selection to reduce overfitting, complexity and the curse of dimensionality. Table 3 shows the summary of the 26 genes used in the PCA. The data were then subjected to PCA, where PC1 and PC2 were plotted (see Figure 2) as it contains the most information.

From Figure 2 and Table 2, several observations can be made. Firstly, there is a clear separation between MPC and PCa from the PCA plot. This suggests that MPC and PCa could be differentiated by looking at their mRNA expression of the 26 genes collectively. Secondly, several genes listed in Table 2 are differentially expressed in certain malignancies. One of them, BOP1, can be linked to PCa. BOP1 is one of the important components for synthesis of the 60S ribosome and maturation of 5.8S and 20S ribosomal RNAs. Mutation and increase of BOP1 expression was demonstrated to lead to aggressive prostate cancer and reduction in patient overall survival (30). Another gene, FAM47E promotes the histone methylation by localizing arginine methyltransferase PRMT5 to chromatin. To date, literature support that links it to cancer is currently limited. However, a member of its family, FAM13C has been shown to be potential

prognostic marker in prostate cancer (37). Several of the genes are involved in other hormone-related cancers such as TRAPPC9, PABPC3, PA2G4 and NDUFA11, which were linked to breast cancer. Thirdly, several of the genes are directly or indirectly linked to NF-κB signalling pathway, which is involved in inflammation, immunity, cell proliferation, differentiation and apoptosis. These genes include EPN1, TRAPPC9 and RBM23.

Table 2: The details of the 26 genes identified through feature selection to construct the PCA between PCa and MPC. RFI refers to relative feature importance where a higher value indicates a higher importance. A value of 100 indicates that the gene is the most important in the group as it had the highest raw feature importance.

Gene	Gene name	RFI	Gene description
PCDHGA7	Protocadherin Gamma-A7	100.0	PCDHGA7 is a neural cadherin-like cell adhesion protein that play a role in specific cell-cell connections in the brain. Down regulation of PCDHGA7 gene was expressed in patients with colorectal cancer and other members of the PCDH families have been found to suppress tumours in certain malignancies where they undergo long-range epigenetic silencing by hypermethylation (24).
EPN1	Epsin-1	89.21	Epsins are ubiquitin-binding adaptor proteins where its overexpression leads to sustained NF-κB signalling, where in breast cancer leads to metastasis and epithelial mesenchymal transition (EMT) (25). Tumour growth and progression are reduced in cases of loss of function of this gene in certain malignancies (26).
NBPF10	Neuroblastoma Breakpoint Family Member 10	82.80	NBPF10 gene is a member of the neuroblastoma breakpoint family (NBPF). Altered expression of some gene family members is associated with several types of cancer, although its role is not fully understood.
DROSHA	Drosha Ribonuclease III	69.70	DROSHA plays an important role as a catalyst for the initial processing step of microRNA (miRNA) synthesis. Somatic mutations of DROSHA have been observed in human patients with kidney cancer where it impairs the expression of tumour suppressing miRNAs such as MYCN, LIN28 and other oncogenes (27).
PCDHGA11	Protocadherin Gamma-A11	65.69	PCDHGA11 is a neural cadherin-like cell adhesion protein that play a role in specific cell-cell connections in the brain. Members of the PCDH families have been found to suppress tumours in certain malignancies where they undergo long-range epigenetic silencing by hypermethylation (24).
PPM1J	Protein Phosphatase 1J	60.56	PPM1J gene plays a role in the catalytic activity to release phosphate from O-phospho-L-seryl-(protein). The function of this gene is not yet fully understood.
SFT2D3	SFT2 Domain-Containing Protein 3	52.20	This gene is involved in the fusion mechanism of transport vesicles that forms from the endocytic compartment with the Golgi complex. The function of this gene is not yet fully understood.
TRAPPC9	Trafficking Protein Particle Complex Subunit 9	50.31	This protein plays a role in the transportation of intra-Golgi and tethering of Golgi vesicle and also the activation of NF-κB signalling pathway. Mutation of this gene has been reported in colon and breast cancer (28).
ZC3H14	Zinc Finger CCCH-Type Containing 14	43.64	ZC3H14 is a gene that is encoded for a poly(A)-binding protein call the Zinc finger CCCH domain-containing protein 14. This protein plays a role in the control of the poly(A) tail length, mRNA stability, nuclear export, and translation. The role of ZC3H14 in cancer is not yet established but a member of its family, ZNF711 are closely associated with ER and HER2 expression. This suggests that ZNF711 is a predictor of poor prognosis in breast cancer (29).
BOP1	BOP1 Ribosomal Biogenesis Factor	43.20	BOP1 is a gene that is encoded for BOP1 ribosome biogenesis protein. This protein is one of the important components for synthesis of the 60S

			ribosome and maturation of 5.8S and 20S ribosomal RNAs. A recent study has found that mutation of the BOP1 gene that causes an increase in the BOP1 expression led to aggressive prostate cancer and reduction in patient overall survival (30).
DYX1C1	Dynein Axonemal Assembly Factor 4	41.56	This gene plays a role in the neuronal migration during the development of cerebral neocortex. Genomic alterations of DNAH family members have been reported in certain malignancies (31).
CXorf38	Chromosome X Open Reading Frame 38 protein	41.13	The function of this gene is not yet fully understood.
IFT122	Intraflagellar transport protein 122	38.41	IFT122 encodes for a member of the WD repeat protein family, which is involved in apoptosis, cell cycle progression, gene regulation and signal transduction. It is unknown if this gene is involved in cancer pathogenesis.
ITM2B	Integral membrane protein 2B	35.33	This protein plays a role in the processing of amyloid-beta A4 precursor protein. It helps to inhibit the amyloid-beta peptide aggregation and fibrils deposition. The inhibition of ITM2B transcription has been found to lead to the activation of PI3K/Akt signalling pathway, which accelerates tumour growth and worsens the prognosis of lung cancer in mice (32).
MFSD1	Major facilitator superfamily domain-containing protein 1	33.98	MFSD1 gene is encoded for the Major facilitator superfamily domain-containing protein 1. No information linking this gene to PCa or MPC has been found.
PABPC3	Polyadenylate-binding protein 3	33.61	This gene plays a role in the stability and initial translation of mRNA. PABPC3 expression have been associated with breast cancer in North African population (33).
PA2G4	Proliferation-associated protein 2G4	33.48	This gene plays important role in the ERBB3-regulated signal transduction pathway. The ERBB3 is also known as the HERS3 (human epidermal growth factor receptor 3). The protein is able to bind and interact with the ERBB3 receptor that causes transduction in the regulatory signal. Mutation of this gene has been found to have association with breast cancer (34). In addition, this gene also plays a role either as tumour suppressor or as an oncogene (35).
PFKL	ATP-dependent 6-phosphofructokinase	33.15	PFKL helps to catalyse the glycolysis metabolism process by converting of D-fructose 6-phosphate to D-fructose 1,6-bisphosphate. The degradation of PFKL leads to decreased glycolysis, which proliferation and metastasis of hepatocellular carcinoma (HCC) cells (36).
FAM47E	Family With Sequence Similarity 47 Member E	32.77	FAM47E promotes the histone methylation by localizing arginine methyltransferase PRMT5 to chromatin. Its role in cancer is not yet known. However, a member of its family, FAM13C have been shown to be potential prognostic marker in prostate cancer (37).
NDUFA11	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 11	32.75	This protein is a subunit of membrane-bound mitochondrial complex I. It plays a role in the mitochondrial electron transport chain. Silencing of NDUFA11 was found to increase oxygen consumption rate of breast cancer cells, as well downregulate expression of IL-6, IL-8, CXCL1, and CXCL3 (38). These lead to tumour metastasis and macrophage infiltration.
RBM23	Probable RNA-binding protein 23	31.90	RBM23 is a gene encoded for the Probable RNA-binding protein 23 which is a part of the U2AF-like family of RNA binding proteins. The RNA binding protein can act as the pre-mRNA splicing factor and as well as a transcription coactivator. In HCC, RBM23was found to promote the angiogenesis via the NF-κB signaling pathway (39).
WBSCR22	BUD23 rRNA methyltransferase and ribosome maturation factor	31.23	This gene has many roles such as it involves in the pre-rRNA processing steps to form small-subunit rRNA, biogenesis end export of the 40S ribosomal subunit, as steroid receptor coactivator, as maintenance of open chromatin and lastly as maintenance of demethylation on histone. Its role in cancer in unknown.
SCAF11	SR-Related CTD Associated Factor 11	30.86	The role of SCAF11 is unclear. However, a member of its family, SCAF1 is involved in pre-mRNA splicing and interacts with RNA polymerase II polypeptide A, specifically at the CTD domain. Overexpression of SCAF1 has been found in breast and ovarian tumours [40].
CENPI	Centromere protein I	30.76	CENPI is a gene that encodes centromere protein I, which is a part of the component of the CENPA-NAC (nucleosome-associated) complex. The complex is crucial in chromosome segregation and alignment, ensuring proper mitotic process. CENPI is overexpressed in colorectal cancer as it regulates cell invasion and migration (41).

MAP2K2	Mitogen-Activated Protein Kinase Kinase 2	30.75	MAP2K2 is a part of the MAP kinase kinase family and plays a role in the mitogen growth factor signal transduction. Mutation of MEK2 has been found to be associated with cancer and drug that limits MAP2K2 has been developed to treat cancer patient (42)
NCAPG2	Condensin-2 complex subunit G2	30.36	NCAPG2 is involved in cell proliferation by regulating the G2/M phase. Its overexpression has been reported in Non-Small Cell Lung Carcinoma, leading to tumour cell growth (43).

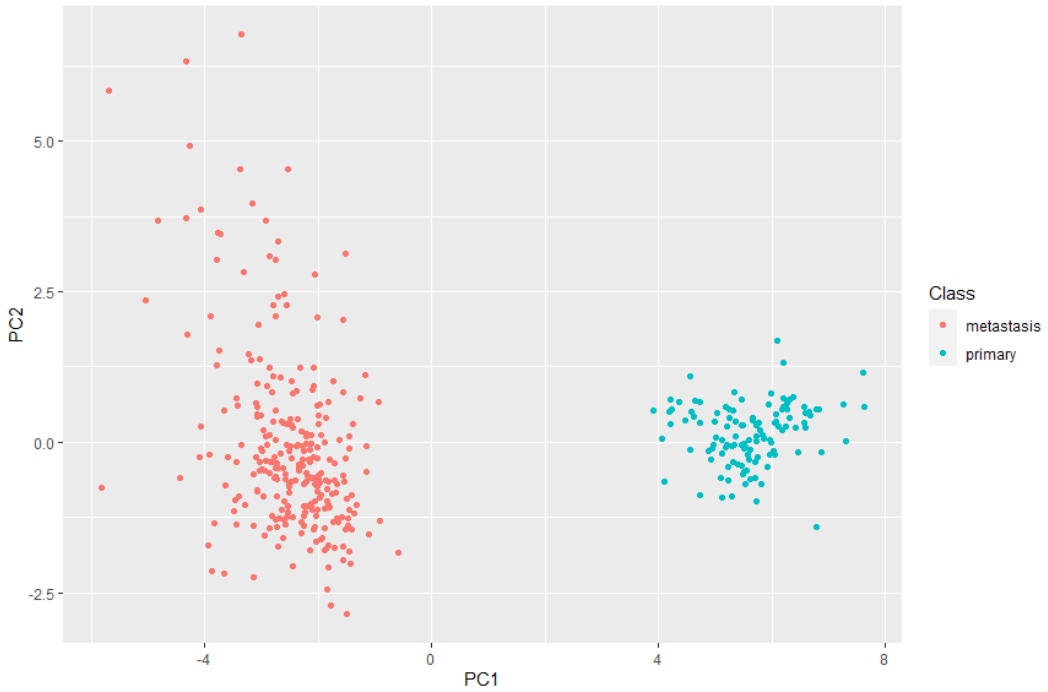

Figure 2: The PCA plot of MPC (labelled metastasis) and PCa (labelled primary) based on mRNA gene expression of selected 26 genes.

3.2 Odds ratio profile of gene mutation of MPC

Table 3 shows the top 20 significantly mutated genes in MPC, compared to PCa. A high odds ratio indicates that the gene mutation is more prominent in MPC than PCa. Several observations can be made from the OR of MPC. Firstly, AR was the second highest significantly mutated gene in MPC. Several studies have demonstrated that gain-of-function mutations and gene amplification of AR take place in adapting to the low androgen level (44). Additionally, AR co-activators such as TRIM24 are also upregulated, where collectively these events may restore the AR signalling pathway after ADT treatment and hence leading to MPC (44).

Secondly, several genes associated with cancers are also listed in Table 3 such as IGSF8, NKX2-5, GLUD2, GRWDI, TRIM32 and TSPYL2. However, the involvement of these genes in PCa or MPC is not yet known. Lastly, similar to the previous section, several of the genes can be found to be directly or indirectly linked to the NF-κB signalling pathway such as CD74 and TRIM40.

Table 3: Details of the top 20 mutated genes of MPC.

Gene	Gene Name	Odd radio	Gene description
ZDHHC20P1	zinc finger DHHC-type containing 20 pseudogene 1	51.82	The function of this gene is not known
AR	Androgen receptor	36.31	Several studies have demonstrated that gain-of-function mutations and gene amplification of AR take place in adapting to the low androgen level. Additionally, AR co-activators such as TRIM24 are also upregulated, where collectively these events may restore the AR signalling pathway after ADT treatment and hence leading to MPC (44).
FBXO24	F-box only protein 24	34.01	This gene is a part of the F-box protein member family that function in phosphorylation-dependent ubiquitination. FBXO24, by mediating ubiquitin-dependent proteasomal degradation, is involved in the regulation of cell proliferation (45). Its role in cancer pathogenesis is not clear.
HIST1H3PS1	H3 Clustered Histone 9, Pseudogene	30.77	HIST1H3PS1is a pseudogene where its role is unclear.
CD74	CD74 Molecule	24.29	CD74 is a gene encoded for the HLA class II histocompatibility antigen gamma chain. This protein plays important role in the MHC class II antigen process by acting as the binding site for cytokine migration inhibitory factor (MIF). CD74 was found to be associated with Mucinous Lung Adenocarcinoma, and related to NF-κB Signaling and Innate Immune System pathways (46).
CEL	Carboxyl ester lipase protein	24.29	This protein plays important role in the absorption and hydrolysis of the cholesterol and lipid-soluble vitamin ester. Recent studies have found that mutation of this gene in pancreatic disease (47).
IGSF8	Immunoglobulin superfamily member 8 protein	24.29	This protein plays many roles such as to regulate proliferation and differentiation of keratinocytes, cell motility, and the neurite outgrowth and maintenance of the neural network. IGSF8 may negatively regulate TGF-β signaling which can lead to invasion and metastasis of cancer cells (48).
MAMDC4	Apical endosomal glycoprotein	24.29	This protein plays a role in the managing the receptors and ligand selective transport on polarised epithelial. Its role in cancer is unknown.
MICF	MHC Class I Polypeptide-Related Sequence F (Pseudogene)	24.29	MICF is a pseudogene and its function is not known.
NKX2-5	Homeobox protein Nkx-2.5	24.29	This protein plays an important function in the heart and spleen development and few studies have found that mutation of this gene is associated with heart disease. NKX2.5 has been found to be expressed in several malignancies such as ovarian yolk sac, papillary thyroid carcinoma, skin squamous cell carcinoma tumor and pediatric acute lymphoblastic leukemia (49).
GLUD2	Glutamate dehydrogenase 2	22.67	It plays important role in the recycling of the glutamate neurotransmitter. A study has found that mutation of the gene has been expressed in cancer patients and other human disorders (50).
S1PR3	Sphingosine 1-phosphate receptor 3	22.67	This protein might play a role in the cell proliferation and help in the suppression of apoptosis.
KRTAP13-3	keratin-associated protein 13-3	21.05	KRTAP13-3 can be found in the hair cortex, forming a rigid and resistant hair shaft. The role of KRTAP13-3 in cancer is unknown. However, a member of its family, KRTAP13-2 was found to be significantly overexpressed in prostate cancer through bioinformatics approaches (51).
GRWD1	Glutamate-rich WD repeat-containing protein 1	21.05	It plays a role in the ribosome biogenesis and histone methylation. A recent study has found that overexpression of this gene increases the risk of oncogenesis (52).
RP11-386P4.1	Antisense RP11-386P4.1	21.05	RP11-386P4.1 is an antisense gene. No further information is currently available

TRIM32	Tripartite Motif containing 32	21.05	This protein is a member of the tripartite Motif (TRIM) family that plays many roles that include differentiation, muscle physiology and regeneration, and tumour suppression. A study has found out the mutation of this gene has an association with hepatocarcinogenesis (53).
TSPYL2	Testis-specific Y-encoded-like protein 2	21.05	This gene plays a role in modulating the gene expression and inhibiting cell proliferation. In addition, a study found that the mutation of this gene is associated with oncogenesis by acting as a proto-oncogene and a tumour suppressor gene (54).
TRIM40	Tripartite Motif Containing 40	20.64	TRIM40 is a member of the TRIM family. This protein plays important role in the innate response. TRIM40 was found to inhibit NF-κB activity via neddylation of IKKγ, which prevents inflammation-associated carcinogenesis in the gastrointestinal tract (55).
ACAD8	Acyl-CoA Dehydrogenase Family Member 8	19.43	ACAD8 plays a role in catalysing the metabolism of dehydrogenation of acyl-CoA derivative.
C16orf71	Dynein Axonemal Assembly Factor 8 protein	19.43	This protein is required for the deployment of outer dynein arm to axoneme in ciliated cells. Its role in cancer is not known.

3.3 PPI predictions of all genes identified

The result of the PPI can be seen in Figure 3 when all 26 genes from the mRNA expression and 20 top mutated genes were analysed. Two main interactions can be seen from Figure 3 where the first involved AR, PAG24, BOP1 and GRWD1. AR is connected to PAG24 whereas potential interaction exists between PAG24, BOP1 and GRWD1. PAG24 is a corepressor of AR and regulated by ERBB3 ligand neuregulin-1/heregulin (HRG). Over 300 coregulator of AR have been identified and they can either be a co-activator or co-repressor of AR. The coregulator can modify AR enzymatically and other components such as transcriptional proteins, histones or other coregulators. These can lead to the initiation of cellular processes such as invasion and proliferation, which drive tumour progression. BOP1 and GRWD1 genes have a similar function where they play a role in ribosomal biosynthesis. Vellky et al., (30) studied the expression of BOP1 in different stages of PCa and found that it is overexpressed in MPC and recurrent PCa. Additionally, the expression was inversely correlated with overall survival. Knockdown of BOP1 showed a decrease in proliferation and motility. The knockdown of GRWD1 also inhibits cell proliferation, invasion and migration, and induced cell cycle arrest but in colon carcinoma (56).

The second interaction from Figure 3 involves NBPF10, PABPC3 and ZC3H14. Both PABPC3 and ZC3H14 have similar function, which is to control and maintain the stability of mRNA strand. PABPC3 expression has been associated with breast cancer in North African population (57). The role of ZC3H14 in cancer is not yet established but a member of its family, ZNF711 has been shown to be closely associated with ER and HER2 expression. This suggests that ZNF711 is a predictor of poor prognosis in breast cancer (29). NBPF10 is a member of the neuroblastoma breakpoint family (NBPF). Altered expression of NBPF family members has been associated with several types of cancer, although its role is not fully understood.

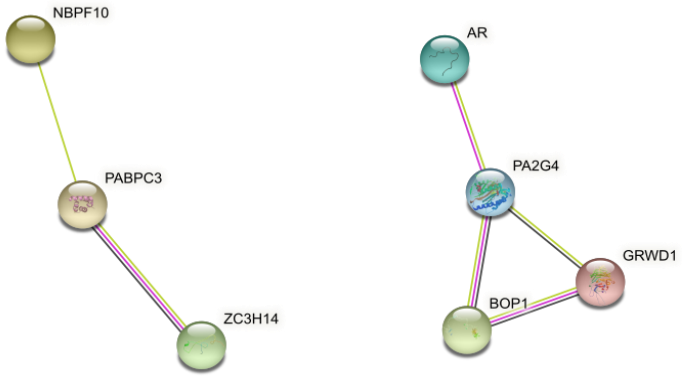

Figure 3: Protein-Protein Interaction of genes identified in this study. For purpose of clarity, only genes that were connected to another gene were shown here. Abbreviations: NBPF10: Neuroblastoma Breakpoint Family Member 10; PABPC3: Polyadenylate-binding protein 3; ZC3H14: Zinc Finger CCCH-Type Containing 14; AR: Androgen Receptor; PA2G4: Proliferation-associated protein 2G4; BOP1: BOP1 Ribosomal Biogenesis Factor; GRWD1: Glutamate-rich WD repeat-containing protein 1

3.4 Predictive model based on mRNA gene expression

Table 4 shows the internal validation of the predictive model built using the mRNA expression of the 26 genes previously mentioned between PCa and MPC patients. The model showed very low specificity (0.017) and good sensitivity (0.616). The predictive model generated 162 true positive (TP) results which means that 162 MPC patients were correctly identified. The predictive model was also only able to correctly identify 2 PCa patients which is the true negative result (TN). While importance is placed on sensitivity over specificity, further improvements involving more data needs to be made to increase the specificity rate.

3.5 Decision tree of mRNA gene expression

Figure 4 shows an example of a single decision tree in a random forest. Note that this is only an example of single decision tree, and a random forest contains hundreds of predictive trees (this is set at 100 in the current model). Here, the gene BOP1 is at the root node (uppermost node) of the decision tree, which is the most important feature for that decision tree. Gini value indicates probability of misclassifying an instance and a lower value indicates a better split. Value indicates the number of data sampled at particular node. This predictive tree differentiates between MPC and PCa based on the BOP1 expression. If the z-score of BOP1 is equal or less than 6.84, it will be classified as primary prostate cancer. If the BOP1 expression is higher than 6.84, it will be classified as metastatic prostate cancer. As both nodes have a value of 0, it is a terminal node.

4.0 Discussion

Based on the result of this study, a few key findings can be further discussed. Firstly, mRNA gene expression of 26 genes identified through feature selection can be used to differentiate between PCa and MPC. This is evident from the clear separation observed in the PCA as well as good sensitivity in the prediction model. The results of the prediction model should be further improved by incorporating more data and externally validated in order to provide a clearer picture on whether a predictive model would be feasible in the future as a diagnostic tool.

Table 4: Cross validation results of the predictive model of mRNA gene expression between PCa and MPC. Abbreviation: TP: True Positive; FP: False Positive; TN: True Negative; FN: False Positive

TP	FP	TN	FN	Sensitivity	Specificity
162	115	2	101	0.616	0.017

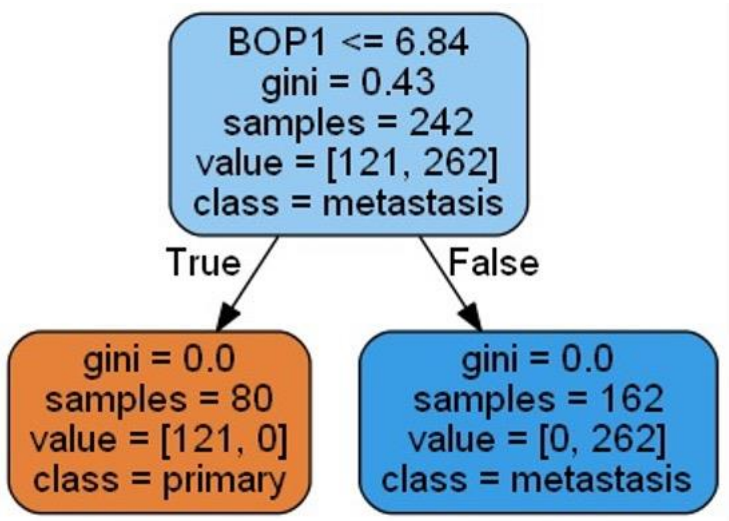

Figure 4: One of the decision trees of the random forest generated in the study. Here, the most important feature is the gene BOP1. Primary refers to PCa and Metastasis refers to MPC.

Secondly, this study highlights potential pathway of MPC involving the mutation of AR, which may be driven by coactivators such as PAG24. PAG24 is an established corepressor of AR, however, its involvement in the pathogenesis of PCa and MPC has not been studied. Several coregulators of AR have been well studied such as SRC1-3 (58, 59), These proteins bind to the amino-terminal domain (NTD) of AR, thereby prompting its transactivation directly through histone acetyltansferase activity and indirectly through recruitment of secondary coactivators to stimulate chromatin re-modelling (59). Several small molecule inhibitors (60) as well as peptides (61) have been developed to target AR coregulators for CPRC. Hence, PAG24 could be a potential target for MPC and future studies should investigate this. BOP1and GRWD1 were predicted to interact with PAG24 where the former is overexpressed in different stages of PCa. The expression of GRWD1 has not been analysed in PCa, but it is overexpressed in colon carcinoma. The knockdown of both genes was shown to reduce the expression of cancer phenotypes such as cell proliferation, migration and invasion.

Thirdly, the result of this study suggests that a potential crosstalk may exist between androgen and NF-κB signalling pathways. This is due to several genes from Tables 2 and 3 being found to directly or indirectly affecting the NF-κB signalling pathway such as TRIM40, EPN1 and RBM23. NF-

κB signalling pathway is involved in inflammation, immunity, cell proliferation, differentiation and apoptosis. The pathway is altered in both hematopoietic and solid malignancies, which promotes the proliferation and survival of tumour cells. Malinen et al., (62) have demonstrated that simultaneous pro-inflammatory and androgen signalling are able to significantly reprogram NF-κB and AR cistromes. Modulation of both cistromes may lead to the progression of PCa. TRIM40 is a member of the TRIM family and plays an important role in the innate response. TRIM40 was found to inhibit NF-κB activity *via* neddylation of IKKγ, which prevents inflammation-associated carcino-genesis in the gastrointestinal tract (55). While dysregulation of AR signalling is the initial driver of PCa, the pathway does not function in isolation. Crosstalk between androgen signalling and other pathways has been demonstrated to be a potential avenue that drives PCa progression. Several intracellular kinases such as SRC, MAPK, PI3K/AKT and ERK1/2 are downstream regulators of nongenomic AR signalling. This mediates a proliferation response and potentially driving PCa progression. Furthermore, several cell surface receptors such as interleukin (IL)-6, IL-8, EGFR, IGF-1 and HER2/NEU were implicated in the cross talk with AR to either sensitize AR at sub-physiological androgen concentrations or drive ligand independent signalling. One of the surface receptors mentioned, HER2, is implicated in breast cancer where it may be overexpressed leading to proliferation of cancer cells. Several genes in Tables 2 and 3 have been connected to breast cancer such as PABPC3 and PA2G4. The similarities between breast and prostate cancer have been explored where it has been shown that males who have female family members with a history of breast cancer are at a higher chance of developing prostate cancer. Follow-up studies have shown that both cancers share the same mutations such as BRCA1, and BRCA2. Recently, Olaparib which was originally prescribed for breast cancer has been approved as treatment for prostate cancer. Hence, the genes PABPC3 and PA2G4 should be further validated and the similarities between the two cancers should be further explored.

5.0 Conclusion

In this study, unsupervised and supervised machine learning methods were employed to differentiate the genetic landscape between PCa and MPC. Several findings can be deduced, which were: (i) mRNA expression can be used to differentiate between PCa and MPC, (ii) the AR-PAG24-BOP1- GRWD1 axis should be investigated further as both potential biomarkers and as well as potential pathways in MPC progression and (iii) androgen signalling pathway may crosstalk with NF-κB signalling pathway and breast cancer pathway. Future studies should include experimental validation of the genes identified here, as well as using more data in the predictive model. One limitation of this study is that a general mutation analysis using odds ratio was performed. A detailed analysis incorporating the type of mutation as well as its location would provide more information. Nevertheless, as this is a preliminary study, the results shown were corroborated by scientific literature and could serve as the foundation for future studies.

Acknowledgements

This research was supported by Universiti Teknologi MARA through the Global Research Reputation Grant (600-RMC 5/3/GRR (003/2020))

Conflict of interest

The authors declare no conflict of interest in the present work.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer J Clin. 2021;71 (3):209-249.

2. Kimura T, Egawa S. Epidemiology of prostate cancer in Asian countries. Int J Urol. 2018;25(6): 524-531.

3. Dai C, Heemers H, Sharifi N. Androgen signaling in prostate cancer. Cold Spring Harb Perspect Med. 2017;7(9):1-19.

4. Sartor O, de Bono JS. Metastatic prostate cancer. N Engl J Med. 2018;378(7):645-657.

5. Jespersen CG, Nørgaard M, Borre M. Androgen-deprivation therapy in treatment of prostate cancer and risk of myocardial infarction and stroke: a nationwide Danish population-based cohort study. Eur Urol. 2014;65(4):704-709.

6. Liao K-M, Huang Y-B, Chen C-Y, Kuo C-C. Risk of ischemic stroke in patients with prostate cancer receiving androgen deprivation therapy in Taiwan. BMC Cancer 2019;19(1):1-9.

7. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, et al. Integrative Clinical Genomics of Advanced Prostate Cancer Cell 2015;161(5):1215-1228.

8. Pritchard CC, Mateo J, Walsh MF, De Sarkar N, Abida W, Beltran H, et al. Inherited DNA-repair gene mutations in men with metastatic prostate cancer. N Engl J Med. 2016;375:443-453.

9. Li R, Dong X, Ma C, Liu L. Computational identification of surrogate genes for prostate cancer phases using machine learning and molecular network analysis. Theor Biol Med Model. 2014;11(1):1-12.

10. Xue M, Liu H, Zhang L, Chang H, Liu Y, Du S, et al. Computational identification of mutually exclusive transcriptional drivers dysregulating metastatic microRNAs in prostate cancer. Nat Commun. 2017;8(1): 1-9.

11. Bello T, Paindelli C, Diaz-Gomez LA, Melchiorri A, Mikos AG, Nelson PS, et al. Computational modeling identifies multitargeted kinase inhibitors as effective therapies for metastatic, castration-resistant prostate cancer. PNAS 2021;118(40):1-11.

12. Gao J, Aksoy BA, Dogrusoz U. Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269):l1-l20

13. Gerhauser C, Favero F, Risch T, Simon R, Feuerbach L, Assenov Y, et al. Molecular evolution of early-onset prostate cancer identifies molecular risk markers and clinical trajectories. Cancer Cell 2018;34 (6):996-1011.

14. Abida W, Cyrta J, Heller G, Prandi D, Armenia J, Coleman I, et al. Genomic correlates of clinical outcome in advanced prostate cancer. PNAS. 2019;116(23): 11428-11436.

15. Armenia J, Wankowicz SA, Liu D, Gao J, Kundra R, Reznik E, et al. The long tail of oncogenic drivers in prostate cancer. Nat Genet. 2018;50(5):645-651.

16. Abida W, Armenia J, Gopalan A, Brennan R, Walsh M, Barron D, et al. Prospective genomic profiling of prostate cancer across disease states reveals germline and somatic alterations that may affect clinical decision making. JCO Precis Oncol 2017;1:1-16.

17. Shaikh, R. Feature selection techniques in machine learning with python. 2018. [cited 2021 Dec 20] Available from: https://towardsdatascience.com/feature-selection-techniques-inmachine-learning-with-python-f24e7da3f36e.

18. Bro R, Smilde A.K. Principal component analysis. Anal Methods. 2014;6(9):2812-2831.

19. Wickham H. An introduction to ggplot: An implementation of the grammar of graphics in R. Statistics 2006 [cited 2021 Dec 1] Available from: http://ftp.auckland.ac.nz/software/CRAN/doc/vignettes/ggplot/introduction.pdf

20. Jaadi Z. A step-by-step explanation of Principal Component Analysis (PCA). 2021 Apr 1 [cited 2021 7 June] Available from:https://builtin.com/data-science/step-

step-explanation-principal-component-analysis.

21. Persoskie A, Ferrer RA. A most odd ratio: interpreting and describing odds ratios. Am J Prev Med. 2017;52(2):224-228.

22. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J., et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607-D613.

23. Breiman L. Random Forests. Mach Learn. 2001;45(1):5-32.

24. Vega-Benedetti AF, Loi E, Moi L, Blois S, Fadda A, Antonelli M, et al. Clustered protocadherins methylation alterations in cancer. Clin Epigenetics 2019;11(1):1-20.

25. Cai X, Brophy M, Hahn S, McManus J, Chang B, Pasula S, et al. The role of epsin in promoting Epithelial-Mesenchymal Transition and metastasis by activating NF-kB signaling in breast cancer. Cancer Res. 2012;72(24 Suppl).

26. Song K, Wu H, Rahman HA, Dong Y, Wen A, Brophy ML, et al. Endothelial epsins as regulators and potential therapeutic targets of tumor angiogenesis. Cell Mol Life Sci. 2017;74(3):393-398.

27. Rakheja D, Chen KS, Liu Y, Shukla AA, Schmid V, Chang T-C, et al. Somatic mutations in DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in Wilms tumours. Nat Commun. 2014;5(1):1-11.

28. Mbimba T, Hussein NJ, Najeed A, Safadi FF. TRAPPC9: Novel insights into its trafficking and signaling pathways in health and disease. Int J Mol Med. 2018;42 (6):2991-2997.

29. Li X, Tian L, Zhang L, Xu B, Zhang Y, Li Q. Clinical Significance of ZNF711 in Human Breast Cancer. Onco Targets Ther 2020;13:6593.

30. Vellky JE, Ricke EA, Huang W, Ricke WA. Expression, Localization, and Function of the Nucleolar Protein BOP1 in Prostate Cancer Progression. Am J Pathol. 2021;191(1): 168-179.

31. Zhu C, Yang Q, Xu J, Zhao W, Zhang Z, Xu D, et al. Somatic mutation of DNAH genes implicated higher chemotherapy response rate in gastric adenocarcinoma patients. J Transl Med. 2019;17(1):109.

32. Zhou J-h, Yao Z-x, Zheng Z, Yang J, Wang R, Fu S-j, et al. G-MDSCs-derived exosomal miRNA-143-3p promotes proliferation via targeting of ITM2B in lung cancer. Onco Targets Ther. 2020;13:9701-9719.

33. Hamdi Y, Boujemaa M, Ben Rekaya M, Ben Hamda C, Mighri N, El Benna H, et al. Family specific genetic predisposition to breast cancer: results from Tunisian whole exome sequenced breast cancer cases. J Transl Med. 2018;16(1):1-13.

34. Hamburger A.W. The role of ErbB3 and its binding partners in breast cancer progression and resistance to hormone and tyrosine kinase directed therapies. J. Mammary Gland Biol. Neoplasia 2008;13 (2):225-233.

35. Stevenson BW, Gorman MA, Koach J, Cheung BB, Marshall GM, Parker MW, et al. A structural view of PA2G4 isoforms with opposing functions in cancer. J Biol Chem. 2020;295(47):16100-16112.

36. Feng Y, Zhang Y, Cai Y, Liu R, Lu M, Li T., et al. A20 targets PFKL and glycolysis to inhibit the progression of hepatocellular carcinoma. Cell Death Dis. 2020;11(2):1-15.

37. Burdelski C, Borcherding L, Kluth M, Hube-Magg C, Melling N, Simon R., et al. Family with sequence similarity 13C (FAM13C) overexpression is an independent prognostic marker in prostate cancer. Oncotarget 2017;8(19):31494-3150.

38. Mao W, Xiong G, Wu Y, Wang C, St Clair D, Li J-D, et al. RORα Suppresses Cancer-Associated Inflammation by Repressing Respiratory Complex I-Dependent ROS Generation. Int J Mol Sci. 2021; 22(19):1-17.

39. Han H, Lin T, Fang Z, Zhou G. RBM23 Drives Hepatocellular Carcinoma by Activating NF-κB Signaling Pathway. BioMed Res Int 2021;1-9.

40. Adamopoulos PG, Raptis GD, Kontos CK, Scorilas A. Discovery and expression

analysis of novel transcripts of the human SR-related CTD-associated factor 1 (SCAF1) gene in human cancer cells using Next-Generation Sequencing. Gene 2018; 670:155-165.

41. Ding N, Li R, Shi W, He C. CENPI is overexpressed in colorectal cancer and regulates cell migration and invasion. Gene 2018;674:80-86.

42. Kidger AM, Sipthorp J, Cook SJ. ERK1/2 inhibitors: New weapons to inhibit the RAS-regulated RAF-MEK1/2-ERK1/2 pathway. Pharmacol Ther 2018;187:45-60.

43. Zhan P, Xi GM, Zhang B, Wu Y, Liu HB, Liu YF, et al. NCAPG 2 promotes tumour proliferation by regulating G2/M phase and associates with poor prognosis in lung adenocarcinoma. J Cell Mol Med 2017;21 (4):665-676.

44. Formaggio N, Rubin MA, Theurillat J-P Loss and revival of androgen receptor signaling in advanced prostate cancer. Oncogene 2021;40(7):1205-1216.

45. Chen W, Gao D, Xie L, Wang A, Zhao H, Guo C, et al. SCF-FBXO24 regulates cell proliferation by mediating ubiquitination and degradation of PRMT6. Biochem Biophys Res Commun 2020;530(1):75-81.

46. Fernandez-Cuesta L, Plenker D, Osada H, Sun R, Menon R, Leenders F, et al. CD74–NRG1 Fusions in Lung Adenocarcinoma. Cancer Discov 2014;4(4):415-422.

47. Dalva M, El Jellas K, Steine SJ, Johansson BB, Ringdal M, Torsvik J, et al. Copy number variants and VNTR length polymorphisms of the carboxyl-ester lipase (CEL) gene as risk factors in pancreatic cancer. Pancreatology 2017;17(1):83-88.

48. Wang H-X, Sharma C, Knoblich K, Granter SR, Hemler ME. EWI-2 negatively regulates TGF-β signaling leading to altered melanoma growth and metastasis. Cell Res. 2015;25(3):370-385.

49. Penha RCC, Buexm LA, Rodrigues FR, de Castro TP, Santos MCS, Fortunato RS, et al. NKX2.5 is expressed in papillary thyroid carcinomas and regulates differentiation in thyroid cells. BMC Cancer 2018;18(1):498-498.

50. Plaitakis A, Kalef-Ezra E, Kotzamani D, Zaganas I, Spanaki C. The Glutamate Dehydrogenase Pathway and Its Roles in Cell and Tissue Biology in Health and Disease. Biology (Basel) 2017;6(1):1-11.

51. Jiang T, Guo J, Hu Z, Zhao M, Gu Z, Miao S. Identification of Potential Prostate Cancer-Related Pseudogenes Based on Competitive Endogenous RNA Network Hypothesis. Med Sci Monit. 2018;24: 4213-4239.

52. Takafuji T, Kayama K, Sugimoto N, Fujita M. GRWD1, a new player among oncogenesis-related ribosomal/nucleolar proteins. Cell cycle (Georgetown, Tex.) 2017;16(15):1397-1403.

53. Cui X, Lin Z, Chen Y, Mao X, Ni W, Liu J, et al. Upregulated TRIM32 correlates with enhanced cell proliferation and poor prognosis in hepatocellular carcinoma. Mol Cell Biochem. 2016;421(1-2):127-137.

54. Lau Y-FC, Li Y, Kido T. Battle of the sexes: contrasting roles of testis-specific protein Y-encoded (TSPY) and TSPX in human oncogenesis. Asian J Androl 2019;21(3):260-269.

55. Noguchi K, Okumura F, Takahashi N, Kataoka A, Kamiyama T, Todo S, et al. TRIM40 promotes neddylation of IKKγ and is downregulated in gastrointestinal cancers. Carcinogenesis 2011;32(7):995-1004.

56. Zhou X, Shang J, Liu X, Zhuang J-F, Yang Y-F, Zhang Y-Y, et al. Clinical Significance and Oncogenic Activity of GRWD1 Overexpression in the Development of Colon Carcinoma. Onco Targets Ther. 2021;14:1565-1580.

57. Hamdi Y, Boujemaa M, Rekaya MB, Hamda CB, Mighri N, El Benna H, et al. Family specific genetic predisposition to breast cancer: results from Tunisian whole exome sequenced breast cancer cases. J Transl Med. 2018;16(1):1-13.

58. Powell S, Christiaens V, Voulgaraki D, Waxman J, Claessens F, Bevan C. Mechanisms of androgen receptor signalling via steroid receptor coactivator-1 in prostate. Endocr.-Relat. Cancer 2004;11(1):117-130.

59. Chakravarti D, LaMorte VJ, Nelson MC, Nakajima T, Schulman IG, Juguilon H, et al. Role of CBP/P300 in nuclear receptor signalling. Nature 1996;383(6595):99-103.

60. Wang Y, Lonard DM, Yu Y, Chow D-C, Palzkill T.G., Wang J., et al. Bufalin is a potent small-molecule inhibitor of the steroid receptor coactivators SRC-3 and SRC-1. Cancer Res. 2014;74(5):1506-1517.

61. Ravindranathan P, Lee T-K, Yang L, Centenera MM, Butler L, Tilley WD, et al. Peptidomimetic targeting of critical androgen receptor–coregulator interactions in prostate cancer. Nat Commun. 2013;4(1):1-11.

62. Malinen M, Niskanen EA, Kaikkonen MU, Palvimo JJ. Crosstalk between androgen and pro-inflammatory signaling remodels androgen receptor and NF-κB cistrome to reprogram the prostate cancer cell transcriptome. Nucleic Acids Res. 2017;45 (2):619-630.