


RESEARCH

Open Access



Identification and validation of smoking-related genes in lung adenocarcinoma using an in vitro carcinogenesis model and bioinformatics analysis

Jin Wang^{1,2†}, Tao Chen^{1,2†}, Xiaofan Yu^{1,2}, Nan OUYang^{1,2}, Lirong Tan^{1,2}, Beibei Jia^{1,2}, Jian Tong^{1,2} and Jianxiang Li^{1,2*} 

Abstract

Background: Lung cancer is one of the most common carcinomas in the world, and lung adenocarcinoma (LUAD) is the most lethal and most common subtype of lung cancer. Cigarette smoking is the most leading risk factor of lung cancer, but it is still unclear how normal lung cells become cancerous in cigarette smokers. This study aims to identify potential smoking-related biomarkers associated with the progression and prognosis of LUAD, as well as their regulation mechanism using an in vitro carcinogenesis model and bioinformatics analysis.

Results: Based on the integration analysis of four Gene Expression Omnibus (GEO) datasets and our mRNA sequencing analysis, 2 up-regulated and 11 down-regulated genes were identified in both S30 cells and LUAD. By analyzing the LUAD dataset in The Cancer Gene Analysis (TCGA) database, 3 of the 13 genes, viz., glycoprotein C (*GYPC*), *NME1* and *SLIT2*, were found to be significantly correlated with LUAD patients' smoking history. The expression levels of *GYPC*, *NME1* and *SLIT2* in S30 cells and lung cancer cell lines were validated by quantitative PCR, immunofluorescence, and western blot assays. Besides, these three genes are associated with tumor invasion depth, and elevated expression of *NME1* was correlated with lymph node metastasis. The enrichment analysis suggested that these genes were highly correlated to tumorigenesis and metastasis-related biological processes and pathways. Moreover, the increased expression levels of *GYPC* and *SLIT2*, as well as decreased expression of *NME1* were associated with a favorable prognosis in LUAD patients. Furthermore, based on the multi-omics data in the TCGA database, these genes were found to be regulated by DNA methylation.

Conclusion: In conclusion, our observations indicated that the differential expression of *GYPC*, *NME1* and *SLIT2* may be regulated by DNA methylation, and they are associated with cigarette smoke-induced LUAD, as well as serve as prognostic factors in LUAD patients.

Keywords: Cigarette smoke, Lung adenocarcinoma (LUAD), DNA methylation, prognosis, *GYPC*, *NME1*, *SLIT2*

*Correspondence: aljxcr@suda.edu.cn

†Jin Wang and Tao Chen contributed equally to this work

¹ Department of Toxicology, School of Public Health, Medical College of Soochow University, Renai Road, Suzhou 215123, China

Full list of author information is available at the end of the article

Background

Lung cancer is one of the most common carcinomas in the world. In 2018, the number of patients newly diagnosed with lung cancer across the globe was 2.09 million,



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

and around 1.76 million patients will die from the disease [1]. Although early diagnosis and treatment of lung cancer have made significant progress, the 5-year relative overall survival (OS) is less than 20% [2]. Lung adenocarcinoma (LUAD) is the most common subtype of non-small cell lung cancer (NSCLC), and NSCLC accounts for approximately 85% of all lung cancer cases [3]. There is a significant and positive correlation between cigarette smoke and lung cancer, and the risk of developing lung cancer in smokers is nearly 10 times higher than that in non-smokers [4, 5]. Nonetheless, it is still unclear how normal lung cells become cancerous in cigarette smokers.

The development of high-throughput sequencing technology has made it possible to identify changes in single bases within the coding sequences of specific genes during lung tumorigenesis. There are plenty of publicly available cancer multi-omic data that we can obtain free from The Cancer Gene Atlas (TCGA; <http://cancergenome.nih.gov/>) and Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>). A meticulous and thorough analysis of these data can identify genes and signaling pathways crucial to lung cancer, which will help for a better understanding of the mechanisms of cancer occurrence and development.

Based on the gene expression profiles, recent studies have identified several genes associated with lung cancer. Spira et al observed that *CYP1B1*, *NEK2* and *CENPF* were significantly correlated with LUAD [6]. Liu et al suggested that *EPHA4*, *FGFR2*, and *EGFR* may be strongly associated with the development and progression of smoking-related LUAD [7]. Landi et al demonstrated that elevated mRNA levels of *NEK2* and *TTK* have the potential to increase the risk of mortality from smoking-related LUAD [8]. Also, numerous genomic and transcriptional alterations in LUAD appeared to be associated with the patient's smoking history [9]. However, there is still a shortage of reliable biomarkers for smoking-related LUAD.

In this study, we aimed to identify novel biomarkers for LUAD in smokers. The workflow of our study is presented in Fig. 1. An in vitro carcinogenesis model was established by exposing BEAS-2B cells to cigarette smoke continuously for 30 passages (S30). In the present study, candidate genes were obtained by integrative analysis of differentially expressed genes (DEGs) according to databases and our mRNA sequencing data. Among these, the smoking-related genes observed in S30 cells and LUAD were further validated by quantitative PCR (qPCR), immunofluorescence assays (IF), and western blotting (WB), and analyzed for a possible association with cancer-related pathways and prognosis. Furthermore, the multi-omics data in the TCGA database were used to explore the regulatory mechanisms of these three genes.

Results

Differentially expressed genes in S30 cells and GEO datasets

Based on the high throughput analysis, a total of 753 differentially expressed genes (DEGs) were identified in cigarette smoke-induced transformed cells (S30) compared with unexposed BEAS-2B cells, including 273 up-regulated and 480 down-regulated genes (Fig. 2a, b). Besides, DEGs in LUAD tissues were screened out from four GEO datasets by differential expression analysis (Fig. 2c–f). Based on the integration analysis, 209 down-regulated genes and 25 up-regulated genes were identified in the GEO datasets (Fig. 2g and Additional file 1: Table S2). A total of 11 down-regulated and 2 up-regulated smoking-related genes were identified by taking the intersection of the DEGs extracted from S30 cells and GEO datasets (Fig. 2f).

Identification of smoking-related genes in lung cancer

Further analysis indicated that 7 of the 13 genes are associated with smoking history ($p < 0.05$) (Additional file 1: Table S3). Notably, the *NME1* expression level in current smokers and reformed smoker for ≤ 15 years was significantly higher compared to life-long non-smokers ($p < 0.01$). Conversely, the expression levels of *SLIT2* and *GYPC* in current smokers were markedly lower than those in life-long non-smokers ($p < 0.01$) (Table 1). As illustrated in Fig. 3, these findings were further validated in two GEO datasets (GSE13213 and GSE41271) with a smoking history.

Validation of mRNA and protein expression in S30 cells and lung cancer cell lines

The mRNA expression levels of *GYPC* and *SLIT2* were found to be dependent on smoke-exposure time and were significantly down-regulated in S30 cells (Fig. 4a, c). On the other hand, the *NME1* expression level was significantly up-regulated in S30 cells (Fig. 4b). Compared with normal BEAS-2B cells, the expression levels of *GYPC* and *SLIT2* in four human lung adenocarcinoma cell lines (PC9, A549, H1975 and H1299) were up-regulated, while *NME1* expression was down-regulated (Fig. 4d). The protein expression levels of *GYPC*, *NME1* and *SLIT2* were further validated. Immunofluorescence staining showed that *NME1* protein expression was increased in S30 cells compared with normal BEAS-2B cells, while *GYPC* and *SLIT2* expression was decreased (Fig. 4e–g). Western Blot analysis further confirmed the downregulation of *GYPC* and *SLIT2* and the upregulation of *NME1* in cigarette smoke-exposed S30 cells compared with unexposed BEAS-2B cells (Fig. 4h).

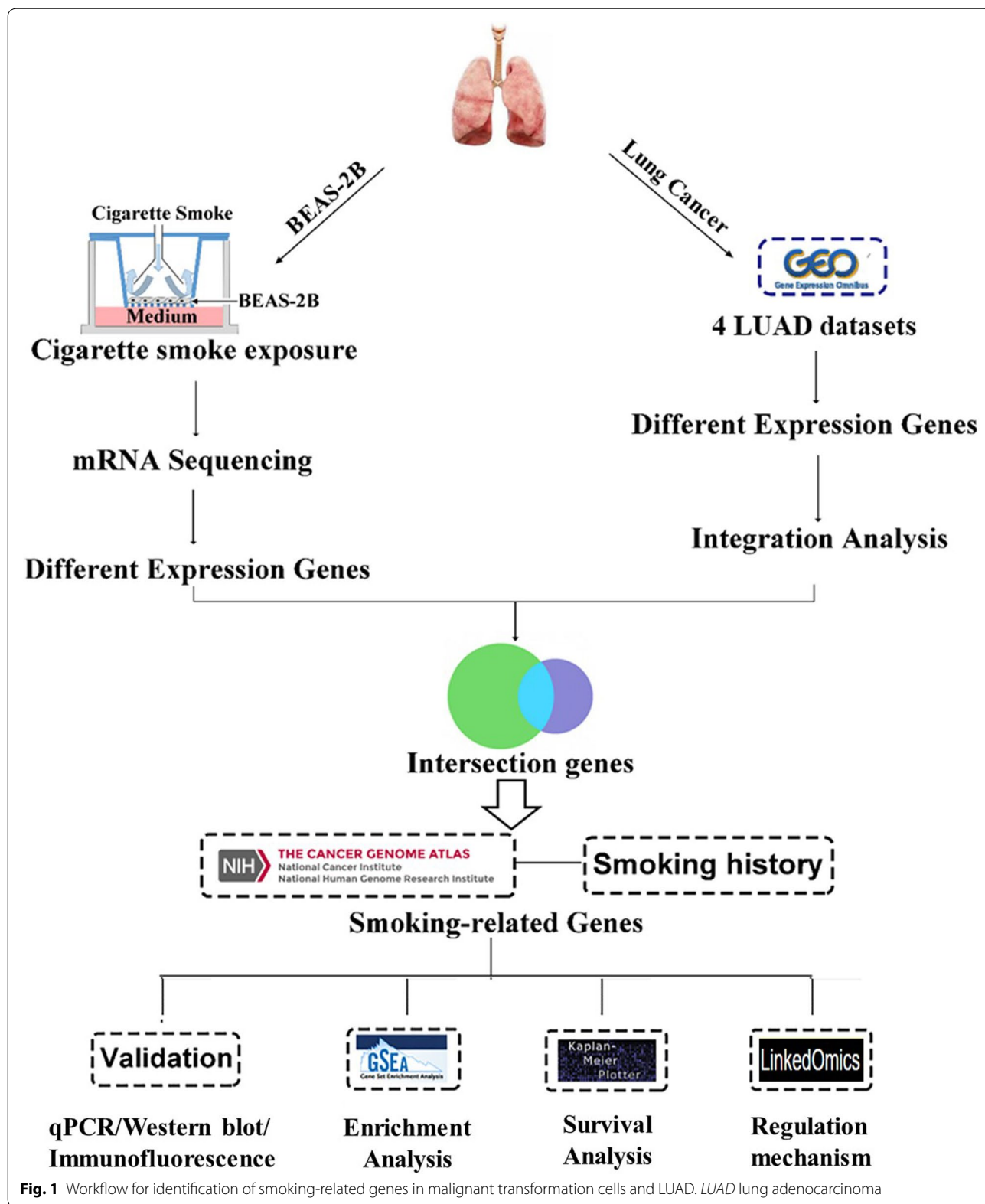


Fig. 1 Workflow for identification of smoking-related genes in malignant transformation cells and LUAD. LUAD lung adenocarcinoma

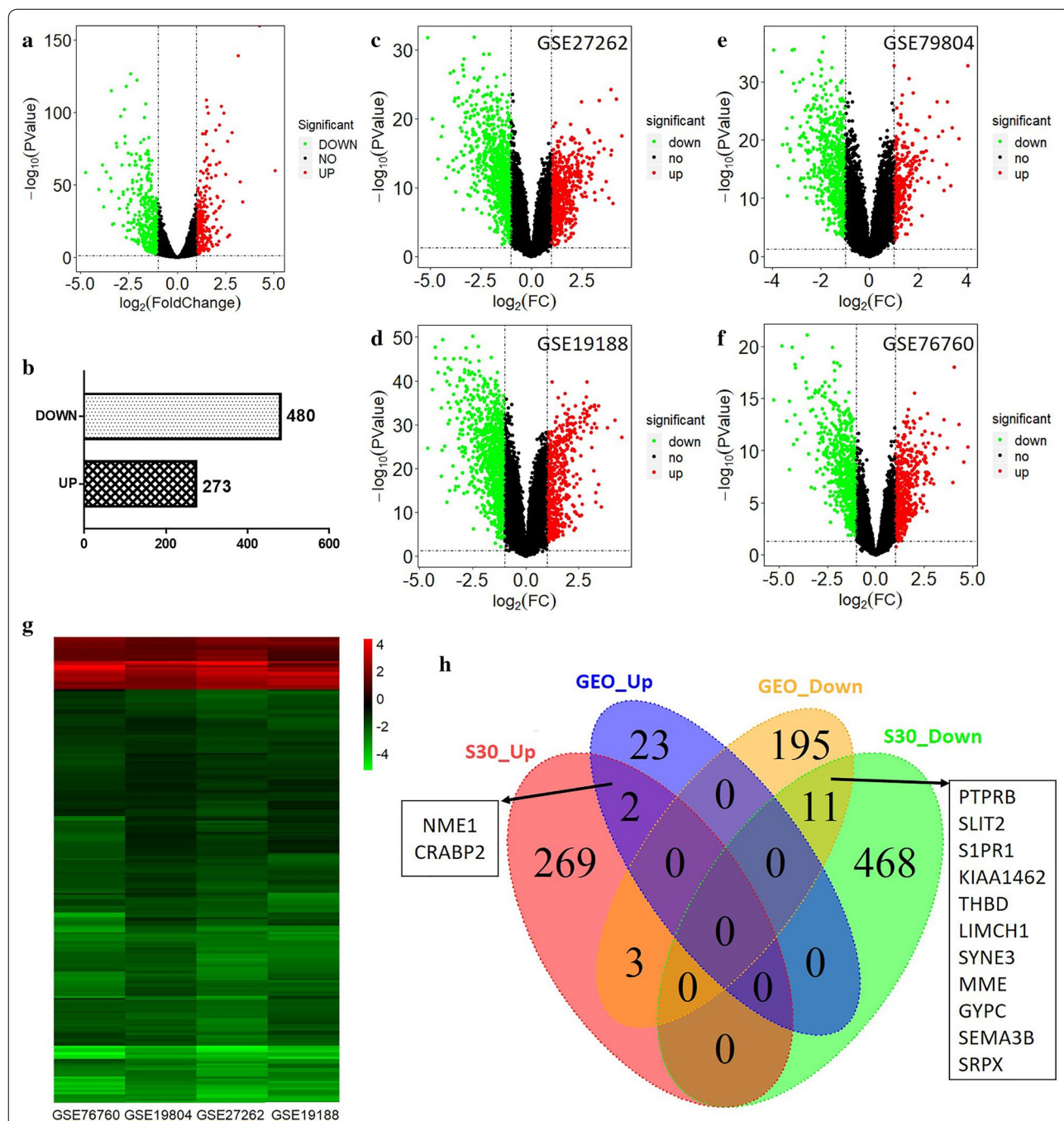


Fig. 2 Identification of smoking-related genes in lung cancer. **a** A volcano plot was generated to visualize the distribution of DEGs. **b** Counts of upregulated or downregulated mRNAs. Volcano plots were generated to visualize the distribution of DEGs between LUAD tissues and adjacent normal tissues from different study cohorts, including GSE27262 (**c**), GSE19804 (**d**), GSE19188 (**e**) and GSE76760 (**f**). The X-axis of volcano plot indicates the fold change (FC, log-scaled), whereas the Y-axis shows the p -values (log-scaled). Each dot represents a different gene, and the red/green color of the dots categorizes the up-regulated/down-regulated genes under the filtering condition. **g** Heatmap of DEGs derived from integrated analysis. Each column represents one dataset and each row represents one gene; the gradual color ranged from green to red represents the changing process from down-regulation to up-regulation. **h** Venn diagram showing the overlap of identified DEGs from GEO datasets and cigarette smoke-induced malignant-transformation-cell model

Association of mRNA expression with pathological characteristics

We further investigated the association between the expression levels of the three genes (*NME1*, *SLIT2* and *GYPE*) and pathological features (Table 2). For invasion depth, the expression levels of *GYPE* and *SLIT2* were significantly decreased in LUAD tissues at the T2 stage compared with T1 stage tissues, while the expression level of *NME1* significantly increased. When lymph node metastasis was considered, the mRNA expression level of *NME1* was greater in N1 and N2 vs. N0. Besides, the mRNA expression level of *NME1* was upregulated in TNM stage III vs. stage I. It is worth mentioning that the expression levels of *GYPE* and *NME1* were significantly different between males and females, and *NME1* and *SLIT2* were substantially different between elder patients (≥ 60 years old) and patients aged less than 60 years old.

Gene Ontology enrichment analysis

Based on the UALCAN online tool, a total of 1182, 1771 and 1822 genes significantly correlated with *GYPE*, *NME1* and *SLIT2* were extracted respectively. Gene ontology enrichment analysis was performed to demonstrate the potential biological functions of these related genes using DAVID. The results showed that these three genes were functionally associated with several critical biological processes. For *GYPE*, the genes co-expressed with it were remarkably enriched in apoptotic signaling pathway and extracellular matrix organization, as well as cell adhesion (Fig. 5a). When *NME1* was considered, the related genes were found to be enriched in cell proliferation, DNA repair and cell cycle, as well as Wnt signaling pathway (Fig. 5d). Besides, the genes correlated with *SLIT2* were significantly associated with extracellular matrix organization, JAK-STAT cascade, and cell adhesion (Fig. 5g). In addition, GSEA enrichment analysis confirmed the three genes signatures, including calcium mediated signaling and regulation of cell-cell adhesion for *GYPE* (Fig. 5b, c), RNA catabolic process and regulation of cell cycle phase transition for *NME1* (Fig. 5e, f), as well as cell matrix adhesion and TGF- β receptor signaling pathway for *SLIT2* (Fig. 5h, i).

KEGG pathway enrichment analysis

To better illustrate the functional role of these three genes in lung cancer, KEGG pathway analysis was performed for the related genes using DAVID online tool. The results showed that these three genes were involved in multiple pathways in lung cancer. When considering *GYPE*, the related genes were involved in JAK-STAT, PI3K-Akt, and Ras/Rap1 signaling pathways, as well as focal adhesion and cell adhesion molecules

(Fig. 6a). Besides, genes related to *NME1* were found to be enriched in base excision repair, mismatch repair, and cell cycle (Fig. 6d). Similarly, the *SLIT2* related genes were significantly associated with TGF- β receptor, VEGF, MAPK and JAK-STAT signaling pathways, as well as lung small lung cancer and focal adhesion (Fig. 6g). The further GSEA enrichment confirmed these pathway signatures of *GYPE* (Fig. 6b, c), *NME1* (Fig. 6e, f), and *SLIT2* (Fig. 6h, i).

Survival analysis

A total of 7 data cohorts, including 1221 LUAD patients, were used to establish univariate Cox models. The univariate Cox model analyses of TCGA dataset suggested that higher *GYPE* expression has a favorable prognosis (HR < 1, $p < 0.05$) (Fig. 7a); The analysis in GSE13213 and GSE30219 datasets revealed that higher *NME1* expression was a risk factor for LUAD patients prognosis (HR < 1, $p < 0.05$) (Fig. 7b). In addition, the analysis in GSE13213 and GSE41271 indicated that increased *SLIT2* expression is associated with a better prognosis (HR < 1, $p < 0.05$) (Fig. 7c). The KM survival analysis is consistent with the univariate Cox analysis (Fig. 8).

Gene dysregulation is mediated by methylation and gene amplification in lung cancer

To understand possible regulation mechanisms of dysregulation of these three genes, we analyzed the public multi-omics datasets in the TCGA database. Based on the UALCAN online tool, the promoter regions of *GYPE* and *SLIT2* were found hyper-methylated significantly, as well as the *NME1* promoter was hypo-methylated (Fig. 9a–c). Also, further person correlation analysis suggested that these three genes mRNA expression levels were remarkably negatively correlated with promoter methylation levels ($r = -0.455$, -0.208 and -0.263 , all $p < 0.001$, Fig. 9d–f). Besides, the up-regulated *NME1* expression was found significantly positively associated with increased gene amplification ($r = 0.349$, $p < 0.001$, Fig. 10b, e). Unexpectedly, the expression of *GYPE* was found negatively correlated to gene amplification ($r = -0.147$, $p < 0.001$, Fig. 10a, d). However, there is no significant correlation between *SLIT2* expression and gene amplification ($r = -0.010$, $p = 0.822$, (Fig. 10c, f).

Discussion

Cigarette smoking is the primary risk factor for lung cancer development, and it is a significant contributor to the morbidity and mortality of LUAD patients [10, 11]. Recent studies have shown that histologically similar lung tumors have different molecular mechanisms of carcinogenesis because of different smoking status [12]. Thus,

the mechanism of lung cancer in smokers and non-smokers needs to be assessed differently.

With the rapid development of sequencing and data analysis technologies, some essential genes related to LUAD have been identified through bioinformatics analysis. For example, *SPPI* has been identified as a prognostic biomarker in four LUAD datasets in the GEO database, which was also validated by the TCGA database [13]. Fan et al suggested 12 significant biomarkers that could distinguish lung cancer patients with different risks from the GEO database [14]. Gan et al identified the aberrantly expressed miR-375 gene involved in LUAD through the comparison of miRNA expression profiles in cancerous tissues based on the analysis and validation from TCGA and GEO datasets and published studies [15]. In the present study, three smoking-related signature genes, namely *GYPC*, *NME1* and *SLIT2*, were identified by an integrated analysis on the LUAD datasets in the GEO database and the high throughput sequencing data of cigarette smoke-induced malignant transformed BEAS-2B cells. In particular, *NME1* was reported increased by cigarette smoking in oral squamous cell carcinoma (OSCC) [16], and the lower expression of *SLIT2* were found in the lungs of cigarette smoke-induced emphysema mice [17].

Glycophorin C (*GYPC*) is an integral membrane glycoprotein. A recent study suggested *GYPC* can be used as a biomarker of breast cancer [18]. Increased *GYPC* gene expression was also reported to correlate with a worse outcome in childhood acute lymphoblastic leukemia [19]. However, there is limited evidence for the function of *GYPC* in the majority of solid tumors, especially in lung cancer. Our present study suggested that increased expression of *GYPC* was associated with a favorable outcome in LUAD patients. GSEA enrichment analysis indicated that *GYPC* was significantly associated with the JAK/STAT and cell adhesion signaling pathways which are essential to lung cancer progression and migration.

The NME/NM23 nucleoside diphosphate kinase 1 (NM23-H1, *NME1*) is a metastasis-associated gene the expression of which was correlated with various tumors. The expression level of *NME1* was found to be inversely proportional to the metastasis potential of several cancers, including breast cancer [20], gastric cancer [21], melanoma [22] and colon adenocarcinoma [23]. These results indicate that *NME1* may act as a metastasis suppressor in these tumors. More importantly, elevated expression of *NME1* has a better prognosis outcome in patients with breast cancer [24] and melanoma [25]. Instead, higher *NME1* expression was significantly associated with poor prognosis in patients with neuroblastoma [26] and osteosarcoma [27], as well as cervical cancer [28]. In this study, increased mRNA expression of *NME1* was found to be related to LUAD invasion

depth and lymph node metastasis. It is noteworthy that decreased expression of *NME1* was associated with an improved prognosis in LUAD patients. Studies have shown that lymphatic metastasis is directly associated with distant recurrence and poor overall survival (OS) in non-small cell lung cancer patients [29]. We further found that *NME1* was involved in the cell cycle pathway, and the defect of cell cycle regulation has been reported to contribute to uncontrolled cancer cell proliferation [30]. Thus, we suggested that *NME1* plays a different role in LUAD than many other cancers, and it may serve as a potential biomarker for LUAD.

SLIT2, a secreted glycoprotein of the SLIT family [31], is involved in the epithelial-mesenchymal transition (EMT) process [32], which permits cancer cells to acquire migratory, invasive, and stem-like properties [33]. The knockdown of the *SLIT2* gene promoted the growth of gastric cancer cells and metastasis through activation of the AKT/ β -catenin-mediated signaling pathway [34]. Another study demonstrated that decreased expression of *SLIT2* is associated with a poor prognosis and brain-specific metastasis in breast cancer patients [35]. The results in the present study showed that *SLIT2* was down-regulated in cigarette smoke-exposed cells and LUAD, and increased expression of *SLIT2* was associated with a better outcome in LUAD patients. Furthermore, *SLIT2* was significantly enriched in the TGF- β signaling and focal adhesion pathways by enrichment analysis. Since both signaling pathways contribute to EMT activation, we suggested that *SLIT2* might serve as a potential tumor metastasis indicator.

The study of epigenetics provides an important clue for understanding the genesis and development of lung cancer [36]. Recent studies suggested that during the genesis of lung cancer, the promoter methylation levels of genes associated with multiple cellular functions are increased [37]. Our analysis determined that these genes expression levels were significantly negatively correlated with the promoter methylation status, which indicated that they may be genomically regulated by DNA methylation. Besides, we analyzed the correlation between gene expression levels and copy numbers, which is another regulation mechanism at genome level. Copy number variation (CNV) is generally considered to be any genomic alteration greater than 50 base pairs in length [38], and it has been shown to play an essential role in human cancer. The loss of tumor suppressor genes and the gain of proto-oncogenes can contribute to cancer development [39, 40]. In this study, only *NME1* was found to have a significant positive correlation among expression and copy numbers, indicating that *NME1* was regulated not only by promoter methylation, but by copy numbers.

Conclusions

In conclusion, our results indicated that *GYPC*, *NME1* and *SLIT2* may play a vital role in the development of smoking-related LUAD, which will be helpful in predicting the prognosis of LUAD patients. Mechanically, these three genes may be regulated by DNA methylation. Further *in vitro* and *in vivo* studies are needed to gain insights into the underlying molecular mechanisms of these three genes in LUAD.

Methods

BEAS-2B cell culture and cigarette smoke exposure

Human bronchial epithelial cells (BEAS-2B) and human LUAD cell lines (PC9, A549, H1975 and H1299) were purchased from the American Type Culture Collection (ATCC, USA). They were maintained in basal LHC-8 nutrient medium (Thermo Fisher Scientific, Waltham, MA, USA) or Dulbecco's Modified Eagle Medium (DMEM, Invitrogen, MA, USA) in an incubator maintaining at 37 °C and humidified atmosphere of 5% CO₂. The *in vitro* model for malignant transformation was established by exposing BEAS-2B cells to cigarette smoke continuously for 5, 10, 20 and 30 passages (S5, S10, S20 and S30, separately), and this has been described in detail previously [41, 42].

RNA isolation and high-throughput sequencing

Total RNA was isolated from normal BEAS-2B and S30 cells with TRIzol RNA isolation reagent (Invitrogen, MA, USA) according to the manufacturer's protocol. Three biological replicates per group were used for mRNA sequencing analysis. A total of 1.5 μg RNA per sample was used as input material for the RNA library construction. The mRNA sequencing procedure has been described in detail in our previous article, and the raw data has been deposited in the Sequence Read Archive (SRA) database (<https://trace.ncbi.nlm.nih.gov/Traces/sra/>) with identifier SRP181756 [43]. Genes were computed by summing the fragments per kilo-base of exon per million fragments mapped (FPKM) of transcripts in each gene group.

Human database extraction

Four datasets, GSE27262 [44], GSE19188 [45], GSE76760 [46] and GSE19804 [47], were downloaded from the Gene Expression Omnibus (GEO) database for differential expression analysis (Additional file 1: Table S1). All datasets met the following two criteria: (1) tissue samples obtained from human LUAD and adjacent normal tissues; and (2) each set included at least 50 samples. An additional six other datasets with survival information were downloaded for survival analysis (Additional

file 1: Table S1). In these datasets, only LUAD and normal samples were retrieved and analyzed. The RNA-seq by the expected maximization (RSEM) data and the corresponding clinical information of LUAD in The Cancer Gene Atlas (TCGA) database were obtained from Xena (<https://xena.ucsc.edu>). The RSEM gene expression measurements for LUAD cases were transformed by using log₂ (RSEM + 1).

Differential expression and integration analysis

For our RNA sequencing data, differential expression analysis of normal BEAS-2B and S30 cells was performed using the “DESeq2” R package [48]. For GEO and TCGA datasets, the “Limma” package was subsequently employed for identifying differentially expressed genes (DEGs) in each dataset [49]. |log₂FC| >1 and a *p*-value < 0.05 were considered statistically significant for the DEGs. Gene integration for the DEGs identified from the four datasets was conducted using another R package “RobustRankAggreg” [50]. The expression levels of integrated genes in four GEO datasets were visualized with the R package “pheatmap” (<https://cran.r-project.org/web/packages/pheatmap/index.html>). Furthermore, a Venn diagram was generated by the “VennDiagram” R package (<https://cran.r-project.org/web/packages/VennDiagram/index.html>) to visualize the genes with the consistent change in S30 cells and LUAD samples.

Real-time quantitative PCR

The total RNA of cells (including Beas-2B, S5, S10, S20 and S30 cells, as well as 4 LUAD cell lines) was isolated using TRIzol reagent (Invitrogen, MA, USA) according to the manufacturer's protocol. A total amount of 1.5 μg of total RNA from each sample was reversely transcribed into complementary DNA (cDNA) using Revert Aid First Strand Complementary DNA Synthesis Kit (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. Quantitative PCR (qPCR) was performed using NovoScript[®] SYBR Two-Step qRT-PCR Kit (novoprotein, China) on QuantStudio[™] 6 Flex qRT-PCR system (Applied Biosystems, Foster City, CA, USA). GAPDH was used as a reference. The primer pairs used for qPCR in this study were listed in Table 3.

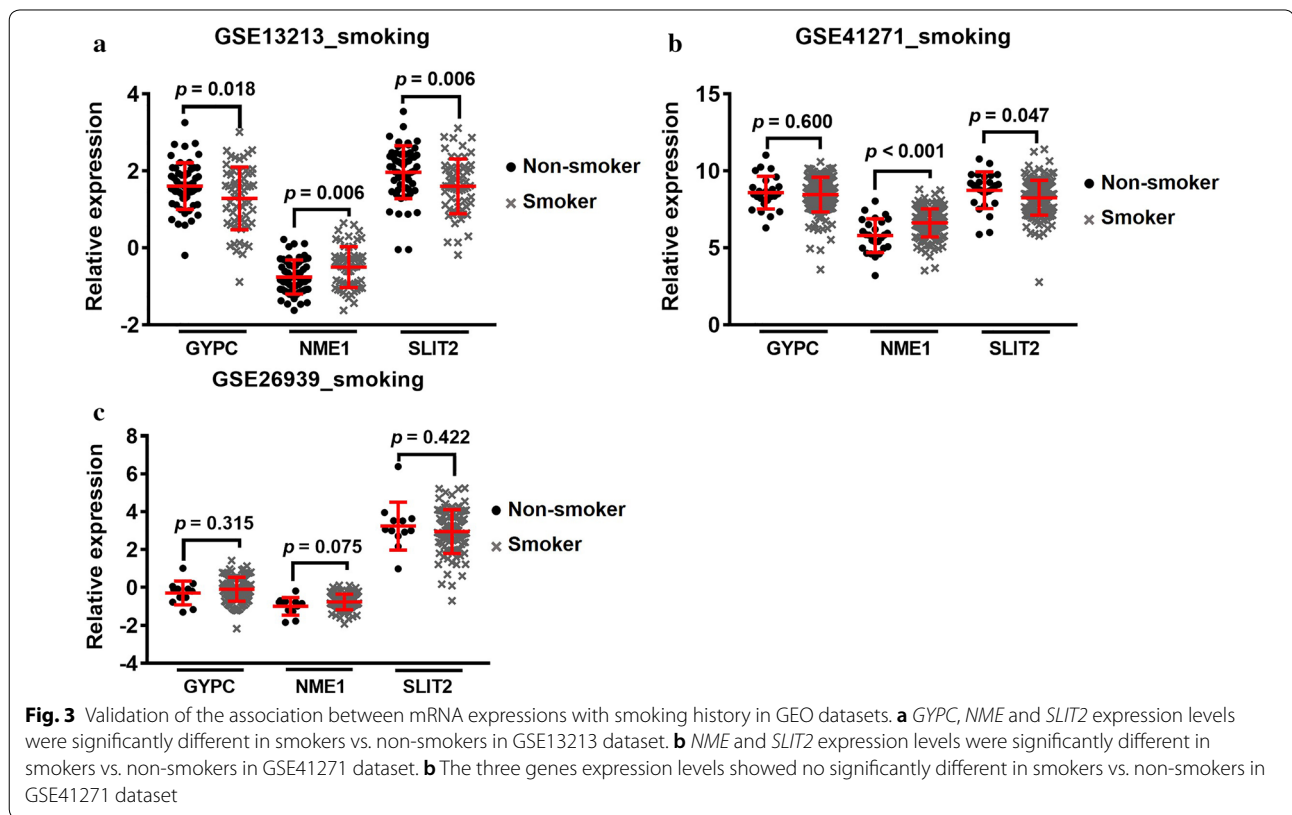
Immunofluorescence analysis

The S30 and unexposed BEAS-2B cells were fixed in PBS containing 4.0% paraformaldehyde without methanol. The cells were washed and permeabilized with 0.2% Triton X-100 and blocked with 5% goat serum for 1 hour at room temperature. Diluted antibodies for human NME1 (11086-2-AP, Proteintech, Chicago, IL, USA), SLIT2 (20217-1-AP, Proteintech, Chicago, IL, USA) or GYPC (ab108619, Abcam, Cambridge, MA, USA) were added

Table 1 Association of mRNA expression with LUAD patient tobacco smoking history in the TCGA database

Smoking history	Cases (N)	<i>GYPC</i> (log ₂ (RSEM+1))	<i>NME1</i> (log ₂ (RSEM+1))	<i>SLIT2</i> (log ₂ (RSEM+1))
1	75	8.64 ± 0.93	10.64 ± 0.98	7.85 ± 1.43
2	119	8.20 ± 1.00**	11.06 ± 0.70**	7.17 ± 1.53**
3	135	8.39 ± 0.82	10.62 ± 0.91##	7.69 ± 1.50##
4	168	8.47 ± 0.92#	10.99 ± 0.80**	7.54 ± 1.55#

1 = Lifelong Non-smokers (less than 100 cigarettes smoked in Lifetime), 2 = Current smokers (includes daily smokers and non-daily smokers or occasional smokers), 3 = Current reformed smokers for >15 years (greater than 15 years), 4 = Current reformed smokers for ≤15 years (less than or equal to 15 years). Data represented are Mean ± SD, n depend on how many valid LUAD samples with corresponding factors. SD indicates standard deviation. **p* < 0.05, versus Lifelong Non-smokers; ***p* < 0.01, versus Current smokers; #*p* < 0.05, versus Lifelong Non-smokers; ##*p* < 0.01, versus Current smokers



drop by drop and the slides were kept in a wet box at 4 °C overnight. Following incubated with FITC-conjugated goat anti-rabbit IgG for 1 hour at room temperature, the

slides were washed, and the nuclei were counter-stained with 4,6-diamidino-2-phenylindole (DAPI). Photographs were taken and visualized using an FV1200MPE

(See figure on next page.)

Fig. 4 Validation of mRNA and protein expression levels of the three genes. The mRNA expression levels of *GYPC* (**a**), *NME1* (**b**) and *SLIT2* (**c**) in cigarette smoke-exposed cells. 2B, normal BEAS-2B cells serve as a control. S5 to S30, BEAS-2B cells exposed to cigarette smoke for different passages. **d** Heatmap of *GYPC*, *NME1* and *SLIT2* in BEAS-2B and four LUAD cell lines. Each column represents one cell line, and each row represents one gene; the gradual color ranged from blue to red represents the changing process from down-regulation to up-regulation. **e** Immunofluorescence staining of *GYPC*. **f** Immunofluorescence staining of *NME1*. **g** Immunofluorescence staining of *SLIT2*. **h** Western blotting results. S5 to S30, BEAS-2B cells exposed to cigarette smoke for different passages. **p* < 0.05, versus normal BEAS-2B cells; ***p* < 0.01, versus normal BEAS-2B cells

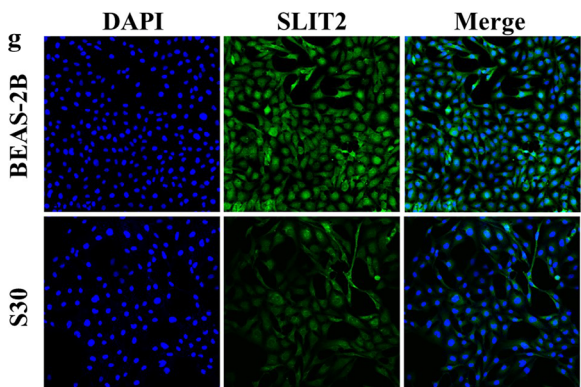
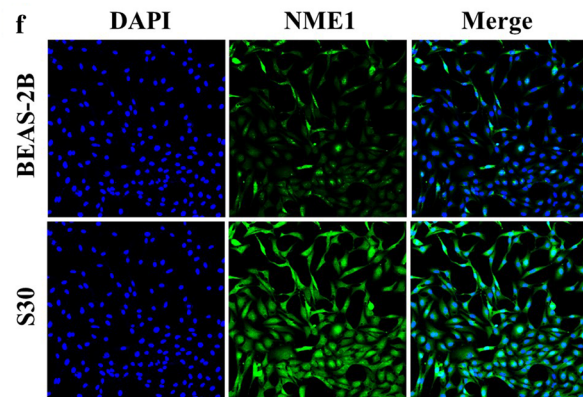
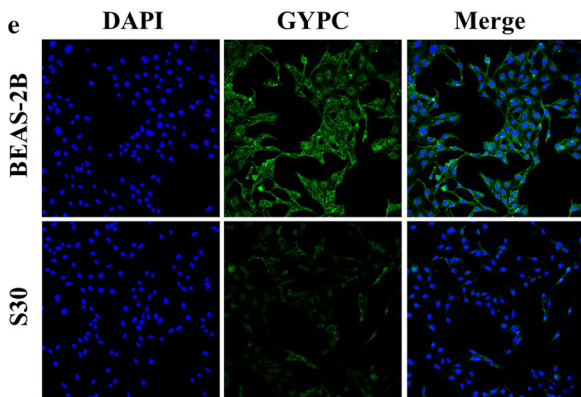
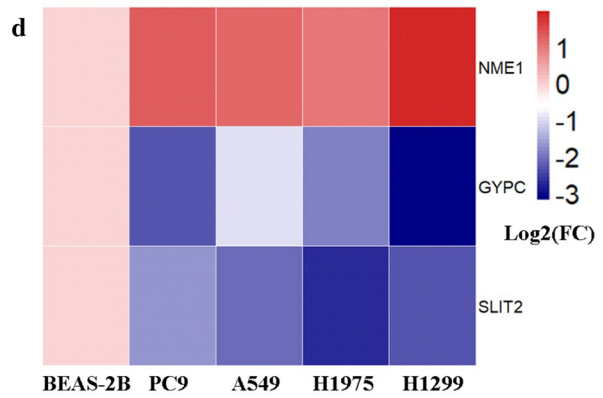
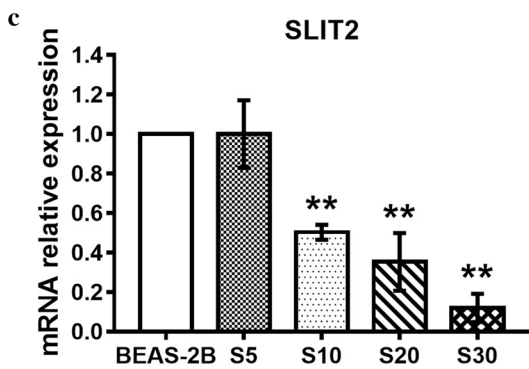
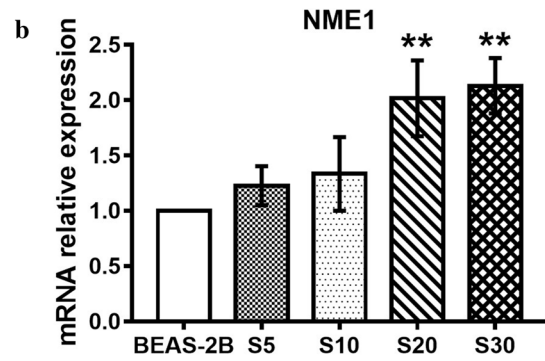
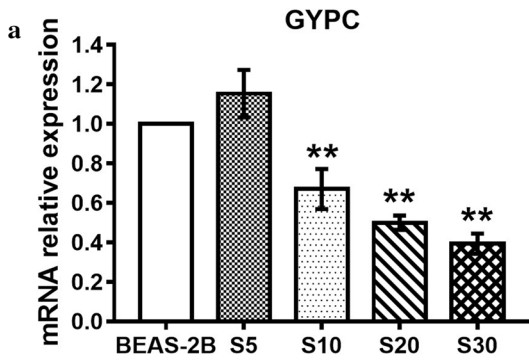


Table 2 Association of mRNA expression levels of GYPC, NME1 and SLIT2 with the pathological features

Factor	Case	GYPC	NME1	SLIT2
Gender				
Male	238	8.31 ± 0.91	10.95 ± 0.86	7.44 ± 1.57
Female	277	8.52 ± 0.94*	10.80 ± 0.84*	7.63 ± 1.47
Age				
< 60	136	8.37 ± 0.99	11.00 ± 0.78	7.32 ± 1.61
≥60	360	8.45 ± 0.91	10.82 ± 0.88*	7.63 ± 1.48*
Invasion depth				
T1	169	8.63 ± 0.77	10.67 ± 0.88	7.89 ± 1.31
T2	277	8.32 ± 0.99**	10.96 ± 0.76**	7.35 ± 1.57**
T3	47	8.38 ± 0.92	10.97 ± 1.11	7.35 ± 1.69*
T4	19	7.99 ± 0.87**	11.07 ± 0.98	7.34 ± 1.29
Lymph node metastasis				
N0	331	8.43 ± 0.95	10.78 ± 0.85	7.58 ± 1.55
N1	96	8.43 ± 0.89	11.04 ± 0.85**	7.34 ± 1.46
N2	74	8.27 ± 0.85	11.10 ± 0.81**	7.47 ± 1.44
N3	2	8.46 ± 0.19	11.04 ± 1.07	6.76 ± 0.46
Distant metastasis				
M0	346	8.43 ± 0.92	10.95 ± 0.85	7.50 ± 1.48
M1	25	8.06 ± 0.88	10.96 ± 0.91	7.66 ± 1.70
TNM stage				
I	275	8.46 ± 0.92	7.60 ± 1.49	10.78 ± 0.82
II	122	8.47 ± 0.99	7.47 ± 1.59	10.87 ± 0.85
III	84	8.30 ± 0.83	7.36 ± 1.41**	11.12 ± 0.89
IV	26	8.17 ± 1.02	7.78 ± 1.78	10.92 ± 0.91

Note: The data was presented as in $\log_2(x+1)$ transformed RSEM normalized count. LUAD, lung adenocarcinoma; TCGA: The Cancer Gene Atlas; TNM: tumor-node-metastasis; stage RSEM: RNA seq by expectation-maximization. * $p < 0.05$, versus the first group of the corresponding feature. ** $p < 0.01$, versus the first group of the corresponding feature

multiphoton laser scanning microscope (FV1200, OLYMPUS, Japan). The acquisition parameters were held constant for all the experiments.

Western blot analysis

Total protein was extracted with RIPA buffer, and 20 μg of extracted total proteins were separated on SDS-PAGE gel and transferred onto a PVDF membrane (Millipore, Billerica, MA). After blocking with 5% bovine serum album (BSA, solarbio, China), the membrane was incubated at 4 °C overnight with various primary antibodies, including NME1, SLIT2 and GYPC. The HRP-labeled secondary antibody was used according to the host species of the primary antibody. Western blots were developed using electrochemiluminescence (ECL) substrate and visualized using the GeneTools GBox (Syngene, Frederick, MD, USA) system. The intensity of each spot was quantified using NIH ImageJ software (NIH, Bethesda, MD, USA).

Analysis of patient smoking and pathological features

The mRNA expression levels of integrated genes in LUAD patients with different smoking histories were examined to identify the genes related to smoking. All samples in the TCGA LUAD and GEO datasets were divided into different groups according to the patient's clinical features, including gender, age and invasion depth, as well as the status of lymph node metastasis, distant metastasis and tumor-node-metastasis (TNM) stage to find an association with mRNA expression levels.

Enrichment analysis

To investigate the potential biological function of these three smoking-related genes, genes related to them were extracted from UALCAN (<http://ualcan.path.uab.edu/index.html>) with the Pearson coefficient ≥ 0.3 [51]. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed using the DAVID online tool (<https://david>

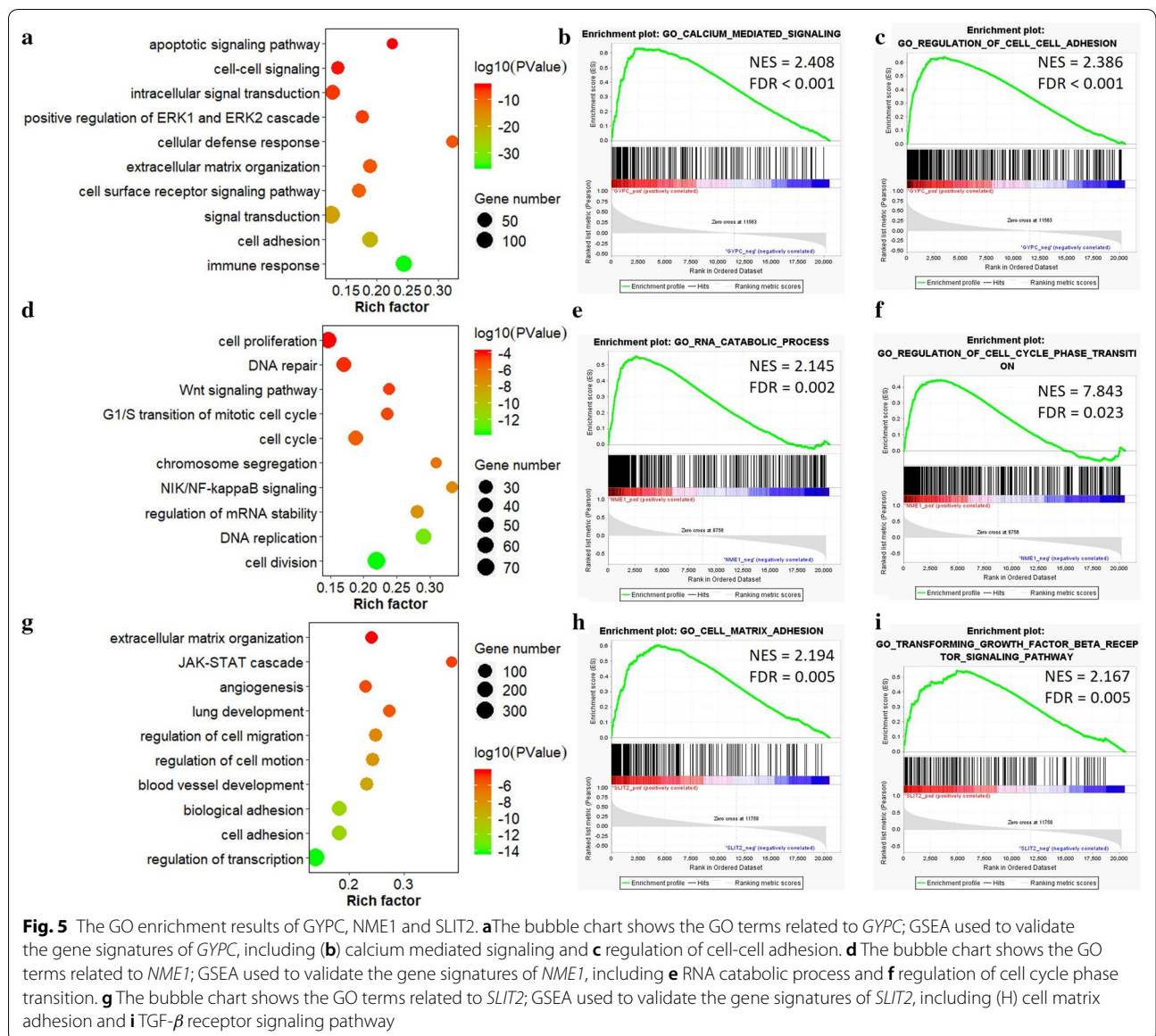


Fig. 5 The GO enrichment results of *GYPC*, *NME1* and *SLIT2*. **a**The bubble chart shows the GO terms related to *GYPC*; GSEA used to validate the gene signatures of *GYPC*, including **(b)** calcium mediated signaling and **(c)** regulation of cell-cell adhesion. **d** The bubble chart shows the GO terms related to *NME1*; GSEA used to validate the gene signatures of *NME1*, including **(e)** RNA catabolic process and **(f)** regulation of cell cycle phase transition. **g** The bubble chart shows the GO terms related to *SLIT2*; GSEA used to validate the gene signatures of *SLIT2*, including **(h)** cell matrix adhesion and **(i)** TGF- β receptor signaling pathway

.ncicrf.gov/). Besides, the gene set enrichment analysis (GSEA, <http://software.broadinstitute.org/gsea/index.jsp>)[52] was used to verify the biological processes and KEGG pathways related to these genes. For the GSEA parameters, “1000”, “gene_set”, “weighted”, and “Pearson” were selected as “Number of permutations”, “Permutation type”, “Enrichment statistic”, and “Metric for ranking genes”, respectively.

Survival analysis

The univariate Cox regression models were used to calculate the hazard ratios (HRs) and the 95% confidence intervals (CIs) based on *GYPC*, *NME1* and *SLIT2* expression levels in LUAD patients in six GEO datasets (GSE13213, GSE26939, GSE30219, GSE41271, GSE42127 and GSE14814) and the TCGA LUAD dataset. The Cox analysis was performed using the “survival” (<https://>

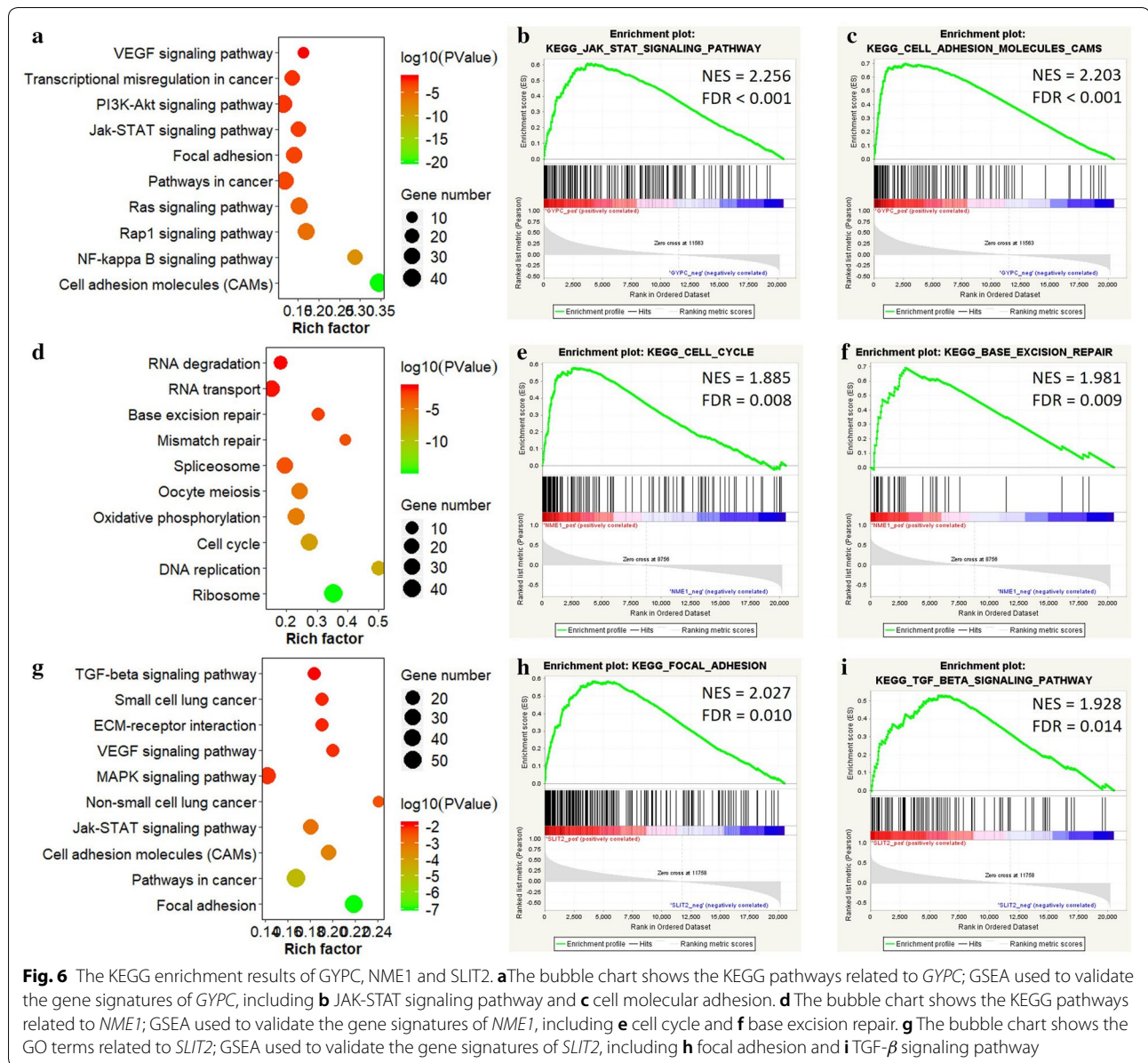


Fig. 6 The KEGG enrichment results of *GYPC*, *NME1* and *SLIT2*. **a**The bubble chart shows the KEGG pathways related to *GYPC*; GSEA used to validate the gene signatures of *GYPC*, including **b** JAK-STAT signaling pathway and **c** cell molecular adhesion. **d** The bubble chart shows the KEGG pathways related to *NME1*; GSEA used to validate the gene signatures of *NME1*, including **e** cell cycle and **f** base excision repair. **g** The bubble chart shows the GO terms related to *SLIT2*; GSEA used to validate the gene signatures of *SLIT2*, including **h** focal adhesion and **i** TGF- β signaling pathway

cran.r-project.org/web/packages/survival/index.html) package, and the “survminer” (<https://github.com/kassambara/survminer>) package was used to generate Kaplan-Meier (KM) survival curves. All LUAD patients were divided into high and low expression groups according to the median values of mRNA expression level.

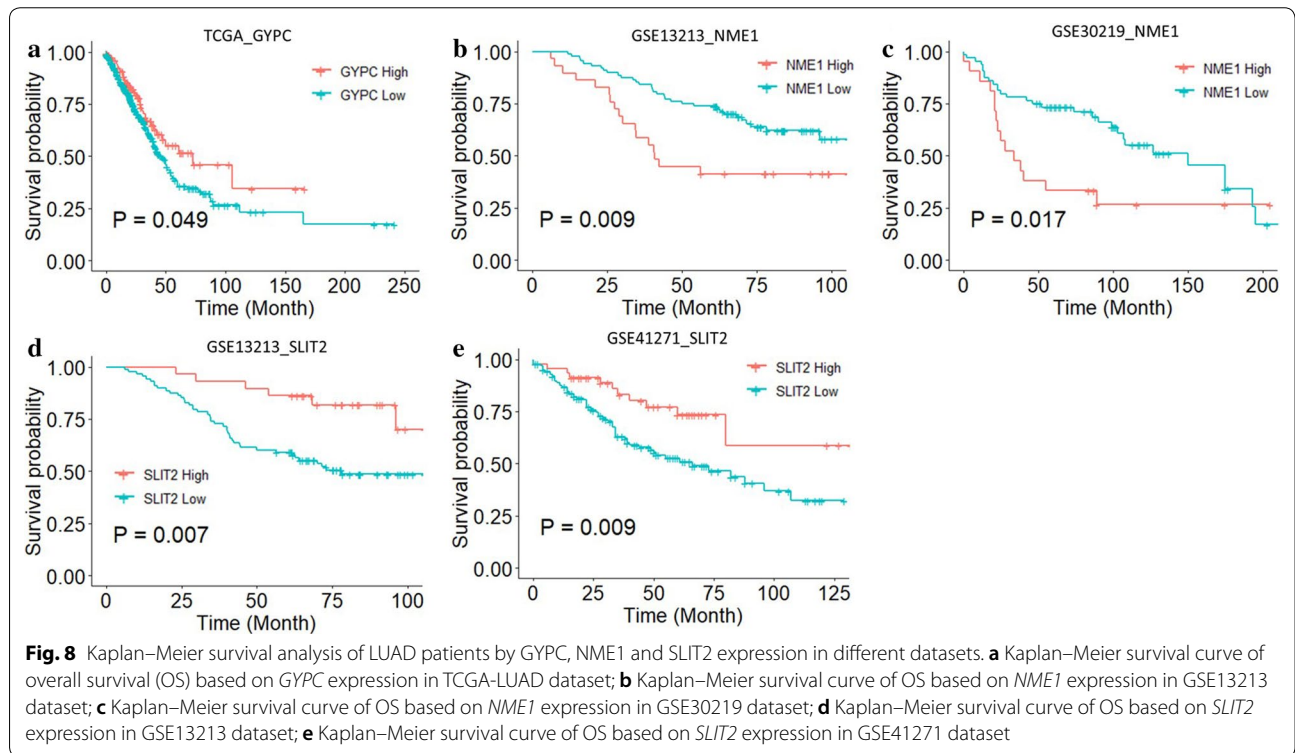
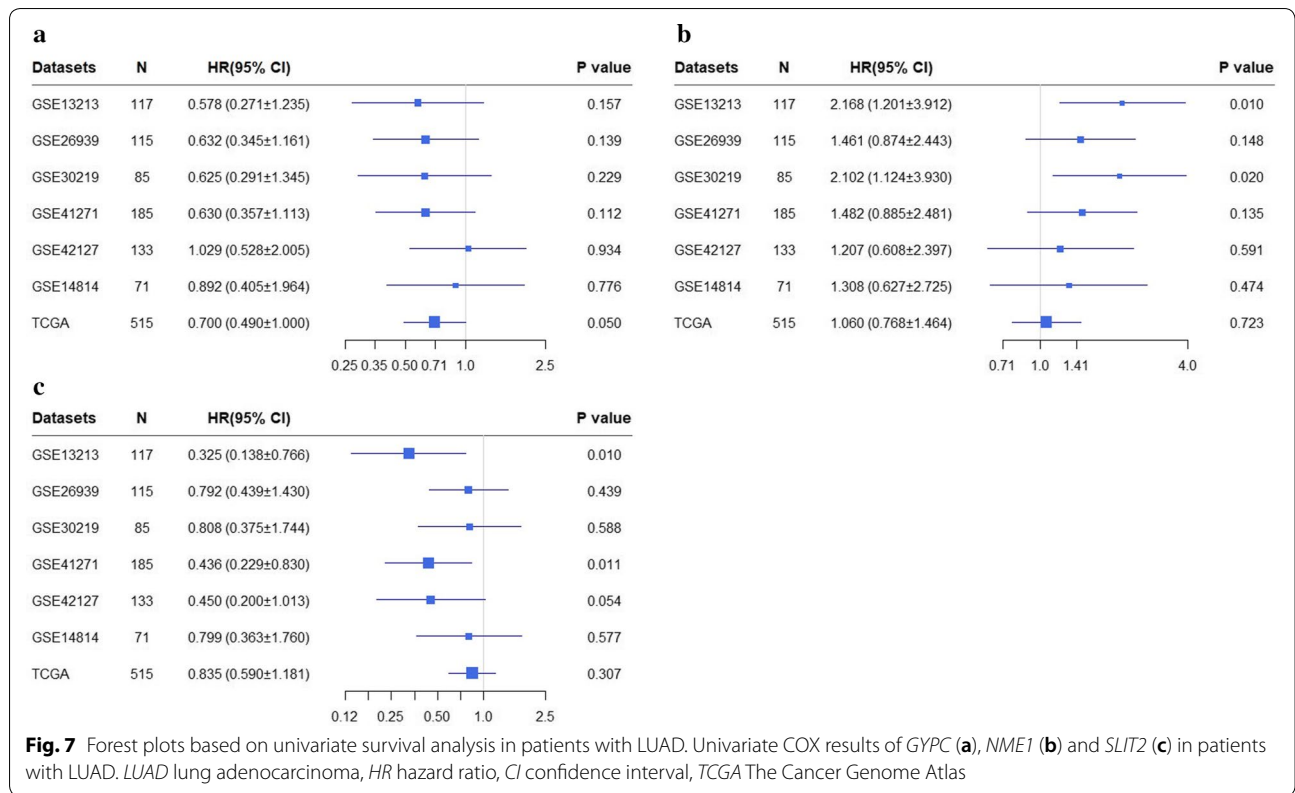
Copy number and methylation analysis

To further understand the regulatory mechanisms of these three genes, we used the multi-omics data in the TCGA database for further correlation analysis. The promoter methylation levels of these three genes in TCGA LUAD dataset were performed using UALCAN

. Besides, the methylation and copy number variation (CNV) data in the TCGA LUAD dataset was obtained from LinkedOmics (<http://www.linkedomics.org/login.php>), and Pearson coefficient was calculated to demonstrate the relationship between them and mRNA expression levels.

Statistical analysis

In this study, heat map, survival, and differential expression analyses were performed with the R software packages. Experimental data were analyzed using GraphPad Prism 7 (GraphPad Software Inc., La Jolla, CA, USA).



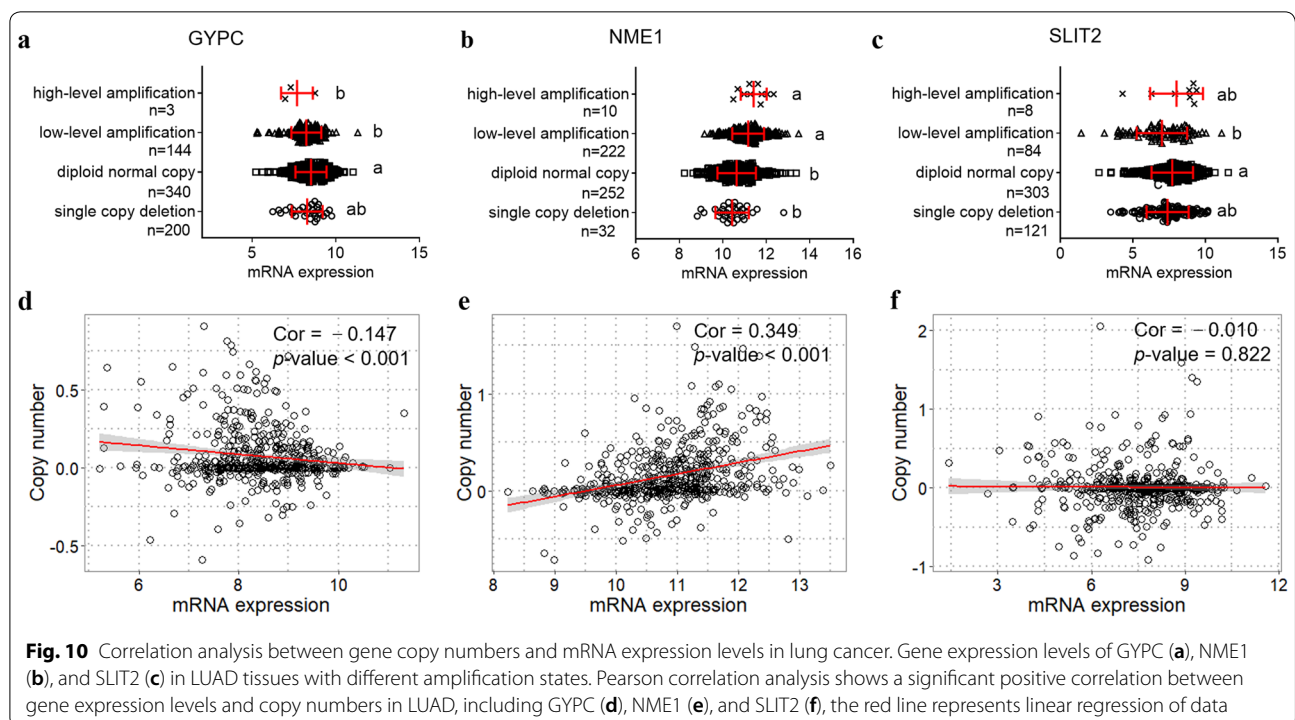
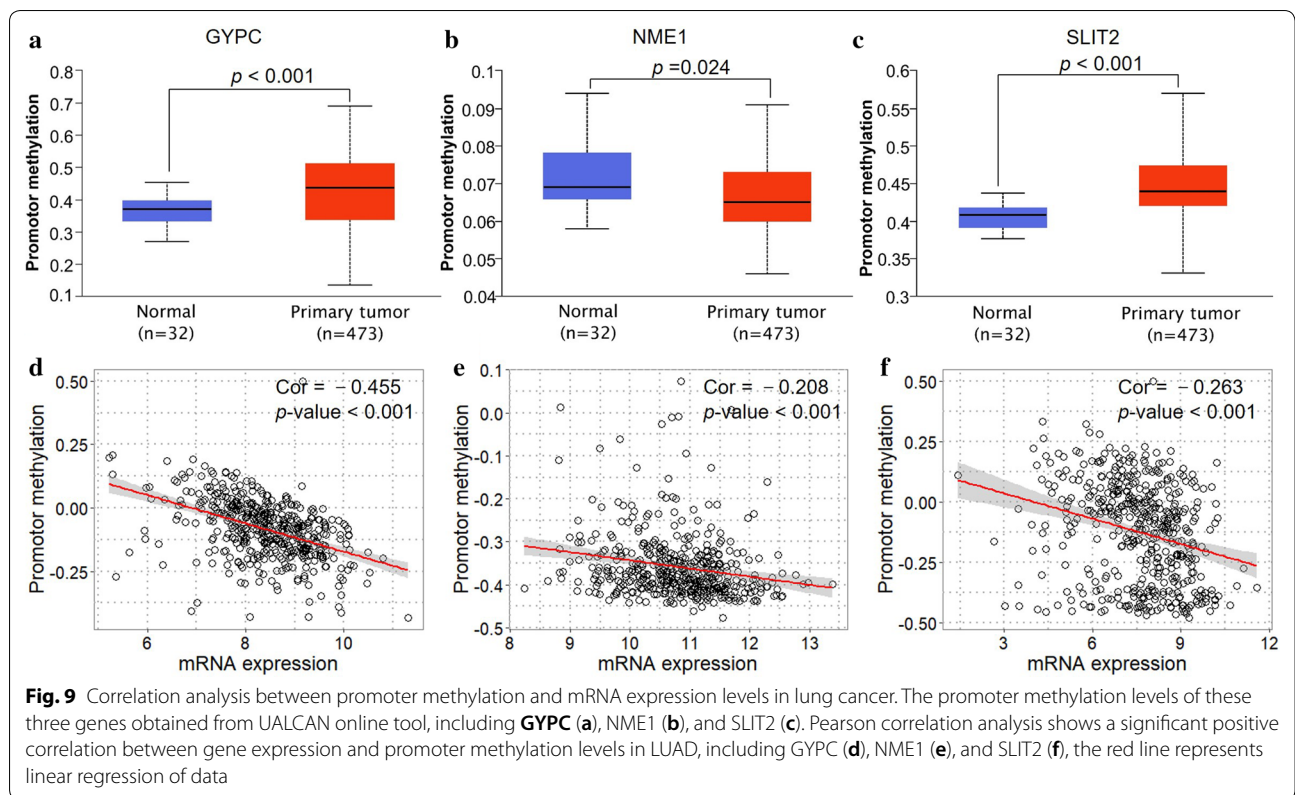


Table 3 Primers used in this study

Primer name	Primer sequence 5' to 3'
F_GYPC	GCCGGATGGCAGAATGGAG
R_GYPC	GGAGGGAGACTAGGACGATGG
F_NME1	AAGGAGATCGGCTTGTGGTTT
R_NME1	CTGAGCACAGCTCGTGAATC
F_SLIT2	AGCCGAGGTTCAAAAACGAGA
R_SLIT2	GGCAGTGCAAAAACACTACAAGA
F_GAPDH	CTGGGTACTACTGAGCACC
R_GAPDH	AAGTGGTCGTTGAGGGCAATG

Data were reported as the mean \pm standard deviation (SD) of three independent experiments. Data were analyzed using Student's *t*-test to compare between two groups. *p*-values < 0.05 were considered significant.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12967-020-02474-x>.

Additional file 1. Additional tables.

Abbreviations

LUAD: Lung adenocarcinoma; BEAS-2B: Normal lung human bronchial epithelial cells; S30: Cells exposed to cigarette smoke continuously for 30 passages; DEGs: Differentially expressed genes; RRA: RobustRankAggreg; GEO: Gene Expression Omnibus; TCGA: The Cancer Gene Analysis; GSEA: Gene Set Enrichment Analysis; qPCR: Quantitative PCR; IF: Immunofluorescence; WB: Western blot; GYPC: Glycophorin C; NME1: NME/NM23 nucleoside diphosphate kinase 1; SLIT2: Slit guidance ligand 2; OS: Overall survival; ATCC: American Type Culture Collection; DMEM: Dulbecco's Modified Eagle Medium; RSEM: RNA-seq by the expected maximization; TNM: Tumor-node-metastasis; TNM: Tumor-node-metastasis; CI: Confidence intervals; KM: Kaplan–Meier; MSigDB: Molecular Signatures Database; cDNA: Complementary DNA; DAPI: 4,6-diamidino-2-phenylindole; SD: Standard deviation; EMT: Epithelial–mesenchymal transition; CNV: Copy number variation.

Acknowledgements

Not applicable.

Authors' contributions

JL conceived this study. JW and TC designed the study. JW and TC acquired and analyzed the high throughput data. JW, TC, XF, NO, LQ and JB contributed to conducting experiments and analyzed the data. JW and TC wrote the paper. JT and JL were of immense help in the preparation of the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (81573178, 81172707). The study was also supported by Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases as well as the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Availability of data and materials

The mRNA sequencing dataset generated for this study can be found in the Sequence Read Archive (SRA) database (<https://trace.ncbi.nlm.nih.gov/Trace/s/sra/>) with identifier SRP181756. The GEO datasets analyzed for this study can be obtained from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>) with corresponding identifier as list in Additional file 1: Table S1.

The TCGA LUAD dataset analyzed for this study can be obtained from UCSC Xena (<https://xenabrowser.net/datapages/>).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Department of Toxicology, School of Public Health, Medical College of Soochow University, Renai Road, Suzhou 215123, China. ² Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, Renai Road, Suzhou 215123, China.

Received: 11 November 2019 Accepted: 30 July 2020

Published online: 14 August 2020

References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424. <https://doi.org/10.3322/caac.21492>.
- Boolell V, Alamgeer M, Watkins DN, Ganju V. The evolution of therapies in non-small cell lung cancer. *Cancers (Basel)*. 2015;7(3):1815–46. <https://doi.org/10.3390/cancers7030864>.
- Gridelli C, Rossi A, Carbone DP, Guarize J, Karachaliou N, Mok T, Petrella F, Spaggiari L, Rosell R. Non-small-cell lung cancer. *Nat Rev Dis Primers*. 2015;1:15009. <https://doi.org/10.1038/nrdp.2015.9>.
- Hecht SS. Tobacco smoke carcinogens and lung cancer. *J Natl Cancer Inst*. 1999;91(14):1194–210.
- Shields PG. Molecular epidemiology of smoking and lung cancer. *Oncogene*. 2002;21(45):6870–6. <https://doi.org/10.1038/sj.onc.1205832>.
- Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, Palma J, Brody JS. Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci USA*. 2004;101(27):10143–8. <https://doi.org/10.1073/pnas.0401422101>.
- Liu Y, Ni R, Zhang H, Miao L, Wang J, Jia W, Wang Y. Identification of feature genes for smoking-related lung adenocarcinoma based on gene expression profile data. *Onco Targets Ther*. 2016;9:7397–407. <https://doi.org/10.2147/OTT.S114230>.
- Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, Mann FE, Fukuoka J, Hames M, Bergen AW, Murphy SE, Yang P, Pesatori AC, Consonni D, Bertazzi PA, Wacholder S, Shih JH, Caporaso NE, Jen J. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS ONE*. 2008;3(2):1651. <https://doi.org/10.1371/journal.pone.0001651>.
- Karlsson A, Ringner M, Lauss M, Botling J, Micke P, Planck M, Staaf J. Genomic and transcriptional alterations in lung adenocarcinoma in relation to smoking history. *Clin Cancer Res*. 2014;20(18):4912–24. <https://doi.org/10.1158/1078-0432.CCR-14-0246>.
- Hecht SS. Lung carcinogenesis by tobacco smoke. *Int J Cancer*. 2012;131(12):2724–32. <https://doi.org/10.1002/ijc.27816>.
- Vucic EA, Thu KL, Pikor LA, Enfield KSS, Yee J, English JC, MacAulay CE, Lam S, Jurisica I, Lam WL. Smoking status impacts microRNA mediated prognosis and lung adenocarcinoma biology. *BMC Cancer*. 2014; <https://doi.org/10.1186/1471-2407-14-778>.
- Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers—a different disease. *Nat Rev Cancer*. 2007;7(10):778–90. <https://doi.org/10.1038/nrc2190>.
- Li SC, Yang RH, Sun X, Miao SC, Lu T, Wang YY, Wo Y, Jiao WJ. Identification of spp1 as a promising biomarker to predict clinical outcome of lung adenocarcinoma individuals. *Gene*. 2018;679:398–404. <https://doi.org/10.1016/j.gene.2018.09.030>.

14. Fan Z, Xue W, Li L, Zhang C, Lu J, Zhai Y, Suo Z, Zhao J. Identification of an early diagnostic biomarker of lung adenocarcinoma based on co-expression similarity and construction of a diagnostic model. *J Transl Med*. 2018;16(1):205. <https://doi.org/10.1186/s12967-018-1577-5>.
15. Gan TQ, Chen WJ, Qin H, Huang SN, Yang LH, Fang YY, Pan LJ, Li ZY, Chen G. Clinical value and prospective pathway signaling of microrna-375 in lung adenocarcinoma: A study based on the cancer genome atlas (tcga), gene expression omnibus (geo) and bioinformatics analysis. *Med Sci Monit*. 2017;23:2453–64.
16. Wang YF, Chow KC, Chang SY, Chiu JH, Tai SK, Li WY, Wang LS. Prognostic significance of nm23-h1 expression in oral squamous cell carcinoma. *Br J Cancer*. 2004;90(11):2186–93. <https://doi.org/10.1038/sj.bjc.6601808>.
17. Lin YZ, Zhong XN, Chen X, Liang Y, Zhang H, Zhu DL. Roundabout signaling pathway involved in the pathogenesis of copd by integrative bioinformatics analysis. *Int J Chron Obstruct Pulmon Dis*. 2019;14:2145–62. <https://doi.org/10.2147/COPD.S216050>.
18. Zhang C, Zhao H, Li J, Liu H, Wang F, Wei Y, Su J, Zhang D, Liu T, Zhang Y. The identification of specific methylation patterns across different cancers. *PLoS ONE*. 2015;10(3):0120361. <https://doi.org/10.1371/journal.pone.0120361>.
19. Zhang JB, Li XH, Ning F, Guo XS. Relationship between expression of gypc and trip3 genes and prognosis of acute lymphoblastic leukemia in children. *Zhongguo Dang Dai Er Ke Za Zhi*. 2009;11(1):29–32.
20. Sirotkovic-Skerlev M, Krizanac S, Kapitanovic S, Husnjak K, Unusic J, Pavelic K. Expression of c-myc, erbB-2, p53 and nm23-h1 gene product in benign and malignant breast lesions: coexpression and correlation with clinicopathologic parameters. *Exp Mol Pathol*. 2005;79(1):42–50. <https://doi.org/10.1016/j.yexmp.2005.02.004>.
21. Muller W, Schneiders A, Hommel G, Gabbert HE. Expression of nm23 in gastric carcinoma: association with tumor progression and poor prognosis. *Cancer*. 1998;83(12):2481–7.
22. Lee CS, Pirdas A, Lee MW. Immunohistochemical demonstration of the nm23-h1 gene product in human malignant melanoma and spitz nevi. *Pathology*. 1996;28(3):220–4.
23. Kapitanovic S, Cacev T, Berkovic M, Popovic-Hadzija M, Radosevic S, Seiwert S, Spaventi S, Pavelic K, Spaventi R. nm23-h1 expression and loss of heterozygosity in colon adenocarcinoma. *J Clin Pathol*. 2004;57(12):1312–8. <https://doi.org/10.1136/jcp.2004.017954>.
24. Toulas C, Mihura J, deBalincoourt C, Marques B, Marek E, Soula G, Roche H, Fabre G. Potential prognostic value in human breast cancer of cytosolic nme1 protein detection using an original hen specific antibody. *Br J Cancer*. 1996;73(5):630–5. <https://doi.org/10.1038/bjc.1996.109>.
25. Leonard MK, McCorkle JR, Snyder DE, Novak M, Zhang QB, Shetty AC, Mahurkar AA, Kaetzel DM. Identification of a gene expression signature associated with the metastasis suppressor function of nme1: prognostic value in human melanoma. *Lab Invest*. 2018;98(3):327–38. <https://doi.org/10.1038/labinvest.2017.108>.
26. Leone A, Seeger RC, Hong CM, Hu YY, Arboleda MJ, Brodeur GM, Stram D, Slamon DJ, Steeg PS. Evidence for nm23 rna overexpression, dna amplification and mutation in aggressive childhood neuroblastomas. *Oncogene*. 1993;8(4):855–65.
27. Oda Y, Naka T, Takeshita M, Iwamoto Y, Tsuneyoshi M. Comparison of histological changes and changes in nm23 and c-met expression between primary and metastatic sites in osteosarcoma: A clinicopathologic and immunohistochemical study. *Hum Pathol*. 2000;31(6):709–16. <https://doi.org/10.1053/hupa.2000.8230>.
28. Hsu CG, Lin LY, Ko JL, Yang SF, Chang H, Lin CY, Tsai HT, Chen SC, Chen SC, Wang PH. High expression of human nonmetastatic clone 23 type 1 in cancer of uterine cervix and its association with poor cell differentiation and worse overall survival. *J Surg Oncol*. 2008;98(6):448–56. <https://doi.org/10.1002/jso.21127>.
29. Varlotto JM, Yao AN, DeCamp MM, Ramakrishna S, Recht A, Flickinger J, Andrei A, Reed MF, Toth JW, Fitzgerald TJ, Higgins K, Zheng X, Shelkey J, Medford-Davis LN, Belani C, Kelsey CR. Nodal stage of surgically resected non-small cell lung cancer and its effect on recurrence patterns and overall survival. *Int J Radiat Oncol Biol Phys*. 2015;91(4):765–73. <https://doi.org/10.1016/j.ijrobp.2014.12.028>.
30. Otto T, Sicinski P. Cell cycle proteins as promising targets in cancer therapy. *Nat Rev Cancer*. 2017;17(2):93–115. <https://doi.org/10.1038/nrc.2016.138>.
31. Georgas K, Burrige L, Smith K, Holmes GP, Chenevix-Trench G, Ioannou PA, Little MH. Assignment of the human slit homologue slit2 to human chromosome band 4p15.2. *Cytogenet Cell Genet*. 1999;86(3–4):246–7. <https://doi.org/10.1159/000015351>.
32. Gonzalez DM, Medici D. Signaling mechanisms of the epithelial–mesenchymal transition. *Sci Signal*. 2014;7(344):8. <https://doi.org/10.1126/scisignal.2005189>.
33. Nieto MA, Huang RYJ, Jackson RA, Thiery JP. EMT: 2016. *Cell*. 2016;166(1):21–45. <https://doi.org/10.1016/j.cell.2016.06.028>.
34. Shi R, Yang Z, Liu W, Liu B, Xu Z, Zhang Z. Knockdown of slit2 promotes growth and motility in gastric cancer cells via activation of akt/beta-catenin. *Oncol Rep*. 2014;31(2):812–8. <https://doi.org/10.3892/or.2013.2887>.
35. Qin FX, Zhang HK, Ma L, Liu XL, Dai K, Li WL, Gu F, Fu L, Ma YJ. Low expression of slit2 and robo1 is associated with poor prognosis and brain-specific metastasis of breast cancer patients. *Sci Rep*. 2015; <https://doi.org/10.1038/srep14430>.
36. Belinsky SA. Silencing of genes by promoter hypermethylation: key event in rodent and human lung cancer. *Carcinogenesis*. 2005;26(9):1481–7. <https://doi.org/10.1093/carcin/bgi020>.
37. Tsou JA, Galler JS, Siegmund KD, Laird PW, Turla S, Cozen W, Hagen JA, Koss MN, Laird-Offringa IA. Identification of a panel of sensitive and specific DNA methylation markers for lung adenocarcinoma. *Mol Cancer*. 2007;6:70. <https://doi.org/10.1186/1476-4598-6-70>.
38. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristianson K, MacArthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Styrups K, Valsesia A, Walter K, Wei J. Wellcome Trust Case Control C, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME. Origins and functional impact of copy number variation in the human genome. *Nature*. 2010;464(7289):704–12. <https://doi.org/10.1038/nature08516>.
39. Tran TN, Selinger CI, Kohonen-Corish MRJ, McCaughan BC, Kennedy CW, O'Toole SA, Cooper WA. Fibroblast growth factor receptor 1 (fgfr1) copy number is an independent prognostic factor in non-small cell lung cancer. *Lung Cancer*. 2013;81(3):462–7. <https://doi.org/10.1016/j.lungc.2013.05.015>.
40. Qiu ZW, Bi JH, Gazdar AF, Song K. Genome-wide copy number variation pattern analysis and a classification signature for non-small cell lung cancer. *Genes Chromosomes Cancer*. 2017;56(7):559–69. <https://doi.org/10.1002/gcc.22460>.
41. Du H, Sun J, Chen Z, Nie J, Tong J, Li J. Cigarette smoke-induced failure of apoptosis resulting in enhanced neoplastic transformation in human bronchial epithelial cells. *J Toxicol Environ Health A*. 2012;75(12):707–20. <https://doi.org/10.1080/15287394.2012.690088>.
42. Huang H, Ji Y, Zhang J, Su Z, Liu M, Tong J, Ge C, Chen T, Li J. Aberrant dna methylation in radon and/or cigarette smoke-induced malignant transformation in beas-2b human lung cell line. *J Toxicol Environ Health A*. 2017;80(23–24):1321–30. <https://doi.org/10.1080/15287394.2017.1384156>.
43. Wang J, Yu X-F, Ouyang N, Zhao S, Yao H, Guan X, Tong J, Chen T, Li J-X. MicroRNA and mRNA interaction network regulates the malignant transformation of human bronchial epithelial cells induced by cigarette smoke. *Front Oncol*. 2019;9:1029. <https://doi.org/10.3389/fonc.2019.01029>.
44. Wei TYW, Juan CC, Hsiao JY, Su LJ, Lee YCG, Chou HY, Chen JMM, Wu YC, Chiu SC, Hsu CP, Liu KL, Yu CTR. Protein arginine methyltransferase 5 is a potential oncoprotein that upregulates g1 cyclins/cyclin-dependent kinases and the phosphoinositide 3-kinase/akt signaling cascade. *Cancer Sci*. 2012;103(9):1640–50. <https://doi.org/10.1111/j.1349-7006.2012.02367.x>.
45. Hou J, Aerts J, den Hamer B, van Ijcken W, den Bakker M, Riegman P, van der Leest C, van der Spek P, Foekens JA, Hoogsteden HC, Grosveld F, Philipsen S. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLoS ONE*. 2010;5(4):10312. <https://doi.org/10.1371/journal.pone.0010312>.
46. Su LJ, Chang CW, Wu YC, Chen KC, Lin CJ, Liang SC, Lin CH, Whang-Peng J, Hsu SL, Chen CH, Huang CYF. Selection of ddx5 as a novel internal control for q-rt-pcr from microarray data using a block bootstrap re-sampling scheme. *BMC Genom*. 2007; <https://doi.org/10.1186/1471-2164-8-140>.
47. Lu TP, Tsai MH, Lee JM, Hsu CP, Chen PC, Lin CW, Shih JY, Yang PC, Hsiao CK, Lai LC, Chuang EY. Identification of a novel biomarker, sema5a, for non-small cell lung carcinoma in nonsmoking women. *Cancer Epidemiol*

- Biomarkers Prev. 2010;19(10):2590–7. <https://doi.org/10.1158/1055-9965.EPI-10-0332>.
48. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biol.* 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8>.
 49. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):47. <https://doi.org/10.1093/nar/gkv007>.
 50. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics.* 2012;28(4):573–80. <https://doi.org/10.1093/bioinformatics/btr709>.
 51. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, Varambally S. Ualcan: A portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia.* 2017;19(8):649–58. <https://doi.org/10.1016/j.neo.2017.05.002>.
 52. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005;102(43):15545–50. <https://doi.org/10.1073/pnas.0506580102>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

