Methylation-driven genes PMPCAP1, SOWAHC and ZNF454 as potential prognostic biomarkers in lung squamous cell carcinoma

QINGQING ZHU1,2, JIA WANG2,3*, QIUJING ZHANG1,2, FUXIA WANG4, LIHUA FANG5, BAO SONG6, CHAO XIE2 and JIE LIU2*

1School of Medicine and Life Sciences, University of Jinan-Shandong Academy of Medical Sciences, Jinan, Shandong 250022; 2Department of Oncology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, Shandong 250117; 3Department of Oncology, Zibo Maternal and Child Health Hospital, Zibo, Shandong 255000; 4Department of Oncology, Yun Cheng Country People's Hospital, Heze, Shandong 274700; 5Department of Oncology, Chang Qing District People's Hospital, Jinan, Shandong 250300; 6Basic Laboratory, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, Shandong 250117, P.R. China

Received June 13, 2019; Accepted December 18, 2019

DOI: 10.3892/mmr.2020.10933

Abstract. Of the different types of lung cancer, lung squamous cell cancer (LUSC) has the second highest rates of morbidity and mortality, which have been increasing in recent years. Epigenetic abnormalities may serve as potential biomarkers and diagnostic and/or therapeutic targets, which may help to monitor and improve the prognosis of patients with cancer. In the present study, data were obtained from The Cancer Genome Atlas database and survival and joint survival analyses were conducted using the R MethylMix package. Peptidase, mitochondrial processing a subunit pseudogene 1 (PMPCAP1), sosondowah ankyrin repeat domain family member C (SOWAHC) and zinc finger protein (ZNF) 454 were identified as independent prognosis-related hub methylation-driven genes (MDGs). Of these three genes, PMPCAP1 and SOWAHC, characterized by hypomethylation and high expression levels, were associated with poor prognosis in patients with LUSC, whilst ZNF454 was associated with an improved prognosis. In addition, pathway enrichment analysis suggested that PMPCAP1, SOWAHC and ZNF454 were primarily involved in gene expression or transcription pathways. Furthermore, 5, 1 and 10 key methylation sites of PMPCAP1, SOWAHC and ZNF454, respectively, were confirmed to be significantly relevant to gene expression, establishing a basis for further investigation into the mechanisms and more precise targets of these 3 genes. In conclusion, the MDGs PMPCAP1, SOWAHC and ZNF454 may be potential prognostic biomarkers of LUSC for guiding diagnosis and therapy options, as well as providing a theoretical basis for further investigation.

Introduction

The latest data released by the World Health Organization revealed that lung cancer has the highest global morbidity and mortality rates of all malignant tumors, and this trend is increasing yearly (1). Based on biological characteristics, treatment and prognosis, lung cancer is classified as non-small-cell lung cancer (NSCLC) or small-cell lung cancer (SCLC). NSCLC accounts for ~85% of all lung cancer cases, of which lung squamous cell cancer (LUSC) accounts for 20-30%, and has a five-year survival rate of <15% (2,3). With the recent rapid development of gene detection methods and targeted drugs, the overall survival time (OS) of patients with NSCLC has significantly improved (4). However, not all patients benefit from targeted therapy; in LUSC, the frequency of gene mutations sensitive to targeted drugs is relatively low, and this is accompanied by poor efficacy and the occurrence of drug resistance. In addition, the prognosis and OS of patients with early-stage lung cancer are markedly more favorable compared with those of patients at an advanced disease stage. Therefore,
the identification of novel biomarkers and therapeutic targets is important for improving early diagnosis, treatment strategies and prognostic detection in patients with LUSC.

At present, the pathogenesis and progressive mechanisms of LUSC remain unclear, though the two most important mechanisms of tumorigenesis are gene mutations and epigenetic alterations (5,6). Relatively few gene mutation sites exist, particularly for early-stage patients; owing to severe fragmentation of tumor gene fragments in the blood, gene mutations are not suitable for the monitoring and diagnosis of early-stage cancer, and epigenetic changes provide a more suitable target (7,8). DNA methylation is easily detectable, and therefore the most studied, epigenetic modification, mediating the occurrence and development of cancer by regulating gene expression (8-10).

It has also been indicated that DNA methylation may occur prior to gene mutation, deeming it more suitable for the early detection of cancer. Studies examining methylation and tumors have recently attracted increased attention, including a series of studies concerning targeted epigenetic therapy approaches for acute myeloid leukemia (11). Even in solid tumors, methylated or epigenetic signatures have become an area of increasing interest, in such malignancies as breast cancer (12), esophageal carcinoma (13,14), epithelial ovarian (15) and liver cancer (16). These studies indicated that the methylation of some specific genes may affect gene expression, and is closely associated with the diagnosis and prognosis of some types of cancer. Therefore, the identification of abnormal gene methylation signatures may provide a basis for the early diagnosis, prognosis and targeted therapy for patients with tumors.

In the context of the era of big data, bioinformatics analysis serves an important role in the comprehensive research of carcinomas, utilizing high-throughput databases such as The Cancer Genome Atlas (TCGA). The present study utilized data from TCGA, in which the genetic information profiles and corresponding clinical data of multiple cancer types can be effectively extracted for analysis, bridging the gap between molecular biology research and clinical application (17,18). Methylation-driven genes (MDGs) were the primary focus of the present study, defined as genes of differentially methylated states and significant predictive transcriptional function; as such, MDGs were identified by dissecting and integrating the correlation between methylation state and the level of gene expression. Previous studies have confirmed that MDGs are more comprehensive and representative tumor biomarkers compared with differentially methylated genes (DMGs) (14). In the present study, methylation, gene expression and patient survival information were extracted from TCGA, and the MethylMix algorithm and survival analysis were used to identify hub MDGs and their prognostic signatures, with a view to providing a basis for individualized precision treatment in patients with LUSC.

Materials and methods

Data acquisition and preprocessing. Firstly, the DNA methylation and gene expression quantification data of patients with LUSC were downloaded from TCGA (https://www.cancer.gov/tcga/), along with the corresponding clinical information, which included details of prognostic or survival analysis. According to the TCGA data number, the data were divided into two groups, LUSC samples and normal samples. In these data, the normal samples were tissues adjacent to the tumor. The Illumina Human Methylation 450 k platform was used to transform and normalize the initial DNA methylation data, which were expressed as β-values (range, 0-1) corresponding with low to high methylation states (19). Gene expression quantification data were obtained in RNA-Seq format.

Screening of differentially expressed genes (DEGs), DMGs and MDGs in LUSC. The R edge package (http://bioconductor.org/packages/edgeR/) was used to identify and analyze DEGs by comparing gene expression quantification data between normal and cancerous specimens, with a fold change (FC)=5 and adjusted P-value ($p_{adj}$)=0.01 as the threshold. The limma package (http://bioinf.wehi.edu.au/limma/) was used to compare the methylation states of normal and cancerous specimens, and DMGs were screened out using a false discovery rate of 0.01 and log2FC=1.

Next, MDGs were identified using the R MethylMix package (http://bioconductor.org/packages/3.9/biocon/tools/MethylMix.html). MethylMix is a new algorithm developed by Gevaert et al. (20,21), which uses univariate b mixture modeling to determine the methylation state of genes in cancer samples, and the Wilcoxon rank test to categorize these into hyper- and hypomethylated groups compared with the methylation state in normal tissues. It may also be used to determine the correlation between DNA methylation state and gene expression level, the absolute value of correlation coefficient (ICor) representing the degree of correlation. In the present study, the screening conditions of MDGs with significant inverse correlation were set as $p_{adj}$<0.05, log2FC=0 and Cor<0.5.

Pathway analysis of MDGs in LUSC. To further investigate the dominating functional pathways of MDGs in LUSC, ConsensusPathDB (http://cpdb.molgen.mpg.de/) and Cytoscape.js was utilized to analyze and visualize the genetic interaction of high-throughput expression data (22,23). ConsensusPathDB is currently the most comprehensive database of functional interaction networks, integrating the functional aspects of genes, proteins, complexes and metabolites. Cytoscape.js is a graph library written in JavaScript that is used as visualization software for graph analysis (24). P<0.05 was set as the cut-off for minimum overlap criterion.

Survival and joint survival analysis of MDGs in LUSC. It is important to note that not all MDGs are significantly associated with cancer prognosis. In order to improve the understanding of the association between MDGs and patient survival, the MDGs independently associated with prognosis, classified as hub MDGs, were identified. Firstly, using the R package survival, Kaplan-Meier survival analysis and the log-rank test were conducted to determine the association between the methylation state of MDGs and the survival of patients with LUSC. $P<0.05$ was considered to indicate a statistically significant correlation.

Owing to the complexity of tumor tissue regulation by the combination of multiple factors, it was necessary and important to conduct joint survival analysis between the degree of methylation, the corresponding levels of MDG expression and patient survival. A joint survival curve was generated using the
survival package, with P<0.05 as the cutoff value. It should be noted that the MDGs screened out using MethylMix were all characterized by a significant inverse correlation (Cor <−0.5). Therefore, of the joint survival analysis, there were only two cases to determine survival: Hypermethylation and low expression; and hypomethylation and high expression.

Finally, using the first two steps the common genes were identified. PMPCAP1, SOWAHC, ZNF454 AND LINC00668 were statistically significant in the survival analysis, while PMPCAP1, SOWAHC, ZNF454 and ADH7 were statistically significant in the joint survival analysis. So, PMPCAP1, SOWAHC and ZNF454 were taken as the hub MDGs.

Correlation analysis between methylation sites and the expression of hub MDGs. Finally, to further examine the internal mechanisms and more precise targets of the 3 identified hub MDGs, data corresponding to the initial methylation sites of these genes were downloaded. The present study focused on the correlation between the methylation of abnormal methylation sites and the corresponding gene expression of hub MDGs. Both P<0.05 and |Cor|>0.5 used as the cut-offs for the identification of key methylation sites.

Results

Identification of MDGs in LUSC. Firstly, a total of 370 LUSC samples and 42 normal samples (from 372 cases) with DNA methylation data were downloaded from TCGA database; 502 LUSC samples and 49 normal samples (from 501 cases) with gene expression quantification data were also downloaded. Additionally, 366 of the patients with LUSC also possessed clinical data for prognostic and survival analysis. Secondly, the R edge and limma packages were used to compare data between cancerous and normal samples, respectively, and to screen out 994 DEGs and 356 DMGs. Finally, the MethylMix algorithm (p_adj<0.05, log2FC=0 and Cor <−0.5) was used to identify 30 MDGs with strong inverse correlation between DNA methylation state and gene expression (Fig. 1; Table I). The methylation models are presented in Figs. 2 and S1, and the correlation plots are demonstrated in Figs. 3 and S2.
Pathway analysis of MDGs in LUSC. Pathway analysis of the 30 identified MDGs was conducted using the ConsensusPathDB database (Fig. 4). The results revealed 3 primary pathways: ‘generic transcription’, ‘RNA polymerase II transcription’ and ‘gene expression (transcription)’. The largest numbers of genes were associated with these 3 pathways, and ~100% of all shared genes, and to the most genes from input (P<0.001).

Recognition of hub MDGs in LUSC. Initially, survival analysis between hyper- and hypomethylated MDGs revealed 4 genes with statistical importance: Peptidase, mitochondrial processing a subunit pseudogene 1 (PMPCAP; P=0.00173), sorsodowah ankyrin repeat domain family member C (SOWAH; P=0.04), zinc finger protein (ZNF) 454 (P=0.023) and LINC00668 (P=0.046; Figs. 5A-D and S3). Joint survival analysis was then conducted between the degree of methylation and the corresponding gene expression level of MDGs and survival, and PMPCAP1 (P=0.041), SOWAH (P=0.028), ZNF454 (P=0.00935) and ADH7 (P=0.033) were identified as statistically significant (Figs. 5E-H and S4). By taking the common genes of the first two steps, 3 hub MDGs (PMPCAP1, SOWAH and ZNF454) were deemed to be independently associated with prognosis in LUSC. Of these 3 hub MDGs, PMPCAP1 and SOWAH, characterized by hypomethylation and high expression levels, were associated with poor prognosis in patients with LUSC, whilst ZNF454, characterized by hypermethylation and low expression level, was associated with an improved prognosis.

Key methylation sites of hub MDGs in LUSC. Using the associated R packages, key methylation sites statistically relevant to the expression of hub MDGs in LUSC were identified. The
results of correlation analysis revealed 5 key methylation sites of the PMPCAP1 gene (cg06551022, cg14777507, cg07794230, cg10697010 and cg16254375), 1 key methylation site of the SOWAHC gene (cg19399885) and 10 key methylation sites of the ZNF454 gene (cg17840719, cg16536329, cg23037403, cg20778451, cg24843380, cg03234732, cg02165355, cg10575261, cg10902717 and cg05461386; Fig. 6 and Table II; P<0.001).

Discussion

The morbidity and mortality rates of LUSC are the second highest of all the pathological types of lung carcinoma, with poor prognosis depending on the biological characteristics of the specific subtype. Furthermore, >70% of patients with LUSC present with late-stage disease at the diagnosis, for which treatment options are limited, and clinical outcomes are
far poorer compared with those in patients with early-stage disease (25). In order to decrease the mortality rate of LUSC, novel approaches for early diagnosis and treatment are required, as well as the identification of novel predictive biomarkers and therapeutic targets.

Compared with the rapid development of precise gene-targeted therapy in lung adenocarcinoma, there are a limited number of effective and distinctive targets to improve the prognosis of patients with LUSC. In addition to studies into gene mutations, the association between epigenetic changes (particularly DNA methylation) and LUSC has also attracted great attention. Epigenetic studies have revealed that genome-scale epigenetic modifications, including DNA methylation, histone modification and microRNA interference, are involved in the pathogenic mechanisms of malignancy (26). Due to its stability and ease of detection,
Accumulating evidence has demonstrated that aberrant gene methylation may serve as an effective, non-invasive diagnostic biomarker and therapeutic in carcinoma (27,28). Reports have indicated that the abnormal methylation of certain genes, such as ZNF671 (29), ADAMTS1 (30) and CD36 (31), may alter their functions, including the regulation of the cell cycle and signal transduction pathways, as well as transcriptional inhibition. Kiyozumi et al. (32) demonstrated that indoleamine 2,3-dioxygenase 1 promoter hypomethylation is associated with poor prognosis in esophageal cancer. Therefore, the accurate detection of methylated genes is likely to improve the clinical management of LUSC.

Studies have previously identified DMGs in LUSC (33-35); however, not all genes can be transcriptionally expressed. Therefore, as DMGs are not able to precisely demonstrate the relevance between genetic methylation and oncogenesis, MDGs are considered to be more representative (14,36). In the present study, high-throughput bioinformatics tools were used to identify and analyze MDGs associated with the prognosis of LUSC. Data extracted from TCGA were analyzed using packages from R, including edge, limma and MethylMix, and 30 LUSC-associated MDGs were derived. To improve the understanding of the functional pathways involving these MDGs, significant pathways were visualized using the ConsensusPathDB and Cytoscape.js library in the present study. The results identified 3 primary pathways: 'generic transcription'; 'RNA polymerase II transcription'; and 'gene expression (transcription)', which were affected by MDG interactions at a functional level. In other words, differential methylation of specific MDGs is able affect their expression and transcription. Furthermore, considering that not all MDGs are significantly associated with cancer prognosis, Kaplan-Meier survival and joint survival analyses were conducted using the R survival package, yielding 5 candidate prognosis-associated MDGs: PMPCAP1; SOWAHC; ZNF454; LINC00668; and ADH7 (P<0.05). The common genes showing significance in survival and joint survival analyses were chosen as the 3 hub MDGs (PMPCAP1, SOWAHC and ZNF454), which were identified to function as potential independent prognosis-associated markers for LUSC. The hub MDGs were determined by analyzing the association between hyper- or hypomethylation and survival, but also by integrating the degree of methylation and the expression of MDGs with survival.

PMPCAP1, a pseudogene of PMPCA1, is located on chromosome 4q22.1. To the best of our knowledge, the function of PMPCAP1 has not been investigated thus far. However, a number of studies have suggested that the functions of certain pseudogenes differ from those of normal homologous genes, but that the expression of associated non-coding (nc)RNAs plays an important regulatory role in the development of certain diseases (37-39). For example, PTENP expression may generate ncRNAs that competitively inhibit the function of PTEN, a known tumor-suppressor gene, and therefore inhibit cancer cell proliferation (40). Pseudogenes may also affect oncogenesis through epigenetic changes. The present study indicated that PMPCAP1 was hypomethylated in LUSC compared with normal tissues, which was associated with high expression levels, and ultimately, poor prognosis (P=0.041). Therefore, it may be speculated that the hypomethylation of PMPCA1 in tumor tissue (and the subsequent increase in RNA expression) is an indicator of poor clinical outcome in patients with LUSC, though further investigation is required to confirm this hypothesis.
SOWAHC, also known as ankyrin repeat domain (ANKRD) 57, is a protein-coding gene. The principle biological function of the ANKRDFamily is to mediate interactions between proteins (41). Takahashi et al (42) identified that ANKRD1 was overexpressed in EGFR-TKIs-resistant NSCLC with EGFR mutation, and that by inhibiting ANKRD1 expression, resistant cells were re-sensitized to afatinib and osimertinib. Lei et al (43) also demonstrated that ANKRD1 regulated apoptosis in ovarian cancer cells and functioned as a potential target to increase sensitivity to chemotherapy in ovarian cancer. In addition, the prognostic value of SOWAHC has been confirmed in bladder cancer (41); however, its value in LUSC has not been elucidated thus far, to the best of our knowledge. In the present study, the results of the correlation analysis indicated that the methylation of SOWAHC was negatively associated with its expression, commonly presenting as hypomethylation and high expression. Joint survival analysis revealed a significant association between the combined methylation and expression data and survival (P=0.028), suggesting that hypomethylation and high expression levels denoted improved prognosis in LUSC. Therefore, SOWAHC may be a potential biomarker of LUSC.

ZNF454 is a protein-coding gene, which expresses a protein measuring 522 amino acids. The ZNF454 protein is...
primarily involved in functional pathways associated with gene expression and transcription, namely ‘DNA binding’, ‘DNA-binding transcription factor activity by RNA polymerase II-specific’, ‘nucleic acid binding’ and ‘metal ion binding’. To the best of our knowledge, no published studies of ZNF454 were available until now, while a number of other members of the ZNF family have been investigated. Previous studies have suggested that, as the largest family of transcription factors in humans, ZNFs serve numerous important roles, and were recently confirmed as potential tumor suppressors (44). For example, through ZNF454 promoter methylation-associated deactivation, ZNF454 inhibited tumor proliferation in colorectal cancer via the PI3K/AKT and MAPK/ERK signaling pathways (45). In the present study, ZNF454 was generally hypermethylated and expressed to a low degree, which was determined to be an indication of poor prognosis. By contrast, ZNF454 was hypermethylated and expressed to a low degree, which was associated with improved prognosis. In addition, specific sites of aberrant methylation were investigated to identify more precise targets for clinical application. Although the results require further experimental validation, the present study provides diagnostic, therapeutic and prognostic value for patients with LUSC, and may guide future clinical applications to some extent.

Acknowledgements

The results published here are in whole based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

Funding

The present study was supported by the Key Technology Research and Development Program of Shandong Province (grant no. GG201710060039).

Availability of data and materials

The authors declare that the initial data are available from the TCGA database, and that all data generated and analyzed during the present study are included in this published article.

Authors’ contributions

QZhu, JW and JL conceived and designed the study. JW and QZha performed the data analysis and generated the figures. FW and LF contributed analysis tools. QZhu performed the data analysis and generated the figures. JW and JL conceived and designed the study. JW and QZha performed the data analysis and validated the results. The authors declare that they have no competing interests.

Ethics approval and consent to participate

Not applicable.

Table II. Correlation between key methylation sites and expression of PMPCA P1, SOWAH C and ZNF454.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Methylation site</th>
<th>Cor</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMPCA P1</td>
<td>cg06551022</td>
<td>-0.662</td>
<td>5.84x10^-49</td>
</tr>
<tr>
<td></td>
<td>cg14777507</td>
<td>-0.642</td>
<td>2.84x10^-45</td>
</tr>
<tr>
<td></td>
<td>cg07794230</td>
<td>-0.586</td>
<td>4.46x10^-36</td>
</tr>
<tr>
<td></td>
<td>cg10697010</td>
<td>-0.563</td>
<td>7.21x10^-53</td>
</tr>
<tr>
<td></td>
<td>cg16254375</td>
<td>-0.541</td>
<td>5.27x10^-30</td>
</tr>
<tr>
<td>SOWAHC</td>
<td>cg19399885</td>
<td>-0.578</td>
<td>6.34x10^-35</td>
</tr>
<tr>
<td>ZNF454</td>
<td>cg17840719</td>
<td>-0.698</td>
<td>2.83x10^-56</td>
</tr>
<tr>
<td></td>
<td>cg16536329</td>
<td>-0.633</td>
<td>1.44x10^-43</td>
</tr>
<tr>
<td></td>
<td>cg23037403</td>
<td>-0.595</td>
<td>1.56x10^-57</td>
</tr>
<tr>
<td></td>
<td>cg20778451</td>
<td>-0.592</td>
<td>5.50x10^-37</td>
</tr>
<tr>
<td></td>
<td>cg24843380</td>
<td>-0.586</td>
<td>4.09x10^-36</td>
</tr>
<tr>
<td></td>
<td>cg03234732</td>
<td>-0.573</td>
<td>3.03x10^-34</td>
</tr>
<tr>
<td></td>
<td>cg02165355</td>
<td>-0.561</td>
<td>1.38x10^-32</td>
</tr>
<tr>
<td></td>
<td>cg10575261</td>
<td>-0.555</td>
<td>6.87x10^-32</td>
</tr>
<tr>
<td></td>
<td>cg10902717</td>
<td>-0.503</td>
<td>1.47x10^-25</td>
</tr>
<tr>
<td></td>
<td>cg05461386</td>
<td>0.531</td>
<td>8.16x10^-29</td>
</tr>
</tbody>
</table>

Only the key methylation sites of PMPCA P1, SOWAHC and ZNF454 with both P<0.05 and |Cor|>0.5 are listed. Cor, correlation coefficient; |Cor|, the absolute value of correlation coefficient. PMPCA P1, peptidase, mitochondrial processing a subunit pseudogene 1; SOWAHC, sosondowah ankyrin repeat domain family member C; ZNF454: Zinc finger protein 454.
References


