Identification of hub-methylated differentially expressed genes in lung adenocarcinoma and immunotherapy resistance

Xiaodie Ye
Translational Medicine Center, the Second Affiliated Hospital of Guangzhou Medical University

Weiguo Cai
The Second Affiliated Hospital of Guangzhou Medical University

Jian Zhang
Translational Medicine Center, the Second Affiliated Hospital of Guangzhou Medical University

Zhiru Zhang
Translational Medicine Center, the Second Affiliated Hospital of Guangzhou Medical University

Cheng Zhi
Translational Medicine Center, the Second Affiliated Hospital of Guangzhou Medical University

Manting Liu
Translational Medicine Center, the Second Affiliated Hospital of Guangzhou Medical University

Zhenfeng Zhang (zhangzhf@gzhmu.edu.cn)
Translational Medicine Center, the Second Affiliated Hospital of Guangzhou Medical University

Research Article

Keywords: lung adenocarcinoma, immunotherapy resistant, transcriptome, tumor immunophenotype, DNA methylation

Posted Date: November 30th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2317392/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Incidence and mortality of lung adenocarcinoma are high, and the epigenetic mechanism of DNA methylation has a critical effect on LUAD at all stages. Our work used GEO and TCGA databases to identify differentially methylated genes (DMGs) in LUAD to explore how DNA methylation works in immunotherapy resistance. Candidate pathogenic genes were highly correlated to hub-methylated differentially expressed genes (SLC2A1, HLF, FAM83A, SCARF1, C2orf40). Core genes were correlated with the pathways regulating cancer development. Using the TISIDB database to estimate immune cell infiltration and immune factor levels, a relation of tumor gene levels with immune infiltration suggested the possible effect of core genes on regulating tumor microenvironment (TME). The functional pathways and key genes were analyzed via GESA and GEVA (GO, KEGG) to identify functionally enriched pathways and key genes. According to CMap, there was a significantly negative correlation between drug expression profiles (BX-912, JAK3-inhibitor-VI, panobinostat, purvalanol-A, and scriptaid) and differentially expressed genes. Therefore, we hypothesized that these drugs could enhance LUAD anti-tumor therapy.

Background

Lung cancer (LC) has high morbidity and mortality among cancers, (1) dominated by the pathological subtype of lung adenocarcinoma (LUAD) (2). In recent years, due to the emergence of immunotherapy, immune checkpoint inhibitors (ICIs) have demonstrated superior advantages to the standard chemotherapy in diverse lines of treatment for LUAD. The anti-PD-1/PD-L1 (nivolumab, pembrolizumab, durvalumab, and atezolizumab), as well as anti-CTLA-4 (tremelimumab and ipilimumab) antibodies, can induce anti-tumor immunity in the host (3, 4). Immune checkpoint drug resistance has recently reduced the cure rate for LUAD, which hampered anti-tumor immunity (5). Thus, it is crucial to investigate the resistance mechanisms in ICI therapy. In addition, the constantly evolving immune/cancer cell interactions make resistance mechanisms more complex (6).

DNA methylation, an epigenetic mechanism that plays a role in all stages of lung cancer (6), implicates epigenetic modulation in precision immunotherapy (7). According to previous research, alterations of genomic methylation can offset the effect of a high mutation burden while enhancing immunotherapeutic tolerance (7). Anti-PD-1 immunotherapy can improve NSCLC by identifying differentially methylated regions and developing biomarkers (8). It may also benefit patients with epigenetic treatment (9). DNMT inhibitors decreased the exhaustion of T-cells through de-novo DNA methylation induced by DNMT3a (10). Thus, targeting DNA methyltransferases by combining chemotherapy and immunotherapy has shown great promise (11). In addition, LUAD immunotherapy may be developed by targeting differentially methylated regions (12).

Our work utilized GEO- and TCGA-derived data to identify differentially methylated genes (DMGs) in LUAD to explore DNA methylation's role in resistance to immunotherapy and find potential drugs for immunotherapy resistance. In addition, the Connectivity Map (CMap), formulated by the Broad Institute
and Genomics of Drug Sensitivity in Cancer (GDSC), was adopted to identify possible therapeutic agents based on the LUAD-specific gene signatures to treat LUAD patients. TISIDB was used to explore the relation of the tumor immune microenvironment (TIME) with key genes, and the CancerSEA database provided relationships between functional states and key genes.

**Materials And Methods**

**Sample datasets and clinical profiles**

**Study Datasets and Design**

Our work examined mRNA profiling data of 594 LUAD patients (594 normal and 535 cancer) from the TCGA (https://tcga-data.nci.nih.gov/tcga/). In addition, a matrix file series of microarray datasets (GEO accession numbers: GSE50081 and GSE30219) and corresponding survival and clinical information were obtained from the GEO database (http://www.ncbi.nlm.nih.gov/geo/). The TCGA provided DNA methylation files for LUAD patients (32 normal and 460 cancer groups). This study employed the R software (3.5.2) ChAMP package to analyze differential methylation; differential methylation loci were detected based on (a) |logFC|>0.3 and (b) p > 0.05. In addition, this work carried out differential expression through the limma package of R software (https://bioconductor.org/packages/limma/), with |logFC|>1 and p < 0.05 as the thresholds. In total, 804 patients were enrolled in the AUC analysis, including 127 patients from GSE50081, 83 from GSE30219, 492 from GSE72094, and 594 from TCGA-LUAD.

DNA methylation and transcriptome analysis were combined to control gene expression through transcriptional silencing. Moreover, this work classified EPN-PF as two subtypes based on methylation levels in the CpG island co-expression networks, which displayed markedly distinct prognostic outcomes. As a result, GSE66354 and GSE114523 were adopted to identify hub-methylated genes. In addition, this work utilized the R package Limma for identifying differentially expressed genes (DEGs) upon the |logFC|>1 and P < 0.05 thresholds. This study conducted differential methylation gene (DMG) analysis with ChAMP R package, with screening criteria being β-value < 0.2 in one group while > 0.3 for the second group. Our next step was identifying DMGs associated with EPN-PF, including both methylated-associated genes with high and low expression. This study also adopted the Metascape database (www.metascape.org) to conduct Gene Ontology (GO) functional annotation and Kyoto Genome Encyclopedia (KEGG) analysis. Subsequently, these genes were subject to protein-protein interaction (PPI) network construction. Statistical significance was determined by the minimal overlap $\geq 3$ and $p \leq 0.01$.

**Functional Enrichment Analysis**

Metascape was used to analyze GO and KEGG pathways based on gene names. The enrichment cut-off was set to a min overlap $\geq 3$ & $p \leq 0.001$. 
**Immunocytes Infiltration Analysis**

Our work calculated 22 tumor-infiltrating immune cells (TIICs) abundances based on the CIBERSORT deconvolution algorithm. By normalizing gene expression profiles in GC, TIIC abundances could be distinguished while analyzing its relation with significantly enriched pathways using CIBERSORT (https://cibersortx.stanford.edu/).

**Gene Set Variation Analysis (Gsva)**

Molecular Signatures Database v7.0 was used to download the set of interest genes. GSVA was performed to score the genes of interest and assess the potential biological functions of the Biological Variations of diverse patient samples.

**Gene Set Enrichment Analysis (Gsea)**

Our work performed GSEA to identify DEGs in high compared with low expression groups based on the LUAD datasets downloaded from TCGA. The gene sets ranged from 15 to 500 in size, which was adopted for 10000 permutations, and those showing a false discovery rate (FDR) ≥ 0.25 and p < 0.05 indicated significant enrichment. Finally, the GO and KEGG enrichment analyses were used to generate plots displaying the relationship between genes and KEGG pathways enriched.

**Analysis Based On The Tisidb Database**

TISIDB database (http://cis.hku.hk/TISIDB) covers 988 anti-cancer immune-related genes (IRGs), molecular profiling, high-throughput sequencing (HTS) technology, para-carcinoma multi-omics data, and different immunological data sources collected in 7 free databases. It allows the analysis of core genes’ relations with immunomodulators, lymphocytes, chemokine, and immunophenotyping.

**Bioinformatics Analysis Based On Connectivity Map**

Connectivity Map (CMap, https://www.broadinstitute.org/cmap/), a gene-drug interaction database, was adopted to predict potential cancer treatments. CMap predicted differential mRNA expression (upregulated and downregulated) for targeted therapy agents.

**Analysis Based On Cancersea**

Our work utilized CancerSEA (http://biocc.hrbmu.edu.cn/CancerSEA/) to comprehensively explore the cancer cellular function database at a single-cell level to analyze hub-genes. COMP's t-SEN distribution within lung adenocarcinoma sequencing data was analyzed using functional associations of diverse cell
subsets at the single-cell level. Different colors represented the input gene level. The present work conducted a correlation analysis of key genes with activities, metastasis, angiogenesis, and quiescence to shed more light on these gene functions. Clustering phenographs identified cell types, with labels assigned based on gene expression and visualized using t-SNE. Gene expression distinguished the different clusters in 2D t-SNE plots.

Tide Algorithm

TCGA (https://portal.gdc.com) was employed to obtain level-3 RNA-seq and related clinical data for xx, and the TIDE algorithm was adopted to predict the possible ICB responses.

Cell Culture And Westernblot (Wb)

This work utilized NCI-H23 (H23), Whole cells were cultured in RPMI-1640 (BI) with 1% penicillin-streptomycin (PS, Invitrogen) and 10% FBS (Gibco). After adding azacitidine and rinsing the cells using pre-chilled PBS, RIPA buffer was added for cell lysis. In the WB assay, protein lysates (30 µg) were separated using 12% SDS-PAGE, followed by transfer to PVDF membranes (Millipore). After blocking in PBS that contained 5% defatted milk at ambient temperature, membranes were incubated further with anti-VEGFA and anti-β-tubulin (mMb, 1:1000; Cell Signaling) primary antibodies overnight at 4°C. After washing thrice with TBS-T buffer (10 min each), membranes were incubated with secondary antibodies labeled with Goats-radish peroxidase (1:5000, Beijing Dingguo). Then, membranes were rinsed with TBS-T buffer four times, and the enhanced chemiluminescence kit (Western Lightning™, Western blot Chemiluminescence reagent, Thermo Fisher Scientific) was adopted for immunoreactive protein detection following specific protocols. ImageJ analysis system was utilized to analyze the density.

Statistical analysis

R (version 3.6) was employed for statistical analysis. Each statistical test was two-sided, with p < 0.05 indicating statistical significance.

Ethics Approval

This study did not require ethics approval.

Results

LUAD-related Core Gene Identification

A total of 2,018 differential genes were screened, including 1,303 downregulated and 715 upregulated genes (Fig. 2A-B). The methylation 450K data of 492 LUAD patients from the TCGA database were
downloaded. Based on a differential methylation analysis using ChAMP, 3744 differentially methylated probes (2822 upregulated and 922 downregulated) were identified (Supplementary Fig. 1). Thus, 78 genes (hypermethylation and upregulation) were extracted from the intersection of the differentially expressed and methylated genes (DMG). The differentially expressed genes were mapped into the differentially methylated genes, including 78 with hypermethylation and downregulation, whereas 13 with hypomethylation and upregulation (Fig. 2C-D). Metascape (http://metascape.org) was adopted for GO and KEGG analysis. Gene Set Enrichment Analyses showed that candidate genes were involved in lung cancer by regulating tube morphogenesis and DNA-binding transcription factor binding specifically to RNA polymerase II and pathways in cancer (Fig. 3A-B). Furthermore, AUC values analysis based on differential gene expression was performed in three datasets (TCGA-LUAD, GSE50081, and GSE30219). The ROC analysis exhibited a good predictive value of nine genes (SLC2A1, HLF, TUBB6, FAM83A, SCARF1, C2orf40, LIMCH1, NTM, and PPP1R16B) with the AUC > 60% in more than two datasets (Fig. 3C-D). There were 579 LUAD-associated genes acquired based on the GeneCards database (https://www.genecards.org/). The numbers of pathogenic genes highly related to 5 key genes (SLC2A1, HLF, FAM83A, C2orf40, and SCARF1) were counted (p < 0.05 & |cor| > 0.3), and correlation analysis between gene expression levels and candidate pathogenic genes was performed (Fig. 3E). Workflow of data preparation, processing and analysis in this study was listed (Fig. 1).

**Tisidb Database Analysis**

The tumor microenvironment (TME) represents the complicated non-cancer cell environment, which mainly contains cancer and immune cells (including extracellular matrix ECM, cancer-associated fibroblasts CAFs), carcinoma cells, inflammatory factors, and growth factors. TIME (Tumor immune microenvironment) status, including TIICs, determines Patient Diagnosis, Therapy, and Outcome. Relation of tumor gene level with immune infiltration via datasets showed the potential molecular mechanism of hub-genes during LUAD progression.

Heatmap was prepared by estimating the distribution and level of immune cell infiltration (Fig. 4A) to display the relation between LC and TIICs. Blue and red showed positive and negative relations, respectively, with a darker color indicating a strong relation (i.e., closer to 1 or −1) (Fig. 4B). Relative to the normal group, immune factors (T cells follicular helper, Plasma cells, B cells naïve) were significantly overexpressed in the cancer group, while NK cells resting and T cells CD4 memory resting were low-expressed (Fig. 4C).

Further exploration of the relations of immune cells with core genes showed that immune cell infiltration was related to several gene levels in different immune cells, such as SLC2A1 was positively correlated to macrophages M1 and T cells CD4 memory activated, negatively correlated to mast cells resting and monocytes. C2orf40 was positively correlated to mast cells resting and monocytes but negatively correlated to T cells CD4 memory activated and macrophages M1. SCARF1 was positively associated with mast cells resting and monocytes while negatively correlated with T cells CD4 memory activated and
plasma cells (Fig. 5A). Correlation between different immune factor levels, including immune regulators, chemokines, and chemokine receptors, were obtained using TISIDB database (Fig. 5B-G). Analysis revealed that the core genes were closely related to immune cell infiltration and play a significant role in regulating the tumor microenvironment. Immunostimulators, immunoinhibitors, and MHC molecules were also related to core gene expression in LUAD based on TISIDB analysis, except for Tumor Infiltrating Lymphocytes (TILs) (Fig. 5C-G). Multi-gene correlation between gene expression and the immune score showed that SCARF1 was positively associated with endothelial cells ($r = 0.46, p < 0.001$) (Supplementary Fig. 1). Immune response score distributions across diverse groups are presented in Supplementary Fig. 2 FAM83A and SLC2A1 expression were negatively correlated with immune response scores, while HLF expression was positively correlated.

**Functional Enrichment Analysis And Transcriptional Regulatory Network**

According to the involved genes and signaling pathways, the signaling pathways correlating with which core genes regulated lung cancer progression were investigated. Data based on GSVA presents that SLC2A1 gene expression was enriched remarkably in bile acid metabolism and KRAS signaling pathway, p53 pathway, wnt/beta-catenin signaling related to HLF, bile acid metabolism, interferon alpha response related to FAM83A, NOTCH signaling, HEME metabolism related to C2orf40, TGF-BETA signaling and APICAL junction related to SCARF1(Fig. 6A-E). Sample Level Enrichment Analysis of KEGG Pathways via GESA identifies distinct involvement of the key genes and Functionally Enriched Pathway, which are presented in Fig. 6F-J.

The Transcription factors (TFs) involved were predicted based on data from the Cistrome DB database, including 84 TFs related to SLC2A1, 98 TFs related to HLF, 80 TFs related to FAM83A, 103 TFs related to C2orf40, and 73 TFs related to SCARF1 to explore transcriptional regulatory network related to the five key genes further. The transcriptional regulatory network related to LUAD core genes was visualized using Cytoscape (Supplementary Fig. 3A). The five key genes shared a common set of transcription factors, including CTCF, MYC, BRD4, MAX, EP300, POLR2A, AR, ESR1, JUND, and MED1.

**Functional States**

CancerSEA characterizes various cancer cell functions to provide the atlas regarding cancer functional states at the single-cell level and link them with IncRNAs/protein-coding genes (PCGs) to explore the mechanism underlying different functions of tumor cells. Our work also presented the atlas showing core genes and cancer functional states at the single-cell level for querying related gene functional states in the LUAD single cells (Fig. 4D). C2orf40 was not searched in the database, and SLC2A1 was markedly associated with proliferation and stemness status; Gene expression does not distinguish clusters. HLF was negatively correlated with angiogenesis ($r = 0.44$). In addition, FAM83A was significantly related to quiescence, metastasis, hypoxia, angiogenesis, differentiation, and EMT. However, SCARF1 was
significantly related to quiescence, angiogenesis, metastasis, differentiation, inflammation, DNA repair, and proliferation. HLF, FAM83A, and SCARF1 expressions distinguish the 2 clusters projected onto t-SNE plots. From the western blot result, azacitidine significantly reduced the expression of VEGFA in H23 (Fig. 4E).

**Potential Drugs For Cancer Treatment**

CMap (http://www.broad.mit.edu/cmap/) can predict possible drugs for cancer treatment according to differentially expressed genes (Upregulated and downregulated genes). CMap predicted differential mRNA expression (upregulated and downregulated) for targeted therapy agents. There was a negative correlation between drug expression profiles (BX-912, JAK3-inhibitor-VI, panobinostat, purvalanol-a, and scriptaid) and lung cancer (Supplementary Fig. 3B-F). Thus, these drugs may be effective as therapeutic interventions for LUAD.

**Discussion**

The present work combined methylations with expression profiles in lung adenocarcinoma cases to identify genes contributing to LUAD onset and development. Methylated differentially expressed genes were mainly enriched in regulating tube morphogenesis, DNA-binding transcription factor binding specifically to RNA polymerase II and pathways in cancer. Survival analysis was conducted in three datasets (TCGA-LUAD, GSE50081, and GSE30219) to evaluate the prognostic impact of these differentially methylated CpG sites. The ROC analysis exhibited a good predictive value of 9 genes (SLC2A1, HLF, TUBB6, FAM83A, SCARF1, C2orf40, LIMCH1, NTM, and PPP1R16B) with AUC > 60% in more than two datasets. Relations of gene levels with candidate pathogenic genes were conducted to select core gene sets, and SLC2A1, HLF, FAM83A, C2orf40, and SCARF1 were found to be significantly correlated to lung cancer development. Transcription factors (TFs) involved were predicted based on data from the Cistrome DB database to explore the transcriptional regulatory network related to the five key genes. For analyzing key gene functions in TME of LUAD, this work discovered that immune cell infiltration and immunity Factor were related to several gene levels in different immune cells. For instance, SLC2A1 was positively correlated with macrophages M1 and T cells CD4 memory activated and negatively correlated to mast cells resting and monocytes. C2orf40 was positively correlated to mast cells resting and monocytes, whereas negatively correlated to T cells CD4 memory activated and macrophages M1. SCARF1 was positively associated with mast cells resting and monocytes and negatively correlated with T cells CD4 memory activated and plasma cells (Fig. 5A). To understand the function states of key genes in LUAD, our study showed that HLF, FAM83A, and SCARF1 were associated with angiogenesis. According to 150 differentially mRNA expressions, there was a significantly negative correlation between expression profiles for drugs (BX-912, JAK3-inhibitor-VI, panobinostat, purvalanol-a, and scriptaid) and differentially expressed genes. Thus, these drugs may have clinical treatment benefits for lung cancer.
Solute carrier family 2 member 1 (SLC2A1) is responsible for encoding the glucose transporter (GLUT), which has a critical effect on cancer cell glycolysis and physiological metabolism in the human body (13–16). Further, it was found that SLC2A1 propels tumor progression oncogene in lung cancer (17), consistent with our findings. HLF (Hepatic leukemia factor) is a specific marker gene for HSC, a circadian rhythm regulator (18), and a tumor suppressor gene in NSCLC cells through PPAR/NF-κb signaling (19). Therefore, our research showed that the p53 and wnt/beta-catenin signaling pathways were related to HLF. FAM83A (family with sequence similarity 83 member A) expression increased in LUAD, related to the late LUAD stage and dismal patient survival (20, 21). C2orf40 (Chromosome 2 Open Reading Frame 40, referred to as esophageal cancer-related gene 4 (ECRG4) or augurin), represents one of the tumor suppressors showing hypermethylation in different cancers, such as colorectal cancer (CRC) (22), esophageal cancer (EC), breast cancer (BC) and glioma. Our research showed that C2orf40 was downregulated in LUAD; however, it was not found in CancerSEA, which plots tumor cells. Thus we suggested that C2orf40 plays a tumor suppressor role in LUAD. However, this speculation requires further validation. SCARF1, Scavenger Receptor Class F Member 1, is a scavenger receptor regulating chemically-modified lipoprotein endocytosis, controlling tissue homeostasis, and clearing the apoptotic cells (23). SCARF1 has a specific mediating effect on selectively absorbing acetylated low-density lipoprotein (Ac-LDL) in the endothelial cells (24). A previous study indicated that high methylation of SCARF1 promoters affected NSCLC tumorigenesis (24). Additionally, methylation and expression in SCARF1 and C2orf40 predicted better overall survival (OS) for LC (24).

In this study, HLF, FAM83A, and SCARF1 expression were positively correlated with tumor angiogenesis. Further, SCARF1 had a high positive correlation with endothelial cells, which help orchestrate blood vessels’ formation (25). The methylation of HLF, FAM83A, and SCARF1 may be critical in controlling angiogenesis and tumor growth in LUAD. Thus, H23, was cultured with 10 µmol/mL azacytidine (DNA methylase inhibitors). Then, the VEGFA expression in H23 was downregulated, indicating that it is crucial to explore genomic DNA (gDNA) methylation status to understand angiogenesis further (Fig .5E).

BX-912 (3-phosphoinositide-dependent protein kinase 1 (PDK1)) directly and selectively inhibits PDK1 by competing against ATP. BX-912 inactivates PDK1/Akt pathway in cancer cells and suppresses anchorage-dependent cancer cell growth in the cell culture or promotes apoptosis (26). A previous study found that BX-912 inhibited parasite proliferation and is analyzed as the novel therapeutic target candidate to resist malaria (27). In addition, it is a new finding that it may exhibit significant therapeutic activity in LUAD. As the mutant-specific inhibitor against the epidermal growth factor receptor, the JAK3 inhibitor carries the T790M gatekeeper mutation (28). Our study provided similar insights regarding LUAD patients. Purvalanol A (Cdc2 selective inhibitor) that markedly suppresses Cdc2 kinase (29) can suppress the increased cyclins D and E levels in serum-mediated progress of the G1 phase. It has been reported to inhibit 60 human cancer cell lines in the NCI panel and efficiently promote taxol sensitivity in NCI-H1299 cells while inhibiting cell colony forming and proliferation (30). The CDK inhibitor can rapidly and extensively promote human neutrophil apoptosis and partially abolish the anti-apoptin (GM-CSF) effect (30). Our results provide proof of concept for purvalanol A as a therapeutic strategy in LUAD and
speculate that it can be a therapeutic strategy in LUAD by inhibiting Cdc2 kinase activity and inducing neutrophils in TME.

Panobinostat (Farydak, LBH589, the oral histone deacetylase inhibitor) is used for treating HIV and multiple myeloma (https://pubchem.ncbi.nlm.nih.gov). A recent study showed that LBH589 combined with osimertinib showed a synergistic effect on decreasing diverse osimertinib-resistant cell line survival in EGFR-mutation Lung cancer (31). In addition, panobinostat combined with MICA/B antibody (32) enhanced olaparib efficacy in CD8+ T cell infiltration in the tumor while decreasing M2-like macrophage marker levels in ovarian cancer (33). PD-L1 expression was upregulated on MM cell surface after panobinostat with IFN-γ treatment (34). Thus, our study provides a basis for epigenetic therapy in lung cancer.

Scriptaid, Histone deacetylase inhibitors (HDACIs) can reduce anti-tumor responses in colorectal cancer, ovarian cancer cells, and advanced HCC (35–37). Scriptaid reverses several well-known Cancer-associated fibroblasts (CAFs) characteristics such as increased contractility, ECM upregulation, and TGF-β pathway activation. In addition, scriptaid damages the tumor-promoting features of CAFs both in-vivo and in-vitro; therefore, HDAC inhibitors like scriptaid are the possible molecular treatments that target stromal and tumor cells in solid tumors TME (38). Thus, scriptaid may be an anti-cancer strategy that could reverse the immunosuppressive tumor microenvironment by activating CAF.

Despite the promising findings, this study has several limitations. Firstly, hub-gene levels and prognostic role in TME of LUAD warrant more investigation. Secondly, the prognosis-predicting ability of hub-genes was not studied in patients in immunotherapy. Thus, future research must explore the underlying mechanism between the hub-genes expression and response in LUAD immunotherapy.

**Conclusion**

Our study found core genes that alter DNA methylation and critically affect lung cancer. mRNA expression of core genes (SLC2A1, HLF, FAM83A, C2orf40, and SCARF1) was related to the expression of TIICs and immune factors. HLF, FAM83A, and SCARF1 may be potential factors in anti-angiogenic therapy, and DNMTi changes angiogenic. There was a significantly negative correlation between the expression profiles for drugs (BX-912, JAK3-inhibitor-VI, panobinostat, purvalanol-a, and scriptaid) and differential gene expression. Thus, these drugs can regulate the lung cancer microenvironment and benefit immunotherapy in lung cancer. Moreover, the results shed more light on LUAD treatment.

**Declarations**

**Acknowledgments**

NONE.

**Author Contributions**
Conceptualization, XY and WC; Formal analysis, XY and WC; Funding acquisition, XY and WC; Investigation, XY and WC; Methodology, XY and WC; Writing—original draft, XY and WC; Writing—review & editing, XY, WC; and ZZ All authors have read and agreed to the published version of the manuscript.

Funding

This research was supported by the National Natural Science Foundation of China (No. 81872069 to Z. Z.)

Conflicts of Interest

The authors declare no conflict of interest.

Ethics approval and consent to participate

Not applicable.

Availability of data and materials


References


22. Y. Z et al., Integrated bioinformatics identifies the dysregulation induced by aberrant gene methylation in colorectal carcinoma. 8, 521–530 (2021).


Figures
Figure 1

Workflow of data preparation, processing and analysis in this study.
Figure 2

Dentification of LUAD-related DMGs. A. Volcano plot for DEGs between two subtype of LUAD. B The top 10 up- and down-regulated DEGs analyzed from the dataset of TCGA. C-D. The Venn diagram showed 63 methylated-related highly expressed genes and 83 methylated-related lowly expressed genes. And hypermethylated and downregulated (C) and hypomethylated and upregulated (D) in expression.
Figure 3

GO term, KEGG pathway, and PPI analysis and the ROC analysis exhibited prediction value of the differentially expressed and methylated genes in three datasets (TCGA-LUAD, GSE50081 and GSE30219). A Bar plot of gene ontology biological process enrichment analysis. Gene-ontology analysis was performed using all DMGs. Functional enrichment analysis was performed by Metascape (http://metascape.org). Only terms with P<0.001 and with at least 3 enriched genes were considered as
significant. Color is proportional to their P values. (The significant GO terms and KEGG pathway of the methylated-related highly expressed genes.) B The significant modules of methylated-related differently expressed genes through PPI network analysis. C X-axis represents the differentially expressed and methylated genes and Y-axis represents the AUC values based on prognosis-related genes. D The AUC values of the differentially expressed and methylated gene in three datasets, in which the axis represents the AUC values in different datasets. E Performing correlation analysis between differentially expressed and methylated gene expression levels and candidate pathogenic genes, the numbers of pathogenic genes highly related to 5 key genes were counted (p < 0.05 & |cor| > 0.3).
Immune landscape of LUAD and An Atlas of Cancer single-cell Functional states and core genes. A. the distributions of Immune cell infiltration in lung healthy versus tumor immune cell populations. B. Heatmap of correlation of 22 types of immune cells. Red indicates a negative correlation, and blue indicates a positive correlation. The darker the color, the stronger the correlation (i.e., closer to 1 or −1). C. The levels of immune cell infiltration in lung versus tumor. D. an atlas of cancer single-cell functional states and core genes. E. The expression of VEGFA in H23 cells decreased significantly to azacitidine treatments.
Figure 5

Relations between TME (tumor-infiltrating lymphocytes and different immune factors) and expression of core genes. A. Pearson’s correlation analyses between core genes and the immune cells. B. Correlation analysis between core genes and chemokine. C. Correlation analysis between core genes and immunoinhibitor. D. Correlation analysis between core genes and immunostimulator. E. Correlation analysis between core genes and MHC. F. Correlation analysis between core genes and receptor. An atlas of cancer single-cell functional states and core genes. The expression of VEGFA in H23 cells decreased significantly to azacitidine treatments.

Figure 6

GSEA and GSVA for functional enrichment based on core genes. (A-E) Bar graphs showing results of GSVA respectively. All cascades, except gray, were significantly enriched (p < 0.05). (F-J) GSEA plot
showing the signaling pathways correlates with which core genes regulated in lung cancer progression across GSEA datasets. KEGG pathways enriched from core genes.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- cancerseaPlotDatacorrbubble.csv.xlsx
- additionalfile.pdf
- SFig3.jpg