Identification of key pathways and genes with aberrant methylation in prostate cancer using bioinformatics analysis

Anshika N Singh
Neeti Sharma
Symbiosis School of Biological Sciences, Symbiosis International University, Gram – Lavale, Taluka – Mulshi, Pune, India

Abstract: Prostate cancer (PCa), a multifocal clinically heterogeneous disease, is the most commonly diagnosed non-cutaneous neoplasm in men worldwide. The epigenome of PCa is a typical representation of catastrophic model of epigenetic alterations during tumorigenesis and its progression. Alterations in methylation patterns in tumor suppressors, cell cycle, oncogenes and metabolism-related genes are the most commonly observed epigenetic alterations in PCa. In this study, we have developed a computational strategy to identify methylated biomarker signature panels as potential targets of PCa by screening 160 genes reported to be epigenetically dysregulated, and shortlisted 26 differentially methylated genes (DMGs) that significantly contribute to oncogenesis. The gene ontology and functional enrichment analysis were performed, which showed that identified DMGs contribute to cellular and metabolic processes such as cell communication, cell cycle, response to drugs, apoptosis and p53 signaling. The top hub genes AR, CDH13, CDKN2A, DAPK1, GSTP1, CD44 and RASSF1 identified from protein–protein interaction network construction using Search Tool for the Retrieval of Interacting Genes contributed to hormonal response, inflammatory response, cell cycle, reactive oxygen species activity and receptor kinase activity, which are related to hallmarks of cancer as revealed by their functional enrichment analysis by Cytoscape. These genes were further scrutinized for CpG islands, transcription start sites and positions of methylated cytosines to study their methylation profiles. Our analysis revealed high negative correlation values between methylation frequencies and gene expressions of the hub genes, namely, AR, CDH13, CDKN2A, DAPK1, CD44, GSTP1 and RASSF1, which can be used as potential diagnostic biomarkers for PCa.

Keywords: prostate cancer, CpG islands, epigenetic modification, DNA methylation

Introduction
Prostate cancer (PCa) accounts for being the most commonly diagnosed non-cutaneous neoplasm and third leading cause of cancer mortalities in men worldwide. Reports suggest that men aged >40 years are ~30-fold more susceptible to developing PCa. Although a number of genetic, surface and intracellular markers have been identified to study cancer progression such as CD133, CD44f, Trop2, p63 and PSA, PSA is the only established biomarker for the detection of PCa. However, PSA lacks in specificity and sensitivity; thus, there exists an urgent need to identify biomarkers that are more specific.

Epigenetic alterations in the form of DNA methylation and histone modifications have emerged as a major focus area for cancer treatment considering their contribution in modulation of crucial gene expressions in various malignancies. Tumor suppressor
genes and oncogenes play crucial roles in cancer by regulating functions such as cell cycle control, DNA repair, cell adhesion and apoptosis and are known to be differentially methylated in their promoter regions in several types of cancers. However, till date, there is a limited knowledge of the specific functional mechanisms at the genome level that are being altered by epigenetic dysregulation during tumor progression in PCa. Given their key role in various stages of carcinogenesis by causing abnormal gene expression of tumor suppressor genes and oncogenes, histone modifications and hypermethylation status of genes are being explored as potential biomarkers of cancer progression and prognosis. Nowadays with the advancement of chromatin immunoprecipitation (ChIP) technology along with microarray profiling, scientists can screen thousands of differentially expressed genes (DEGs), analyze their mRNA profiles and methylation profiles computationally and subsequently scrutinize their involvement in various molecular and biological regulatory functions during various stages of tumor initiation and progression.

The present study aims to explore the contribution of epigenetic modulation of genes involved in progression of PCa using computational approach. We screened >160 differentially methylated genes (DMGs), which are reported to be epigenetically dysregulated, and identified 26 genes that significantly contribute to oncogenesis in prostate gland. This was followed by screening the identified DMGs for gene ontology (GO) and their pathway enrichment. The protein–protein interaction (PPI) and functional enrichment analysis further shed light on the crucial roles played by the identified DMGs during tumor initiation and progression. Our analysis showed a high negative correlation value between methylation frequencies and gene expressions of identified hub genes, which have the potential to act as early prognostic and diagnostic biomarkers of PCa.

**Methods**

**Identification of DMGs**

A comprehensive literature survey was performed using PubMed and Google Scholar. The keywords used were “Prostate cancer” and “DNA methylation”. The literature survey was limited to full text articles available in English. A preliminary abstract review was performed to decide relevance of these research articles to our methylation study. Once the literature survey was updated, an extensive data extraction exercise was carried out to retrieve genes that are differentially methylated in PCa using online databases. The Cancer Genome Atlas (TCGA) (cancergenome.nih.gov), Cancer Genetics Web (www.cancerindex.org/geneweb), and MethyCancer (methycancer.psych.ac.cn) were used for identifying the DMGs in PCa.

**GO and functional enrichment analysis**

Databases such as GeneCodis and Panther were used for functional enrichment analysis. GO of identified DMGs included categories such as cellular component (CC), biological process (BP) and molecular function (MF) terms. The parameters were set as p-value < 0.01 for considering the results that were statistically significant. The enriched pathways were identified using Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. To further analyze the interrelationship of identified hub genes in regulating cellular pathways, gene set enrichment analysis module of Cytoscape software was used.

**PPI network construction by search tool for the retrieval of interacting genes** (STRING)

The interactive relationships between DMGs were identified using STRING database. A combined score of >0.7 (high) of only experimentally validated interactions was considered to be statistically significant.

**Epigenetic analysis**

The FASTA sequences of identified differentially methylated hub genes in PCa were retrieved from nucleotide database of National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). CpGProD (doua.prabi.fr/software/cpgprod) and EMBoss CpGPlot (www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot) were used to monitor the presence of CpG islands if any. Information related to GC%, G + C skew, A + T skew, etc. was also retrieved from these online programs. Methylation and Expression database of Normal and Tumor tissues (MENT; mgrc.kribb.re.kr:8080/MENT) is an online integrative database to study DNA methylation and its correlation with gene expression in paired samples gathered from TCGA and Gene Expression Omnibus (GEO). MENT was used to identify the effect of methylation on the gene expression of the identified DMGs in PCa. Analysis of DNA methylation with exon–intron structures, genomic loci of epigenetic changes, positions of methylated cytosines, CpG islands and transcription start sites (TSS) using a graphical description was carried out using Prostate Epigenetic Database (PEpiD; wukong.tongji.edu.cn/pepid). The identified hub genes after epigenetic analysis were further classified using extensive literature survey based on their specificity and sensitivity for determining their performance as prognostic and
Results
Identification of DMGs involved in PCa initiation and progression
A consolidated list of >160 genes from literature survey and online databases with evidences of differential methylation during PCa progression was prepared (Table S1). We shortlisted 26 candidate DMGs that were associated with metastatic and recurrent stages of the disease with at least two publications and were used for further analysis (Table 1).

Functional annotation and pathway enrichment analysis
The altered biological function and MF of the identified DMGs were then analyzed through GO using GeneCodis and Panther. A p-value of <0.01 was considered as significant during the analysis. The enrichment GO terms were categorized into MF, BP and CC (Table 2). During the MF analysis, we found that majority of DMGs contributed to protein binding (16 genes), zinc ion binding (seven genes), metal ion binding (eight genes), DNA binding (six genes) and receptor activity (six genes). Other regulated MFs included enzyme binding, sequence-specific DNA binding and calcium ion binding. The CC analysis revealed that majority of differentially methylated candidate genes were associated with cytoplasm (20 genes), nucleus (16 genes), plasma membrane (10 genes), nucleoplasm (seven genes) and cytosol (six genes). In the BP ontology, the most significantly regulated processes included response to drugs (six genes), negative regulation of cell proliferation (six genes), signal transduction (seven genes) and negative regulation of apoptotic process (five genes). The other regulated processes included response to estradiol stimulus, cell cycle arrest and negative regulation of MAPK cascade. Thus, the most significantly enriched pathways identified using KEGG analysis include pathways in cancer (10 genes), cell cycle (five genes), p53 signaling (three genes) and PCa (three genes; Table 2). Functional gene set enrichment analysis using gene set enrichment analysis (GSEA) module of Cytoscape showed that the identified DMGs most commonly contributed in hormonal response, inflammatory response, cell cycle, reactive oxygen species (ROS) activity and receptor kinase activity, which are all related to hallmarks of oncogenesis (Table S2).

PPI network
The PPI networks between DMGs were investigated using STRING server. In total, 26 nodes with 90 edges and 6.92 as average node degree were analyzed (Figure 1 and Table 3). In the PPI network, seven node proteins, namely, GSTP1, CDH13, CDKN2A, RASSF1C, CD44, AR and DAPK1, showed maximum degree of association with the other proteins (>5), thus indicating that these genes (proteins) have higher hub degrees. Thus, these hub genes (proteins) may play a crucial role in PCa initiation and progression.

Epigenetic analysis
Evidences have shown that the cytosines in the CpG dinucleotides are often methylated to 5-methylcytosines leading to change in gene expression.46 Almost 70%–80% cytosines in CpG mammals are seen to be methylated in mammals and are often located near the promoter regions.47 CpGProd and EMBoss CpGPlot were used to analyze positions of CpG islands, CpG/o/e ratio, G + C skew and A + T skew (Table S3). The positions of methylated cytosines along
with positions of TSS of the identified hub DMGs were also retrieved using PEpID (a prostate epigenetic database in mammals; Table S4). Once the CpG islands and the methylated cytosines positions were retrieved, MENT database was used to scrutinize the correlation between methylation frequencies and gene expressions. The MENT launched in 2012 is an integrated database of DNA methylation and gene expressions. Datasets containing information regarding DNA methylation and gene expression in paired samples were retrieved from TCGA and GEO. We opted for gene search tool to predict the correlation between methylation values and gene expressions of our hub genes. The Illumina methylation datasets consists of Human Methylation27 and Golden Gate Methylation Cancer Panel I, which is composed of 125 prostate tumor samples and 92 normal prostate samples for analysis. Based on the correlation values between mean methylation frequencies and gene expression, it was observed that expressions of hub genes GSTP1, CDH13, CDKN2A, RASSF1, AR, CD44 and DAPK1 were seen to be significantly affected by their hypermethylation during PCa progression due to an observed negative correlation of as high as −0.7 as shown in Table 4.

These identified hub genes, ie, AR, CDH13, CDKN2A, DAPK1, GSTP1, CD44 and RASSF1, have been reported by various other research groups to show high specificity and sensitivity as independent biomarkers in prostatic intraepithelial neoplasia and different stages of PCa, including primary and metastatic tumors (Table 5). These genes have been reported to be differentially regulated due to hypermethylation in PCa cells, ie, PC3, DU145 and LNCaP, as well as bodily samples, including serum, ejaculates, post-massage urine and biopsy samples.\textsuperscript{22–45}
Discussion
Alterations in methylation patterns in tumor suppressor genes, cell cycle genes, oncogenes and metabolism-related genes are the most commonly observed epigenetic alterations in PCa. With the advancement in high-throughput technologies, including microarray and ChIP array, it has become feasible to analyze not only the expression levels of several genes simultaneously but also their methylation frequencies that can eventually lead to discovery of genes that can act as potential diagnostic, prognostic and therapeutic biomarkers. In the present study, we evaluated a panel of 26 DMGs using computational approaches to study their differential methylation and gene annotation profiles that are regulated during different stages of PCa. Consequently, their interrelationships were analyzed using PPI server.

Table 3 Associated genes of identified seven differentially methylated hub genes

<table>
<thead>
<tr>
<th>Hub gene</th>
<th>Connected genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>ESR1, CAV1, ESR2, CDH1, RARβ, CDKN1B</td>
</tr>
<tr>
<td>CDH1</td>
<td>CDH1, RASSF1, DAPK1, GSTP1, CDKN2A, RARβ, FHIT, MGMT, TIMP3, HIC1</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>CCND2, CDKN1B, RASSF1, FHIT, CDH1, MGMT, GSTP1, ESR1, TIMP3, HIC1, RARβ, SFN, AR, ESR2</td>
</tr>
<tr>
<td>DAPK1</td>
<td>RASSF1, TIMP3, RARβ, MGMT, HIC1, FHIT, CDH1</td>
</tr>
<tr>
<td>GSTP1</td>
<td>GPX3, TIMP3, RASSF1, DAPK1, MGMT, CDH1, RARβ, HIC1, ESR1, FHIT, SFN, PTGS2, AR, CCND2, ESR2</td>
</tr>
<tr>
<td>RASSF1</td>
<td>FHIT, SFN, RARβ, MGMT, HIC1, TIMP3, CDH1, CCND2, ESR1, ESR2</td>
</tr>
<tr>
<td>CD44</td>
<td>AR, CDKN1B, CAV1, ESR1, CDKN2A, TIMP3, PTGS2</td>
</tr>
</tbody>
</table>

Table 4 Methylation and Expression database of Normal and Tumor tissues (MENT) data showing correlation between mean methylation values and gene expressions

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Methylation</th>
<th>Probe</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDH1</td>
<td>ILMN_1766925</td>
<td>cg 13759328</td>
<td>−0.53224</td>
</tr>
<tr>
<td>DAPK1</td>
<td>ILMN_1708340</td>
<td>cg 08797471</td>
<td>−0.75606</td>
</tr>
<tr>
<td>RASSF1</td>
<td>ILMN_1808066</td>
<td>cg 06821120</td>
<td>−0.44059</td>
</tr>
<tr>
<td>AR</td>
<td>ILMN_1792540</td>
<td>cg 05786601</td>
<td>−0.56984</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>ILMN_1717714</td>
<td>cg 11653709</td>
<td>−0.68581</td>
</tr>
<tr>
<td>GSTP1</td>
<td>ILMN_1679869</td>
<td>cg 05244766</td>
<td>−0.55374</td>
</tr>
<tr>
<td>CD44</td>
<td>ILMN_1778625</td>
<td>cg 04125208</td>
<td>−0.39674</td>
</tr>
</tbody>
</table>
Using thorough literature survey and online databases such as TCGA, Cancer Genetics Web and MethyCancer, we identified 26 DMGs in PCa. Here, we hypothesized that DMGs may act as potential diagnostic and prognostic biomarkers for PCa. However, to better understand the inter-relationship and contribution of DMGs in cancer progression, we further performed GO and KEGG pathway enrichment analysis. The GO analysis showed that the DMGs significantly contributed in MFs such as protein binding, DNA binding, zinc ion binding and metal ion binding, which may contribute to carcinogenesis. Our hypothesis was further strengthened by the BP analysis which showed that these 26 DMGs significantly contributed to signal transduction, negative regulation of cell proliferation and negative regulation of apoptosis. These identified BPs pointed toward the involvement of the identified DMGs in major hallmarks of cancer. CDKN2A, CDKN1B and CDKN1C among the identified genes are known to be critical regulators of cell cycle, acting as checkpoints at various stages of cell cycle. aberrantly, methylated cyclin-dependent kinase genes lead to carcinogenesis and progression of cancer due to uncontrolled cell growth.46 CDKN2A has been seen to be frequently methylated in PCa progression.48–50 CCND2 is another important identified cell cycle regulator required for cell cycle G1/S transition.51 Other cell cycle genes such as CD44,52 HIC153 and PTGS254 identified in our analysis have been reported to be frequently methylated during PCa progression. Another important DMG identified was stratifin, which is known to inhibit the cyclin B1–CD2 complex from entering the nucleus by enforcing a G2/M arrest. The loss of stratifin due to promoter hypermethylation during tumor progression has been seen in various cancers, including PCa.55 Our analysis also identified endothelin B receptor type B, protein encoded by the EDNRB gene, which is known to induce apoptosis and inhibit tumor progression by interacting with endotherins. Several studies have reported promoter hypermethylation of EDNRB in higher frequencies in prostate tumors as compared to normal tissues.56–58 KEGG analysis showed that the major enriched pathways included pathways in cancer, cell cycle, p53 signaling, focal adhesion and Wnt signaling (Table 2). Wnt signaling is reported to play integral roles in cell fate decisions, proliferation, neural patterning, migration, differentiation, tumor aggression and stem cell renewal during many malignancies including PCa.59,60 In addition, differential regulation of p53 signaling as regulated by CCND2, SFN and CDKN2A has been reported in various cases of PCa.61 Reports have suggested that expressions of AR and p53, which are balanced in PCa development.

The PPI network constructed with identified DMGs led to the identification of high-degree hub genes (Table 3 and Figure 1) in the local network, which included AR, CDH13, CDKN2A, DAPK1, CD44, GSTP1 and RASSF1 that exhibited the highest degree of connectivity. Androgens that are responsible for stimulating growth of PCa cells mediate their effect in various cancers, including PCa.56,57
through androgen receptors (ARs). Silencing of AR can lead to decreased growth and induce apoptosis.\(^{32-45}\) AR promoter hypermethylation has been reported by various groups to be between 8% and 39% in PCa tumor tissues.\(^{2}\) The results of our analysis were in concordance with several published research papers that have reported the downregulated expression of AR due to hypermethylation during progression of PCa from androgen-dependent to androgen-independent stages.\(^{42-45}\) T-cadherin (CDH13), another identified hub gene located on 16q24, is a tumor suppressor gene known to play a critical role in cell–cell adhesion.\(^{70}\) CDH13 expression can restrain the invasive potential and proliferation rate of tumor cells. Anomalous expression of CDH13 due to promoter methylation has been reported in various malignancies, including breast cancer, lung cancer and bladder cancer.\(^{70}\) Yet, another gene identified in our analysis was glutathione S-transferase Pi 1 (GSTP1), which is known to be the most frequently epigenetically altered metabolic gene in PCa. DNA hypermethylation of GSTP1 is reported to be as high as 90% in prostate cancerous samples and 70% in prostatic interepithelial neoplasia.\(^{22-34}\) GSTP1 located on chromosome 11q3 eliminates foreign chemicals and detoxifies electrophilic toxic carcinogens and oxidants. Methylation in GSTP1 has been detected with high frequencies in PCa cell lines, urine samples, patient ejaculates and blood samples as compared to low or no methylation frequencies in healthy control patients.\(^{71}\) Ras association domain family protein 1 isoform A (RASSF1A), another gene identified, was found to be associated with more than eight identified DMGs. RASSF1A is a tumor suppressor gene usually associated with DNA repair and apoptosis.\(^{72}\) Evidences have shown that aberrant methylation of RASSF1A may hamper cell cycle control and DNA repair in tumors.\(^{72}\) Silencing of RASSF1A has been observed in various cancers such as non-small-cell lung cancer, colorectal cancer, breast cancer and PCa. In fact, some researchers have shown the frequency of hypermethylation of RASSF1A to be correlated with PCa progression and aggressiveness.\(^{72}\)

**Conclusion**

We identified seven DMGs, namely, AR, CDH13, CDKN2A, DAPK1, GSTP1, CD44 and RASSF1, with high negative correlation (~0.7), which depicts that the increase in methylation directly leads to decrease in the expression of these genes. Biological and functional annotation showed that this panel of selected candidate genes played a crucial part in cell proliferation, binding, Wnt regulation, enzyme regulator activities and drug response. Although nonspecific epigenetic targeting might have short-term undesirable effects, gene-specific demethylation approaches offer promising prognostic and therapeutic targets for PCa diagnosis. Thus, results of the present study may be used for basic and translational treatment; however, further in vitro and in vivo investigations are needed to confirm our hypothesis.

**Acknowledgments**

This work was supported by grants from the Symbiosis Centre for Research and Innovation (SCRI) and Symbiosis School of Biological Sciences (SSBS), Symbiosis International University (SIU), Lavale, Pune, India and Junior Research INSPIRE Fellowship (JRF) from the Department of Science and Technology, Government of India, to Miss Anshika Nikita Singh.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


