


# A two-phase epigenome-wide four-way gene–smoking interaction study of overall survival for early-stage non-small cell lung cancer

Leyi Chen<sup>1</sup> , Xiang Wang<sup>1</sup>, Ning Xie<sup>1</sup>, Zhongwen Zhang<sup>1</sup>, Xiaowen Xu<sup>1</sup>, Maojie Xue<sup>1,2</sup>, Yuqing Yang<sup>1</sup>, Liya Liu<sup>3</sup>, Li Su<sup>4,5</sup>, Maria Bjaanæs<sup>6</sup> , Anna Karlsson<sup>7</sup>, Maria Planck<sup>7</sup>, Johan Staaf<sup>7</sup>, Åslaug Helland<sup>6,8</sup>, Manel Esteller<sup>9,10,11,12</sup> , David C. Christiani<sup>4,5</sup>, Feng Chen<sup>1</sup>  and Ruyang Zhang<sup>1,13,14,15</sup> 

1 Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, China

2 Department of Health Inspection and Quarantine, Center for Global Health, School of Public Health, Nanjing Medical University, China

3 School of Public Health, Health Science Center, Ningbo University, China

4 Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA

5 Pulmonary and Critical Care Division, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

6 Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Norway

7 Division of Oncology, Department of Clinical Sciences Lund and CREATE Health Strategic Center for Translational Cancer Research, Lund University, Sweden

8 Institute of Clinical Medicine, University of Oslo, Norway

9 Josep Carreras Leukaemia Research Institute, Barcelona, Spain

10 Centro de Investigacion Biomedica en Red Cancer, Madrid, Spain

11 Institutio Catalana de Recerca i Estudis Avançats, Barcelona, Spain

12 Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona, Spain

13 China International Cooperation Center for Environment and Human Health, Nanjing Medical University, China

14 Changzhou Medical Center, Nanjing Medical University, Changzhou, China

15 Information Center, The Affiliated Changzhou Second People's Hospital of Nanjing Medical University, Changzhou, China

## Keywords

epigenomics; four-way interaction;  
gene–smoking interaction; NSCLC

## Correspondence

R. Zhang, Department of Biostatistics,  
Center for Global Health, School of Public  
Health, Nanjing Medical University, SPH  
Building Room 406, 101 Longmian Avenue,  
Nanjing, Jiangsu 211166, China  
Tel: +86 025 86868436

E-mail: [zhangruyang@njmu.edu.cn](mailto:zhangruyang@njmu.edu.cn)

F. Chen, Department of Biostatistics, Center  
for Global Health, School of Public Health,  
Nanjing Medical University, SPH Building  
Room 412, 101 Longmian Avenue, Nanjing,  
Jiangsu 211166, China

High-order interactions associated with non-small cell lung cancer (NSCLC) survival may elucidate underlying molecular mechanisms and identify potential therapeutic targets. Our previous work has identified a three-way interaction among pack-year of smoking (the number of packs of cigarettes smoked per day multiplied by the number of years the person has smoked) and two DNA methylation probes (cg05293407<sub>TRIM27</sub> and cg00060500<sub>KIAA0226</sub>). However, whether a four-way interaction exists remains unclear. Therefore, we adopted a two-phase design to identify the four-way gene–smoking interactions by a hill-climbing strategy on the basis of the previously detected three-way interaction. One CpG probe, cg16658473<sub>SHISA9</sub>, was identified with  $FDR-q \leq 0.05$  in the discovery phase and  $P \leq 0.05$  in the validation phase. Meanwhile, the four-way interaction improved the discrimination ability for the prognostic prediction model, as indicated by the area under the receiver operating characteristic curve (AUC) for both 3- and 5-year survival. In summary, we identified a

## Abbreviations

AUC, area under the receiver operating characteristic curve; CI, confidence interval; C-index, concordance index; FDR, false discovery rate; G × E interaction, gene–environment interaction; G × G interaction, gene–gene interaction; GEO, Gene Expression Omnibus; GO, gene ontology; HR, hazard ratio; IKKε, the inhibitor of nuclear factor κB kinase subunit epsilon; KEGG, Kyoto Encyclopedia of Genes and Genomes; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; NSCLC, non-small cell lung cancer; QC, quality control; ROC, receiver operating characteristic; SD, standard deviation; SNP, single nucleotide polymorphisms; TCGA, The Cancer Genome Atlas; TFEB, transcription factor e-box binding; TRIUNE, integrated covariates and four-way interaction score.

Tel: +86 025 86868414  
E-mail: [fengchen@njmu.edu.cn](mailto:fengchen@njmu.edu.cn)

Leyi Chen, Xiang Wang, Ning Xie, Zhongwen Zhang and Xiaowen Xu contributed equally to this article. David C. Christiani is the senior author who supervised the work.

(Received 3 August 2024, revised 5 October 2024, accepted 7 November 2024, available online 4 December 2024)

doi:10.1002/1878-0261.13766

four-way interaction associated with NSCLC survival among pack-year of smoking, cg05293407<sub>TRIM27</sub>, cg00060500<sub>KIAA0226</sub> and g16658473<sub>SHISA9</sub>, providing novel insights into the complex mechanisms underlying NSCLC progression.

## 1. Introduction

Lung cancer is one of the most prevalent malignancies worldwide and a leading cause of cancer-related mortality. According to global cancer statistics, almost 2.5 million new cases are diagnosed and 1.8 million new deaths are registered annually [1]. Most of lung cancer patients were non-small-cell lung cancer (NSCLC) [2,3], of which two major histological subtypes were lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) [4,5]. Nowadays, lung cancer survival rates remain low, with a 3-year survival rate of 13–40% and a 5-year rate around 25% [2]. However, various clinical outcomes are observed among early-stage NSCLC patients, even among those with similar clinical characteristics [6]. This indicates that there are still hidden mechanisms behind the development of lung cancer.

Epigenetic alterations, notably DNA methylation, are recognized as key factors regulating molecular mechanisms of tumorigenesis and cancer progression [7,8]. As a reversible epigenetic modification process, DNA methylation primarily involves the substitution of the 5'-hydroxyl group of DNA with a methyl group, known as 5-methylcytosine (5mC). This modification holds promise as a tool for early detection [9,10], providing pivotal insights into tumor metastasis and prognosis [11,12] and guiding therapeutic strategies, including immunotherapy and chemotherapy [6,13,14], for NSCLC patients. Additionally, gene–gene ( $G \times G$ ) interactions between methylation traits have been reported for their pivotal insights into the biological mechanisms underlying NSCLC [15–17]. Over the past two decades, epigenetic characteristics have been proved to be influenced by environmental exposure such as smoking, which affects the course of cancer progression [18]. This complex interaction pattern

between smoking and DNA methylation, referred to as the gene–environment ( $G \times E$ ) interaction, parallels  $G \times G$  interactions, both recognized for their crucial roles in elucidating the intricate molecular mechanisms underlying various complex diseases [19,20], including NSCLC [21–23]. Our previous studies have found the patterns under two-way interaction (pack-year of smoking  $\times$  cg05293407<sub>TRIM27</sub>) and three-way interaction (pack-year of smoking  $\times$  cg05293407<sub>TRIM27</sub>  $\times$  cg0060500<sub>KIAA0226</sub>) between DNA methylation and smoking. These patterns may reveal the histologically heterogeneous effect of molecular mechanisms of cancer development in NSCLC [22,23]. Considering that the progression of tumors constitute a complex biological process characterized by multifaceted, multistep, multigenic, and multifactorial mechanisms, low-order interactions are limited in correctly interpreting complex composite patterns related to cancer progression [24]. Higher-order interactions deserve to be tested to confirm our previous work and explore new interactions, if they exist [25,26]. Therefore, we performed the first epigenome-wide four-way gene-smoking interaction study of NSCLC survival and identified one novel epigenetic biomarker associated with lung cancer prognosis.

In this study, we integrated the epigenomic and demographic data of multiple cohorts and employed a two-phase study to elucidate the robust four-way interaction influencing the overall survival of NSCLC patients. First, build upon the previously established three-way interaction involving cg05293407<sub>TRIM27</sub>, cg00060500<sub>KIAA0226</sub>, and smoking, we conducted a four-way interaction analysis involving pack-year of smoking, cg05293407<sub>TRIM27</sub>, cg00060500<sub>KIAA0226</sub> and an additional CpG probe, using samples from four

international study centers (USA-Harvard, Spain, Norway, Sweden). Second, we validated these significant signals by data obtained from the Cancer Genome Atlas (TCGA) database.

## 2. Materials and methods

### 2.1. Study populations of DNA methylation data

The DNA methylation data from early-stage (stage I and II) NSCLC patients were collected from five international consortium: USA-Harvard, Spain, Norway, Sweden study and TCGA database [27–30]. Cases from USA-Harvard, Spain, Norway, and Sweden study were assigned to discovery phase, while cases from TCGA were assigned to validation phase. All these studies were approved by each institutional review board, and patients provided written informed consent. The study methodologies conformed to the standards set by the Declaration of Helsinki and was approved by the local ethics committee.

#### 2.1.1. USA-Harvard

The Harvard Lung Cancer Study cohort was described previously [27]. Patients were recruited at Massachusetts General Hospital between 1992 and 2016, and all were newly diagnosed and histologically confirmed with primary NSCLC at the time of recruitment. Snap-frozen tumor samples were taken from patients during complete resection of the therapeutic procedure. A total of 149 early-stage patients were selected in this study with complete survival information. Tumor DNA was extracted from 5- $\mu$ m-thick histopathologic sections. Each specimen was evaluated by a pathologist for amount (tumor cellularity > 70%) and quality of tumor cells. All specimens were histologically classified using the World Health Organization (WHO) criteria. The study was approved by the Institutional Review Boards of the Massachusetts General Hospital (Partners Human Research Committee, Protocol #1999P004935/MGH).

#### 2.1.2. Spain

As previously described [28], tumors were surgically collected from 207 early-stage NSCLC patients between 1991 and 2009. These samples were obtained from several research institutions, including Spain (Catalan Institute of Oncology; Center for Applied Medical Research; and Bellvitge Biomedical Research Institute), Italy (IRCCS Foundation National Cancer Institute; and University of Turin), UK (University of

Liverpool Cancer Research Centre), France (CHU Albert Michallon), and the United States (University of Michigan Medical School). Patients provided written consent and tumors were surgically collected. This study was approved by the Bellvitge Biomedical Research Institute institutional review boards (PR055/10).

#### 2.1.3. Norway

There were 132 LUAD patients participated in Norway cohort with operable lung cancer tumors enrolled at Oslo University Hospital, Rikshospitalet, Norway, between 2006 and 2011 [29]. The project was developed with the approval of the Oslo University Institutional Review Board and regional ethics committee (S-05307). All patients provided informed consent. Tumor tissues were snap-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until DNA isolation.

#### 2.1.4. Sweden

Tumor DNA were collected from 36 patients with early-stage NSCLC, at Skane University Hospital, Lund, Sweden [30]. DNA extraction was performed on tumor specimens (10  $\mu$ m-thick, tumor cellularity > 50%). The study was developed under the approval of the Regional Ethical Review Board in Lund, Sweden (Registration no. 2004/762 and 2008/702).

#### 2.1.5. TCGA

TCGA database consisted of 227 LUAD and 241 LUSC cases with complete DNA methylation, overall survival time and covariates data. Level-1 HumanMethylation450 DNA methylation data (image data) of each patient were downloaded at October 1, 2015.

## 2.2. Quality control procedures for DNA methylation data

The quality control (QC) procedures outlined here have been consistently applied in our previous studies [16,22,23]. DNA methylation was profiled using the Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, CA, USA). All studies adhered to the same QC procedures before the association analysis. Raw image data were converted into beta values (continuous numbers ranging from 0% to 100%) using the GenomeStudio Methylation Module V1.8 (Illumina Inc.) for background subtraction and control normalization. Unqualified probes meeting any one of the following criteria were excluded: (a) failed detection ( $P > 0.05$ ) in

more than 5% samples; (b) coefficient of variance (CV) < 5%; (c) methylated or unmethylated in all samples; (d) common single nucleotide polymorphisms (SNP) located in the probe sequence or 10-bp flanking regions; (e) cross-reactive probes or cross-hybridizing probes [31]; or (f) did not pass quality control in all centers. Samples with > 5% undetectable probes or missing of pack-year of smoking were excluded. Methylation signals were further processed for quantile normalization (*betan* function in R package *minfi*) as well as type I and II probes correction (*BMIQ* function in R package *lumi*), and were adjusted for batch effects (*ComBat* function in R package *sva*) according to the best pipeline by a comparative study [32]. The QC processes are detailed in Fig. S1.

### 2.3. Study populations and quality control of gene expression data

Gene expression data for early-stage NSCLC were derived from the USA-Harvard, Norway, Sweden, and TCGA database. While, Spanish study center did not profile gene expression data. The USA-Harvard, Norway, Sweden and TCGA database had 26, 93, 34, 456 individuals, respectively. The TCGA workgroup completed the mRNA sequencing data processing and QC. Level 3 gene quantification data were downloaded from the TCGA data portal and were further checked for quality. Gene probes were excluded if the missing rate > 80%, and batch effects were corrected with *ComBat*. The expression value of each gene was transformed on  $\log_2$  scale and standardized subsequently.

### 2.4. Study design and statistical analysis

#### 2.4.1. A two-phase design for four-way gene-smoking interaction study

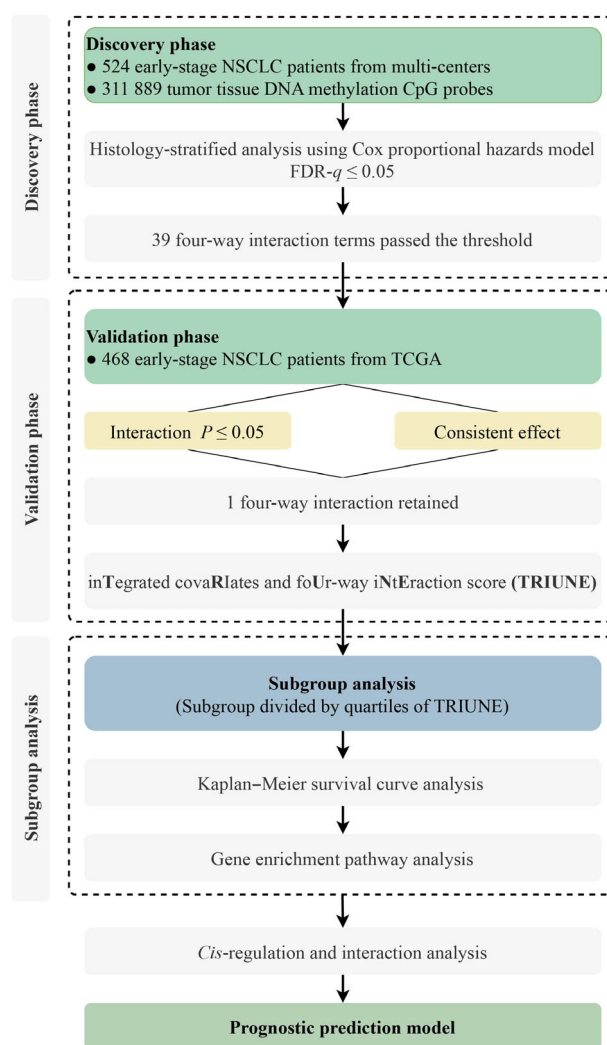
The flowchart illustrating the analysis was depicted in Fig. 1, outlining a two-phase design aimed at exploring

the association between four-way interactions and early-stage NSCLC overall survival on an epigenome-wide scale. In the discovery phase, we constructed a histology-stratified Cox proportional hazards model adjusted for age, sex, smoking status, clinical stage, and study centers. In current study, we employed a hill-climbing algorithm to detect high-order interactions. Originally, we identified a significant two-way interaction, i.e., pack-year of smoking  $\times$  cg05293407<sub>TRIM27</sub>, associated with NSCLC survival [22]. Then, we further recognized a significant three-way interaction, i.e., pack-year of smoking  $\times$  cg05293407<sub>TRIM27</sub>  $\times$  cg0060500<sub>KIAA0226</sub> [23]. Here, we aimed to identify another biomarker contributing to the significant four-way interaction, i.e., pack-year of smoking  $\times$  cg05293407<sub>TRIM27</sub>  $\times$  cg0060500<sub>KIAA0226</sub>  $\times$  CpG probe, among samples from USA-Harvard, Spain, Norway, and Sweden. This algorithm allows us to systematically explore significant signals from low-order interaction to high-order interaction [22,23,33]. Hazard ratio (HR) and 95% confidence interval (CI) was computed for per 1% level of the methylation increment. Multiple testing corrections were applied to control the false discovery rate (FDR) at a 5% level [34]. In the validation phase, the identified four-way interactions will be tested in TCGA dataset. Significant interactions should meet the following criteria: (a)  $\text{FDR-}q \leq 0.05$  in the discovery phase and  $P \leq 0.05$  in the validation phase; and (b) with consistent direction effect between two phases.

#### 2.4.2. TRIUNE subgroup analysis

An integrated covariates and four-way interaction score (TRIUNE) was defined by a weighted linear combination of all components inside the significant four-way interaction, as well as all covariates, with weights derived from the aforementioned Cox proportional hazards model. Specifically,

$$\begin{aligned} \text{TRIUNE} = & \sum \alpha_i \times \text{Covariate}_i + \beta_1 \times \text{Pack-year of smoking} + \beta_2 \times \text{cg05293407}_{\text{TRIM27}} + \beta_3 \times \text{cg0060500}_{\text{KIAA0226}} + \beta_4 \times \text{cg16658473}_{\text{SHISA9}} \\ & + \beta_{12} \times \text{Pack-year of smoking} \times \text{cg05293407}_{\text{TRIM27}} + \beta_{13} \times \text{Pack-year of smoking} \times \text{cg0060500}_{\text{KIAA0226}} \\ & + \beta_{14} \times \text{Pack-year of smoking} \times \text{cg16658473}_{\text{SHISA9}} \\ & + \beta_{23} \times \text{cg05293407}_{\text{TRIM27}} \times \text{cg0060500}_{\text{KIAA0226}} + \beta_{24} \times \text{cg05293407}_{\text{TRIM27}} \times \text{cg16658473}_{\text{SHISA9}} \\ & + \beta_{34} \times \text{cg0060500}_{\text{KIAA0226}} \times \text{cg16658473}_{\text{SHISA9}} \\ & + \beta_{123} \times \text{Pack-year of smoking} \times \text{cg05293407}_{\text{TRIM27}} \times \text{cg0060500}_{\text{KIAA0226}} \\ & + \beta_{124} \times \text{Pack-year of smoking} \times \text{cg05293407}_{\text{TRIM27}} \times \text{cg16658473}_{\text{SHISA9}} \\ & + \beta_{234} \times \text{cg05293407}_{\text{TRIM27}} \times \text{cg0060500}_{\text{KIAA0226}} \times \text{cg16658473}_{\text{SHISA9}} \\ & + \beta_{1234} \times \text{Pack-year of smoking} \times \text{cg05293407}_{\text{TRIM27}} \times \text{cg0060500}_{\text{KIAA0226}} \times \text{cg16658473}_{\text{SHISA9}} \end{aligned}$$



**Fig. 1.** Flow chart of study design and statistical analysis. Patients diagnosed with lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) from the Harvard, Spain, Norway, and Sweden cohorts were utilized in the discovery phase for screening, whereas data from the Cancer Genome Atlas (TCGA) were employed for validation. Subgroup analyses were conducted based on quartiles of TRIUNE. FDR, false discovery rate; NSCLC, non-small lung cancer.

where  $\alpha_i$  and  $\beta_i$  were the coefficient estimated for each variable in the multivariable Cox proportional hazards model.

To further evaluate the prognostic value of the TRIUNE, we performed the following analyses. (a) Patients were divided into four subgroups by the quartiles of the TRIUNE, and their survival difference was compared using Kaplan–Meier survival curves. (b) To explore the molecular mechanisms underlying the TRIUNE, potential genes regulated by CpG probes involving in TRIUNE were identified by genome-wide DNA methylation and gene expression correlation analysis using the linear regression model adjusted for the covariates aforementioned, with significant significance at

FDR- $q \leq 0.05$ . (c) Functional annotation and gene enrichment pathway analyses, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG), were conducted to evaluate potential biological functions of significantly correlated genes using the *WebGestaltR* package [35–38]. FDR- $q \leq 0.05$  was considered statistically significant.

#### 2.4.3. Four-way gene-smoking interaction analysis in gene expression level

We examined the four-way gene-smoking interaction in gene expression level using genes *cis*-regulated by CpG probes. Specifically, the CpG probe located



within 1 kb window of a gene and significantly ( $P \leq 0.05$ ) correlated with its expression were defined as candidate having significant epigenetic *cis*-regulation of transcription. The four-way gene-smoking interaction at the gene expression level was tested by Cox model adjusted for the covariates aforementioned. Significant threshold was defined as  $P \leq 0.05$ .

#### 2.4.4. Development of a prognostic prediction model

We utilized clinical variables and CpG probes involved in the four-way interaction to develop a prognostic prediction model in the combined dataset. The discrimination ability of the prediction model was illustrated using time-dependent receiver operating characteristic (ROC) curves, with the area under the ROC curve (AUC) calculated by *timeROC* package. The 3- and 5-year overall survival time of cancer patients are well recognized as important clinical outcomes in practice [2,39,40]. Therefore, we gave 3- and 5-year ROC and AUC to present model discrimination. We applied DeLong test to compare the AUCs of two models with and without the four-way interaction. Also, the concordance index (C-index), an average accuracy of predictive survival across follow-up years, which ranges from 0.5 to 1.0, was calculated to estimate the predictive performance.

#### 2.4.5. Statistical analysis

Continuous variables were expressed as mean  $\pm$  standard deviation (SD), and categorical variables were expressed in frequency (*n*) and proportion (%). Multiple comparisons were adjusted using the false discovery rate (FDR) method, as measured by the FDR-*q* value (false discovery rate-adjusted *q* value) [34], to control the overall false-positive rate at 5% level. Statistical analysis was performed using R version 4.3.1 (The R Foundation of Statistical Computing, <https://www.r-project.org/>).

### 3. Results

#### 3.1. Sample characteristics of the study population

After quality control, our epigenome-wide DNA methylation data were composed of 311 891 CpG probes from 992 early-stage NSCLC patients; 524 patients ( $N_{\text{LUAD}} = 425$  and  $N_{\text{LUSC}} = 99$ ) were in the discovery phase and 468 patients ( $N_{\text{LUAD}} = 227$  and  $N_{\text{LUSC}} = 241$ ) were in the validation phase. A total of 609 patients had gene expression data. Demographic

and clinical information for these patients is detailed in Table 1 and Table S1.

#### 3.2. A significant four-way interaction involving cg16658473<sub>SHISA9</sub> identified in a two-phase study

In the discovery phase, 39 pairs of significant four-way interactions were identified ( $\text{FDR-}q \leq 0.05$ ); whereas only one pair of interactions remained significant in the validation phase ( $P \leq 0.05$ ) and showed robust results in the combined data (Table S2). There was a statistically significant four-way interaction effect of cg16658473<sub>SHISA9</sub> with pack-years of smoking, cg05293407<sub>TRIM27</sub>, and cg00060500<sub>KIAA0226</sub> on NSCLC survival (discovery set:  $\text{HR}_{\text{interaction}} = 0.9993$ , 95% CI: 0.9990–0.9996,  $P = 3.08 \times 10^{-6}$ ,  $\text{FDR-}q = 0.027$ ; validation set:  $\text{HR}_{\text{interaction}} = 0.9992$ , 95% CI: 0.9986–0.9998,  $P = 0.014$ ; combined data:  $\text{HR}_{\text{interaction}} = 0.9995$ , 95% CI: 0.9993–0.9997,  $P = 3.06 \times 10^{-6}$ ). However, the identified CpG probe (cg16658473) did not have significant marginal effects (Table S3). Annotation information for three CpG probes located in the genes *TRIM27*, *KIAA0226*, and *SHISA9* was presented in Table S4.

To visualize the four-way interactions, we generated a 3D figure. Samples were divided into two subgroups based on the median level of cg00060500<sub>KIAA0226</sub>. Then, the effect of cg16658473<sub>SHISA9</sub> on NSCLC survival measured by HR at different combinations of pack-year of smoking and cg05293407<sub>TRIM27</sub> methylation level was exhibited as a 3-D surface in the subgroup with high and low expression of cg00060500<sub>KIAA0226</sub>, respectively (Fig. 2). In the low expression subgroup of cg00060500<sub>KIAA0226</sub>, as smoking intensity increases and methylation levels of cg05293407<sub>TRIM27</sub> decrease, we observed an increased risk associated with cg16658473<sub>SHISA9</sub>. Conversely, in the high expression subgroup of cg00060500<sub>KIAA0226</sub>, as smoking intensity decreases and methylation levels of cg05293407<sub>TRIM27</sub> increase, the effects of cg16658473<sub>SHISA9</sub> are completely reversed.

#### 3.3. Significant survival heterogeneity among TRIUNE subgroups

As depicted in Fig. 3, patients were stratified into four distinct subgroups according to the quartiles of TRIUNE, wherein Group 1 and Group 4 denoted as the subgroup with the smallest and largest risk, respectively. For median survival years, Group 1 was 5.63 years, followed by a decrease in each subsequent group, with the Group 4 having the shortest median survival (0.94, 95% CI: 0.75–1.18). Furthermore, in line with this trend, patients

**Table 1.** Demographic and clinical descriptions for early-stage NSCLC patients with DNA methylation data in five international study centers. 95% CI, 95% confidence interval; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; TCGA, The Cancer Genome Atlas.

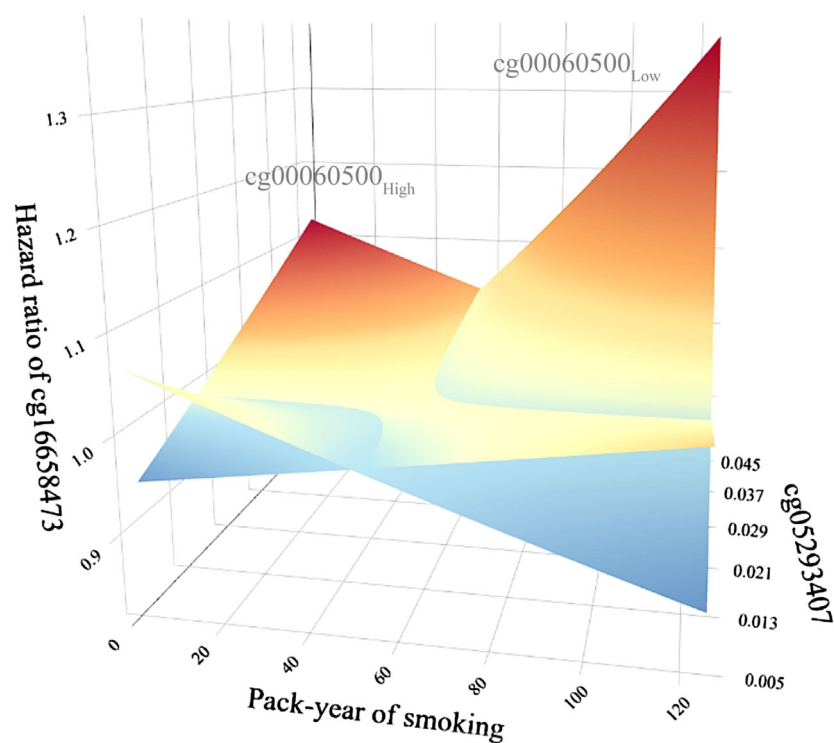
Variable	Discovery phase					Validation phase	Combined dataset
	USA-Harvard (N = 149)	Spain <sup>a</sup> (N = 207)	Norway (N = 132)	Sweden (N = 36)	All (N = 524)	TCGA (N = 468)	Overall (N = 992)
Age (years)	67.78 ± 9.92	65.77 ± 10.66	65.4 ± 9.28	72.25 ± 7.22	66.7 ± 10.04	66.52 ± 9.3	66.61 ± 9.69
Pack-year of smoking	51.49 ± 40.62	43.24 ± 31.72	26.93 ± 17.62	22.64 ± 25.19	40.06 ± 33.00	47.05 ± 28.40	43.36 ± 31.10
Sex							
Female	66 (44.30%)	99 (47.83%)	70 (53.03%)	22 (61.11%)	257 (49.05%)	185 (39.53%)	442 (44.56%)
Male	83 (55.70%)	108 (52.17%)	62 (46.97%)	14 (38.89%)	267 (50.95%)	283 (60.47%)	550 (55.44%)
Smoking status							
Never	18 (12.08%)	28 (13.53%)	16 (12.12%)	18 (50.00%)	80 (15.27%)	0 (0.00%)	80 (8.06%)
Ever	81 (54.36%)	113 (54.59%)	74 (56.06%)	11 (30.56%)	279 (53.24%)	323 (69.02%)	602 (60.69%)
Current	50 (33.56%)	66 (31.88%)	42 (31.82%)	7 (19.44%)	165 (31.49%)	145 (30.98%)	310 (31.25%)
Clinical stage							
I	102 (68.46%)	167 (80.68%)	92 (69.70%)	34 (94.44%)	395 (75.38%)	301 (64.32%)	696 (70.16%)
II	47 (31.54%)	40 (19.32%)	40 (30.30%)	2 (5.56%)	129 (24.62%)	167 (35.68%)	296 (29.84%)
Histology							
LUAD	96 (64.43%)	169 (81.64%)	132 (100.00%)	28 (77.78%)	425 (81.11%)	227 (48.50%)	652 (65.73%)
LUSC	53 (35.57%)	38 (18.36%)	0 (0.00%)	8 (22.22%)	99 (18.89%)	241 (51.50%)	340 (34.27%)
Chemotherapy							
No	140 (93.96%)	166 (91.21%)	101 (76.52%)	25 (89.29%)	432 (87.98%)	151 (75.88%)	583 (84.49%)
Yes	9 (6.04%)	16 (8.79%)	31 (23.48%)	3 (10.71%)	59 (12.02%)	48 (24.12%)	107 (15.51%)
Unknown	0	25	0	8	33	269	302
Radiotherapy							
No	130 (87.25%)	172 (94.51%)	131 (99.24%)	28 (100.00%)	461 (93.89%)	190 (95.48%)	651 (94.35%)
Yes	19 (12.75%)	10 (5.49%)	1 (0.76%)	0 (0.00%)	30 (6.11%)	9 (4.52%)	39 (5.65%)
Unknown	0	25	0	8	33	269	302
Adjuvant therapy <sup>b</sup>							
No	125 (83.89%)	157 (86.26%)	100 (75.76%)	25 (89.29%)	407 (82.89%)	146 (73.37%)	553 (80.14%)
Yes	24 (16.11%)	25 (13.74%)	32 (24.24%)	3 (10.71%)	84 (17.11%)	53 (26.63%)	137 (19.86%)
Unknown	0	25	0	8	33	269	302
Survival year <sup>c</sup>							
Median (95% CI)	6.61 (5.14–7.49)	3.83 (3.21–4.46)	5.40 (5.16–5.77)	3.25 (2.09–4.39)	5.01 (4.53–5.25)	0.58 (0.50–0.70)	2.31 (2.00–2.64)
Censoring rate	0.1879	0.5459	0.6894	0.5278	0.479	0.7607	0.6119

<sup>a</sup>Spain center is a collaborative study center, containing samples from Spain, Italy, the UK, France, and the United States.<sup>b</sup>Adjuvant therapy included chemotherapy and/or radiotherapy.<sup>c</sup>Restricted mean survival time was given because median was not available; proportion of samples lost to follow-up or alive at the end of study.

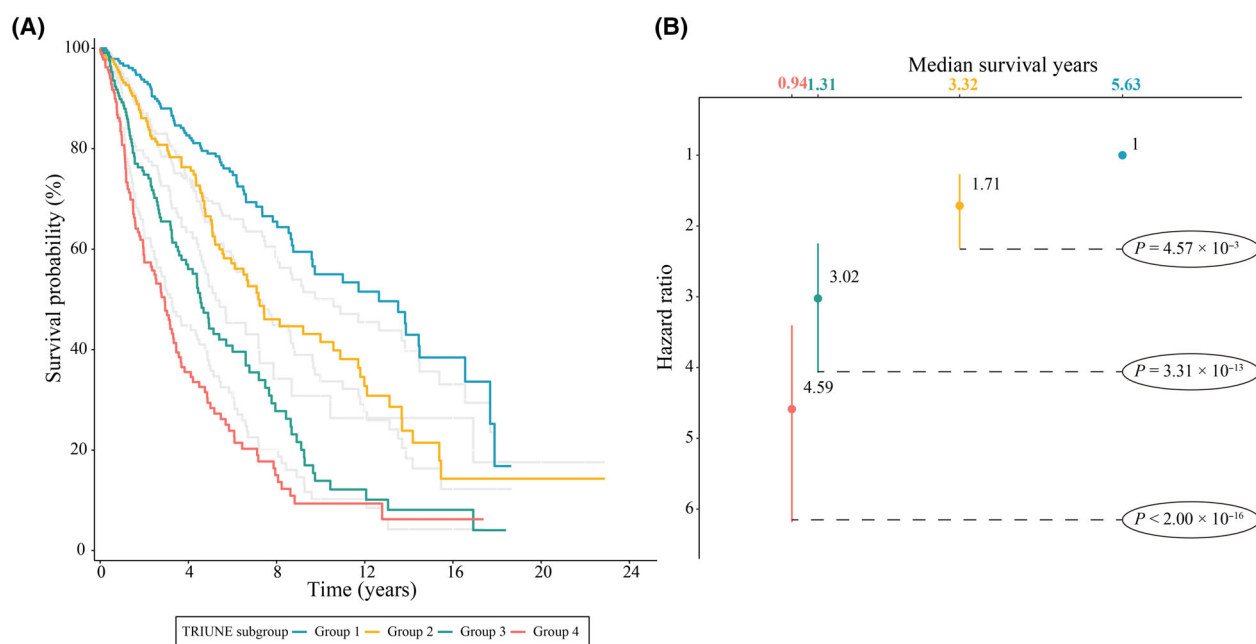
with higher TRIUNE exhibited a higher risk of mortality ( $HR_{4 \text{ vs } 1} = 4.59$ , 95% CI: 3.40–6.18,  $P < 2.00 \times 10^{-16}$ ;  $HR_{3 \text{ vs } 1} = 3.02$ , 95% CI: 2.25–4.07,  $P = 3.31 \times 10^{-13}$ ;  $HR_{2 \text{ vs } 1} = 1.71$ , 95% CI: 1.27–2.31,  $P = 4.57 \times 10^{-4}$ ). For comparison, we also divided patients into four subgroups based on the clinical score, a weighted linear combination of clinical variables, as shown in gray lines in Fig. 3. The risk stratification ability of TRIUNE was obviously superior to that of clinical scores, and the association between TRIUNE and overall survival remained significant in all subgroup analyses stratified by different clinical factors (Fig. 4), indicating it was a robustly significant score of lung cancer survival.

### 3.4. *Cis*-regulatory genes were enriched in biological pathways

Genome-wide correlation analyses identified 892 genes whose expression levels were significantly associated with the DNA methylation levels of these CpG probes involved in TRIUNE ( $FDR-q \leq 0.05$ ). These genes were further found to be enriched in 20 KEGG pathways, which were primarily related to immune system processes (Fig. S2A). Moreover, GO enrichment analysis identified 83 significant biological process pathways, 8 cellular component pathways, and 12 molecular function pathways, many of which

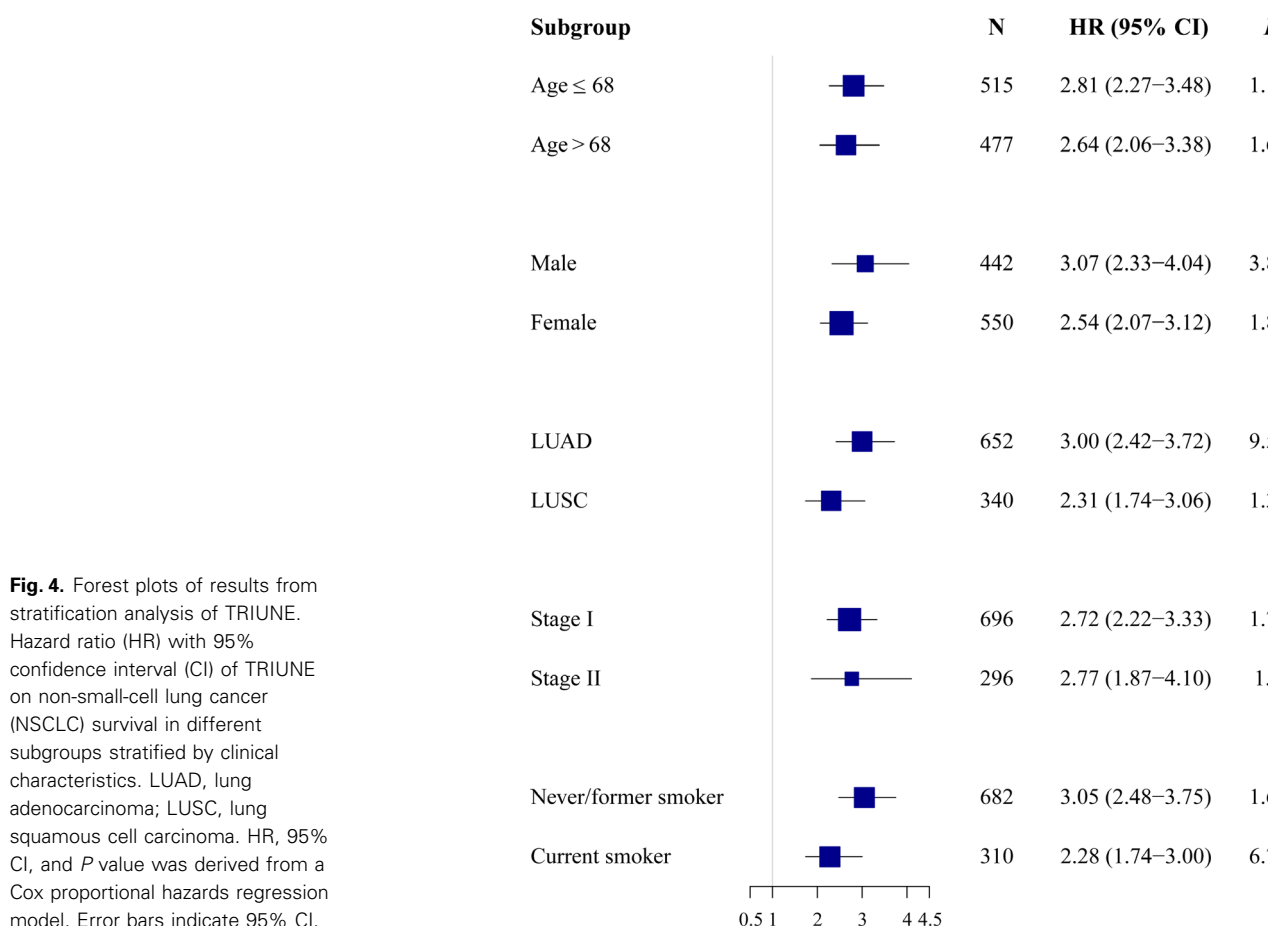


**Fig. 2.** The 3D figure for visualization of four-way interactions. Results of four-way interaction among pack-years of smoking and cg16658473<sub>SHISA9</sub>, cg05293407<sub>TRIM27</sub>, and cg00060500<sub>KIAA0226</sub>. The varying effects of cg16658473<sub>SHISA9</sub> measured using HR at different smoking and cg05293407<sub>TRIM27</sub> methylation levels in the subgroup with low and high level of cg00060500<sub>KIAA0226</sub>, respectively. HR, hazard ratio.



**Fig. 3.** Comparison of overall survival time of patients among TRIUNE subgroups. (A) Kaplan–Meier survival curves for patients grouped by TRIUNE. Patients were categorized into four subgroups by using the quartiles of TRIUNE as the cutoffs. The number of patients in Group 1 to Group 4 was 248. The gray lines indicate the groups of patients based on the quartiles of the risk scores derived from a clinical model without the addition of the TRIUNE. (B) Hazard ratios and  $P$  values were derived from the Cox proportional hazards model for four groups of patients, where Group 1 was set to be reference group. Error bars indicate 95% confidence interval (CI).





**Fig. 4.** Forest plots of results from stratification analysis of TRIUNE. Hazard ratio (HR) with 95% confidence interval (CI) of TRIUNE on non-small-cell lung cancer (NSCLC) survival in different subgroups stratified by clinical characteristics. LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma. HR, 95% CI, and *P* value was derived from a Cox proportional hazards regression model. Error bars indicate 95% CI.

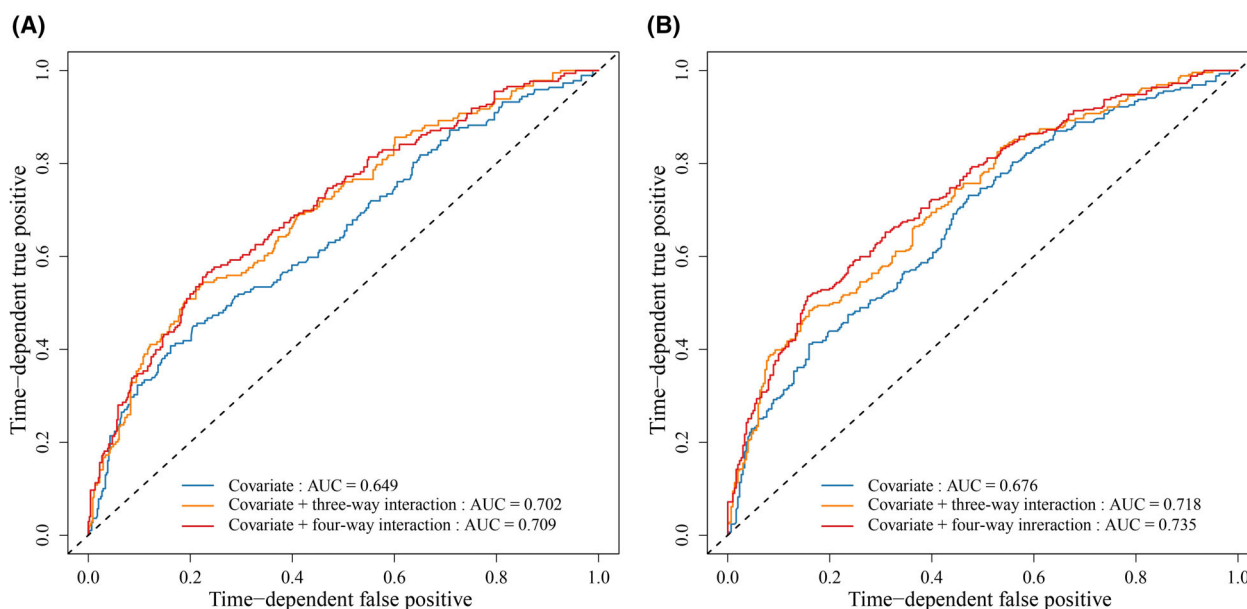
are related to immune system regulation (Fig. S2B–D).

### 3.5. A significant four-way interaction involving *TRIM27*, *FYTTD1*, and *SHISA9* identified in genes *cis*-regulated by CpG probes

We observed that *TRIM27*, *KIAA0226*, *FYTTD1*, and *SHISA9* were significantly *cis*-regulated by CpG probes within 1 kb window (Table S5). For CpG probe cg00060500, it was mapped to both *TRIM27* and *KIAA0226*. Meanwhile, there was a significant correlation ( $r = 0.455$ ,  $P = 5.21 \times 10^{-33}$ ) between *KIAA0226* and *FYTTD1* (Fig. S3). We did not observe significant ( $P = 0.530$ ) four-way interaction among pack-year of smoking, *TRIM27*, *KIAA0226*, and *SHISA9* (Table S6). Anyway, there is significant four-way interaction among pack-year of smoking, *TRIM27*, *FYTTD1*, and *SHISA9* (Table S7).

### 3.6. Four-way interaction empowered prognostic prediction model

As depicted in Fig. 5, initially, the predictive capacity of the model comprising solely demographic and clinical variables was found to be limited ( $AUC_{3\text{-year}} = 0.649$ , 95% CI: 0.601–0.694;  $AUC_{5\text{-year}} = 0.677$ , 95% CI: 0.633–0.719, *C*-index = 0.625). Nonetheless, a slight but discernible improvement of AUC was observed with the inclusion of all elements of the three-way interaction (pack-year of smoking  $\times$  cg05293407<sub>*TRIM27*</sub>  $\times$  cg00060500<sub>*KIAA0226*</sub>), with  $AUC_{3\text{-year}} = 0.702$  (95% CI: 0.656–0.745),  $AUC_{5\text{-year}} = 0.718$  (95% CI: 0.676–0.758), *C*-index = 0.661. Moreover, the AUC was further improved through the incorporation of all elements of the four-way interaction (pack-year of smoking  $\times$  cg05293407<sub>*TRIM27*</sub>  $\times$  cg00060500<sub>*KIAA0226*</sub>  $\times$  cg16658473<sub>*SHISA9*</sub>), with  $AUC_{3\text{-year}} = 0.709$  (95% CI: 0.663–0.753),  $AUC_{5\text{-year}} = 0.735$  (95% CI: 0.693–0.775), and



**Fig. 5.** ROC curves were generated to compare the predictive performance of different models. Three models were: (a) one using only covariates (clinical variables); (b) one incorporating covariates and all elements of three-way interactions (cg05293407<sub>TRIM27</sub>, pack-year of smoking and another CpG probe) with  $FDR-q \leq 0.05$ ; and (c) another one including covariates and a four-way interaction (pack-year of smoking, cg05293407<sub>TRIM27</sub>, cg00060500<sub>KIAA0226</sub> and cg16658473<sub>SHISA9</sub>) with  $FDR-q \leq 0.05$ . (A) ROC for 3-year survival prediction. (B) ROC for 5-year survival prediction. AUC, area under the receiver operating characteristic curve; ROC, receiver operating characteristic. The dashed line represents the performance of a random classifier, corresponding to an AUC of 0.5.  $FDR-q$ , false discovery rate-adjusted  $q$  value.

$C$ -index = 0.673. The addition of the four-way interaction term, compared to the clinical model, resulted in AUC increases of 9.23% ( $P = 1.74 \times 10^{-4}$ ) and 8.52% ( $P = 2.00 \times 10^{-4}$ ) for 3- and 5-year survival, respectively.

#### 4. Discussion

To our knowledge, we were among the first to conduct an epigenomic study on four-way gene-smoking interactions related to overall survival in early-stage NSCLC patients. We identified one CpG probe (cg16658473<sub>SHISA9</sub>), by interacting with cg05293407<sub>TRIM27</sub>, cg00060500<sub>KIAA0226</sub> and pack-year of smoking, displaying a significant four-way interaction effect on the overall survival of NSCLC. Moreover, the TRI-UNE we developed revealed substantial survival heterogeneity among subgroups. We also identified a four-way gene-smoking interaction in gene expression level using genes *cis*-regulated by CpG probes. Finally, the four-way interaction moderately, but significantly, improved the discrimination ability of the prognostic prediction model.

Statistical  $G \times G$  and  $G \times E$  interactions, particularly at the epigenetic level, are crucial for understanding

the molecular mechanisms underlying cancer progression [15,16,21–23,33,41–46]. Taking the two-way interaction as an example, the statistical interaction is the deviation between the joint effect of two factors and sum of their main effects, indicating a synergistic or antagonistic effect. On the other way, it can be also defined as the variation of the marginal effect of one factor across strata of another factor [41]. In this study, we focused on investigating these interactions through a rigorous two-phase study and identified one four-way  $G \times E$  interaction. It is a common phenomenon that interaction effect among several factors is usually weaker than their own main effects [22,23,33]. Although two-, three-, and four-way interaction effects are small, but they remain statistically significant, underscoring their robustness and relevance [22,23]. Moreover, the 3D figure we generated visualized the differential effects of cg16658473<sub>SHISA9</sub> at different levels of these variables. These findings underscore the presence of a complex four-way interaction, which may offer valuable insights into the epigenetic regulation of lung cancer prognosis. Building on our previous work with two- and three-way interactions, we found that including the four-way interaction moderately enhanced the prediction accuracy for both 3- and

5-year survival. However, the underlying biological mechanisms of four-way interactions warrant further investigation.

In current study, we identified one novel gene, *SHISA9* (Shisa Family Member 9), contributing to the four-way interaction. It was also called *CKAMP44*, acting as a protein coding gene whose coding proteins belongs to the Shisa protein family [47] and constitute the Cystine-knot AMPA receptor-modulating proteins (*CKAMPs*) family [48]. Emerging evidence indicates that *SHISA9* is intricately involved in immune modulation, particularly through its interactions with autophagy processes and the inhibitor of nuclear factor  $\kappa$ B kinase subunit epsilon (IKK $\epsilon$ ) signaling pathway [49]. NF- $\kappa$ B signaling pathway, including IKK $\epsilon$  pathway, supports tumor survival by regulating apoptosis, promoting the formation of an inflammatory microenvironment, and affect survival outcomes in NSCLC [50,51]. In 2021, Ke et al. [52] discovered that *SHISA9* was an independent risk factor for poor prognosis in LUSC patients. Therefore, *SHISA9* expression may modulate the IKKi signaling pathway, ultimately impacting NSCLC patients.

*TRIM27* played a crucial role in enhancing autophagy flux by activating transcription factor e-box binding (TFEB), thereby contributing to the host's immune defense against intracellular pathogens [53]. Our previous study found that *KIAA0226* may interact with cigarette smoke, regulating the transcription of *Beclin-1* and affecting the cross-regulation between apoptosis and autophagy [23]. Thus, it is plausible that *SHISA9* interacts with *TRIM27*, *KIAA0226* and smoking, influencing the autophagy process and ultimately affecting lung cancer prognosis [54,55]. Despite these findings, biological mechanisms underlying the four-way interaction need to be more systematically studied.

Since DNA methylation may directly affect gene expression and impact cancer prognosis [56,57], we further test the four-way interaction at the gene expression level using genes *cis*-regulated by CpG probes. Anyway, we merely observed interaction among *TRIM27*, *FYTTD1*, *SHISA9*, and pack-year of smoking. *FYTTD1* (Forty-two-three Domain Containing 1), also called *UIF*, was found to play a crucial role on mRNA export [58,59]. Given the interplay between mRNA export and autophagy [60], the interaction of *TRIM27*, *FYTTD1*, *SHISA9* and pack-years of smoking may influence autophagy and, in turn, affect lung cancer prognosis [61]. Interestingly, while cg00060500 was associated with both *KIAA0226* and *FYTTD1*, only *FYTTD1* showed a significant interaction. Despite the significant correlation between *FYTTD1* and

*KIAA0226*, their independent contributions to the interaction model likely differ due to distinct biological functions and the complexity of four-way interactions. *FYTTD1* likely influenced mRNA transcription through its role in mRNA export, directly affecting gene expression and interacting more significantly with the other three factors to impact NSCLC survival. In contrast, while *KIAA0226* may interact with smoking to influence autophagy, the complexity of autophagy likely modulates or dilutes its effect, potentially explaining the lack of significance in the four-way interaction. This difference in biological functions and interaction patterns may explain why *FYTTD1* showed a significant four-way interaction, while *KIAA0226* did not. These findings underscore the variability in gene-specific regulatory effects, highlighting the need for further investigation in NSCLC.

Notably, alongside the heterogeneous survival among the four subgroups, genes significantly associated with CpG probes were enriched in immune-related pathways, including NF- $\kappa$ B, T cell, and B cell signaling pathways, which may affect overall survival in NSCLC [62–64]. Moreover, enriched terms such as cytokine receptor activity and MHC protein binding, were reported to play essential roles in modulating immune responses, which could impact lung cancer prognosis [65–68]. Our subgroup analysis further indicates that immune-related pathways and biological functions may play pivotal roles in the progression of lung cancer.

There are some strengths in our study. First, we were among the first to conduct an epigenomic study on four-way gene-smoking interactions related to overall survival in early-stage NSCLC patients. Next, the false positives were undergone critically appraised in our two-phase study. During the discovery phase, the significant threshold was defined as  $FDR-q \leq 0.05$  and meanwhile as  $P \leq 0.05$  in the validation phase. Therefore, the final false-positive rate is already around 0.0025 ( $0.05 \times 0.05$ ). Then, we generated the 3-D plot and forest plot to facilitate the understanding and interpretation of four-way interaction. Moreover, incorporating four-way interaction, we construct a prognostic model, which may aid in clinical decision-making. Finally, in addition to identifying interactions between CpG probes, we also investigated the interactions between the genes *cis*-regulated by these CpG probes.

We also have some limitations. First, we have only made a stratified analysis of TRIUNE and have not elucidated the biological mechanisms. Next, the TCGA cohort has a high censoring rate (76.07%) of survival time and thus the validation phase based on samples from TCGA population had low statistical power.

Anyway, these signals confirmed significant in TCGA indicates what we observed is conservative and robust. Ultimately, as the majority of population included in this study was Caucasian (89%), the extrapolation of conclusions requires caution.

## 5. Conclusion

We identified a significant CpG probe (cg16658473<sub>SHISA9</sub>), together with cg05293407<sub>TRIM27</sub>, cg00060500<sub>KIAA0226</sub> and pack-year of smoking, to constitute a four-way gene-smoking interaction associated with NSCLC survival.

## Acknowledgements

This study was supported by the National Natural Science Foundation of China (82220108002 to FC and 82273737 to RZ), the US National Institutes of Health (CA209414, CA249096, CA092824, and ES000002 to DCC), Zhejiang Provincial Natural Science Foundation of China (ZCLY24H2601 to LL), One Health Interdisciplinary Research Project of Ningbo University, Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). RZ was partially supported by the Qing Lan Project of the Higher Education Institutions of Jiangsu Province and the Outstanding Young Teachers Training Program of Nanjing Medical University.

## Conflict of interest

The authors declare no conflict of interest.

## Author contributions

LC, DCC, RZ, and FC contributed to conception and design. LL, DCC, RZ, and FC contributed to financial support. LS, MMB, AK, MP, JS, ÅH, ME, DCC, and RZ contributed to data collection and quality control. LC, XW, NX, ZZ, XX, MX, YY, LL, DCC, and RZ contributed to data analysis and interpretation. LC and RZ contributed to manuscript writing. LC, XW, NX, ZZ, XX, MX, YY, LL, DCC, and RZ contributed to critical revision. All authors contributed to final approval of manuscript and accountable for all aspects of the work.

## Peer review

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/1878-0261.13766>.

## Data accessibility

The DNA methylation image data from the USA-Harvard cohort can be requested from DCC. Data from the Spain, Norway, and Sweden cohorts can be requested from ME, ÅH, and JS, respectively. Alternatively, it can be retrieved from gene expression omnibus database (GSE39279, GSE66836 and GSE56044). TCGA: <https://tcga-data.nci.nih.gov>; now hosted at GDC: <https://portal.gdc.cancer.gov>. GEO: <https://www.ncbi.nlm.nih.gov/gds/>.

## References

- Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024;**74**(3):229–63.
- Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin.* 2024;**74**(1):12–49.
- Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc.* 2008;**83**(5):584–94.
- Davidson MR, Gazdar AF, Clarke BE. The pivotal role of pathology in the management of lung cancer. *J Thorac Dis.* 2013;**5**(Suppl 5):S463–78.
- Langer CJ, Besse B, Gualberto A, Brambilla E, Soria JC. The evolving role of histology in the management of advanced non-small-cell lung cancer. *J Clin Oncol.* 2010;**28**(36):5311–20.
- Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer.* 2014;**14**(8):535–46.
- Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 2012;**13**(7):484–92.
- Esteller M. Epigenetics in cancer. *N Engl J Med.* 2008;**358**(11):1148–59.
- Anglim PP, Alonzo TA, Laird-Offringa IA. DNA methylation-based biomarkers for early detection of non-small cell lung cancer: an update. *Mol Cancer.* 2008;**7**:81.
- Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology.* 2013;**38**(1):23–38.
- Yu J, Bulk E, Ji P, Hascher A, Tang M, Metzger R, et al. The EPHB6 receptor tyrosine kinase is a metastasis suppressor that is frequently silenced by promoter DNA hypermethylation in non-small cell lung cancer. *Clin Cancer Res.* 2010;**16**(8):2275–83.
- Ma R, Feng N, Yu X, Lin H, Zhang X, Shi O, et al. Promoter methylation of Wnt/beta-catenin signal

- inhibitor TMEM88 is associated with unfavorable prognosis of non-small cell lung cancer. *Cancer Biol Med*. 2017;**14**(4):377–86.
- 13 Mok T, Camidge DR, Gadgeel SM, Rosell R, Dziadziuszko R, Kim DW, et al. Updated overall survival and final progression-free survival data for patients with treatment-naïve advanced ALK-positive non-small-cell lung cancer in the ALEX study. *Ann Oncol*. 2020;**31**(8):1056–64.
  - 14 Pan Y, Liu G, Zhou F, Su B, Li Y. DNA methylation profiles in cancer diagnosis and therapeutics. *Clin Exp Med*. 2018;**18**(1):1–14.
  - 15 Zhang R, Chen C, Dong X, Shen S, Lai L, He J, et al. Independent validation of early-stage non-small cell lung cancer prognostic scores incorporating epigenetic and transcriptional biomarkers with gene-gene interactions and Main effects. *Chest*. 2020;**158**(2):808–19.
  - 16 Chen J, Song Y, Li Y, Wei Y, Shen S, Zhao Y, et al. A trans-omics assessment of gene-gene interaction in early-stage NSCLC. *Mol Oncol*. 2023;**17**(1):173–87.
  - 17 Xu Z, Chen X, Song X, Kong X, Chen J, Song Y, et al. ATHENA: an independently validated autophagy-related epigenetic prognostic prediction model of head and neck squamous cell carcinoma. *Clin Epigenetics*. 2023;**15**(1):97.
  - 18 Kim DH, Kim JS, Ji YI, Shim YM, Kim H, Han J, et al. Hypermethylation of RASSF1A promoter is associated with the age at starting smoking and a poor prognosis in primary non-small cell lung cancer. *Cancer Res*. 2003;**63**(13):3743–6.
  - 19 Rau CD, Lusis AJ, Wang Y. Systems genetics for mechanistic discovery in heart diseases. *Circ Res*. 2020;**126**(12):1795–815.
  - 20 Noble AJ, Nowak JK, Adams AT, Uhlig HH, Satsangi J. Defining interactions between the genome, epigenome, and the environment in inflammatory bowel disease: progress and prospects. *Gastroenterology*. 2023;**165**(1):44–60.e2.
  - 21 Chen C, Wei Y, Wei L, Chen J, Chen X, Dong X, et al. Epigenome-wide gene-age interaction analysis reveals reversed effects of PRODH DNA methylation on survival between young and elderly early-stage NSCLC patients. *Aging (Albany NY)*. 2020;**12**(11):10642–62.
  - 22 Ji X, Lin L, Shen S, Dong X, Chen C, Li Y, et al. Epigenetic-smoking interaction reveals histologically heterogeneous effects of TRIM27 DNA methylation on overall survival among early-stage NSCLC patients. *Mol Oncol*. 2020;**14**(11):2759–74.
  - 23 Ji X, Lin L, Fan J, Li Y, Wei Y, Shen S, et al. Epigenome-wide three-way interaction study identifies a complex pattern between TRIM27, KIAA0226, and smoking associated with overall survival of early-stage NSCLC. *Mol Oncol*. 2022;**16**(3):717–31.
  - 24 Zhou F, Ren J, Lu X, Ma S, Wu C. Gene-environment interaction: a variable selection perspective. *Methods Mol Biol*. 2021;**2212**:191–223.
  - 25 Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet*. 2001;**69**(1):138–47.
  - 26 Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet*. 2009;**10**(6):392–404.
  - 27 Asomaning K, Miller DP, Liu G, Wain JC, Lynch TJ, Su L, et al. Second hand smoke, age of exposure and lung cancer risk. *Lung Cancer*. 2008;**61**(1):13–20.
  - 28 Sandoval J, Mendez-Gonzalez J, Nadal E, Chen G, Carmona FJ, Sayols S, et al. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. *J Clin Oncol*. 2013;**31**(32):4140–7.
  - 29 Bjaanaes MM, Fleischer T, Halvorsen AR, Daunay A, Busato F, Solberg S, et al. Genome-wide DNA methylation analyses in lung adenocarcinomas: association with EGFR, KRAS and TP53 mutation status, gene expression and prognosis. *Mol Oncol*. 2016;**10**(2):330–43.
  - 30 Karlsson A, Jonsson M, Lauss M, Brunnström H, Jönsson P, Borg Å, et al. Genome-wide DNA methylation analysis of lung carcinoma reveals one neuroendocrine and four adenocarcinoma epitypes associated with patient outcome. *Clin Cancer Res*. 2014;**20**(23):6127–40.
  - 31 Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;**8**(2):203–9.
  - 32 Marabita F, Almgren M, Lindholm ME, Ruhrmann S, Fagerström-Billai F, Jagodic M, et al. An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics*. 2013;**8**(3):333–46.
  - 33 Zhang R, Lai L, Dong X, He J, You D, Chen C, et al. SIPA1L3 methylation modifies the benefit of smoking cessation on lung adenocarcinoma survival: an epigenomic-smoking interaction analysis. *Mol Oncol*. 2019;**13**(5):1235–48.
  - 34 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol*. 1995;**57**(1):289–300.
  - 35 Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res*. 2017;**45**(W1):W130–7.
  - 36 Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res*. 2015;**43**(Database issue):D1049–56.



- 37 Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017;**45**(D1):D353–61.
- 38 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;**25**(1):25–9.
- 39 Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WEE, et al. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol.* 2016;**11**(1):39–51.
- 40 Chen J, Shen S, Li Y, Fan J, Xiong S, Xu J, et al. APOLLO: an accurate and independently validated prediction model of lower-grade gliomas overall survival and a comparative study of model performance. *EBioMedicine.* 2022;**79**:104007.
- 41 Zhang R, Shen S, Wei Y, Zhu Y, Li Y, Chen J, et al. A large-scale genome-wide gene-gene interaction study of lung cancer susceptibility in Europeans with a trans-ethnic validation in Asians. *J Thorac Oncol.* 2022;**17**(8):974–90.
- 42 Wang LY, Cui JJ, Liu JY, Guo AX, Zhao ZY, Liu YZ, et al. Gene-gene and gene-environment interaction data for platinum-based chemotherapy in non-small cell lung cancer. *Sci Data.* 2018;**5**:180284.
- 43 Zhu J, Guan J, Ji X, Song Y, Xu X, Wang Q, et al. A two-phase comprehensive NSCLC prognostic study identifies lncRNAs with significant main effect and interaction. *Mol Genet Genomics.* 2022;**297**(2):591–600.
- 44 Zhang R, Chu M, Zhao Y, Wu C, Guo H, Shi Y, et al. A genome-wide gene-environment interaction analysis for tobacco smoke and lung cancer susceptibility. *Carcinogenesis.* 2014;**35**(7):1528–35.
- 45 Xu Z, Gu Y, Chen J, Chen X, Song Y, Fan J, et al. Epigenome-wide gene-age interaction study reveals reversed effects of MORN1 DNA methylation on survival between young and elderly oral squamous cell carcinoma patients. *Front Oncol.* 2022;**12**:941731.
- 46 Chu M, Zhang R, Zhao Y, Wu C, Guo H, Zhou B, et al. A genome-wide gene-gene interaction analysis identifies an epistatic gene pair for lung cancer susceptibility in Han Chinese. *Carcinogenesis.* 2014;**35**(3):572–7.
- 47 Karataeva AR, Klaassen RV, Stroder J, Ruiperez-Alonso M, Hjorth JJ, van Nierop P, et al. C-terminal interactors of the AMPA receptor auxiliary subunit Shisa9. *PLoS One.* 2014;**9**(2):e87360.
- 48 von Engelhardt J, Mack V, Sprengel R, Kavenstock N, Li KW, Stern-Bach Y, et al. CKAMP44: a brain-specific protein attenuating short-term synaptic plasticity in the dentate gyrus. *Science.* 2010;**327**(5972):1518–22.
- 49 Zheng Y, Wang L, Liu Q, Xian H, Zhang C, Cai S, et al. Modulation of virus-induced neuroinflammation by the autophagy receptor SHISA9 in mice. *Nat Microbiol.* 2023;**8**(5):958–72.
- 50 Dimitrakopoulos FD, Kottorou AE, Kalofonos M, Kalofonos HP. The fire within: NF-kappaB involvement in non-small cell lung cancer. *Cancer Res.* 2020;**80**(19):4025–36.
- 51 Mizushima N, Levine B. Autophagy in human diseases. *N Engl J Med.* 2020;**383**(16):1564–76.
- 52 Ke D, Guo Q, Fan TY, Xiao X. Analysis of the role and regulation mechanism of hsa-miR-147b in lung squamous cell carcinoma based on the cancer genome atlas database. *Cancer Biother Radiopharm.* 2021;**36**(3):280–91.
- 53 Zhao D, Qiang L, Lei Z, Ge P, Lu Z, Wang Y, et al. TRIM27 elicits protective immunity against tuberculosis by activating TFEB-mediated autophagy flux. *Autophagy.* 2024;**20**(7):1483–504.
- 54 Khayati K, Bhatt V, Lan T, Alogaili F, Wang W, Lopez E, et al. Transient systemic autophagy inhibition is selectively and irreversibly deleterious to lung cancer. *Cancer Res.* 2022;**82**(23):4429–43.
- 55 Han T, Guo M, Gan M, Yu B, Tian X, Wang JB. TRIM59 regulates autophagy through modulating both the transcription and the ubiquitination of BECN1. *Autophagy.* 2018;**14**(12):2035–48.
- 56 Hong J, Rhee JK. Genomic effect of DNA methylation on gene expression in colorectal cancer. *Biology (Basel).* 2022;**11**(10):1388.
- 57 Carrot-Zhang J, Chambwe N, Damrauer JS, Knijnenburg TA, Robertson AG, Yau C, et al. Comprehensive analysis of genetic ancestry and its molecular correlates in cancer. *Cancer Cell.* 2020;**37**(5):639–54.e6.
- 58 Hautbergue GM, Hung ML, Walsh MJ, Snijders APL, Chang CT, Jones R, et al. UIF, a new mRNA export adaptor that works together with REF/ALY, requires FACT for recruitment to mRNA. *Curr Biol.* 2009;**19**(22):1918–24.
- 59 Ehrnsberger HF, Pfaff C, Hachani I, Flores-Tornero M, Sørensen BB, Längst G, et al. The UAP56-interacting export factors UIF1 and UIF2 function in mRNA export. *Plant Physiol.* 2019;**179**(4):1525–36.
- 60 Shi SL, Fukuda H, Chujo T, Kouwaki T, Oshiumi H, Tomizawa K, et al. Export of RNA-derived modified nucleosides by equilibrative nucleoside transporters defines the magnitude of autophagy response and Zika virus replication. *RNA Biol.* 2021;**18**(Suppl 1):478–95.
- 61 Hama Y, Ogasawara Y, Noda NN. Autophagy and cancer: basic mechanisms and inhibitor development. *Cancer Sci.* 2023;**114**(7):2699–708.
- 62 Wang SS, Liu W, Ly D, Xu H, Qu L, Zhang L. Tumor-infiltrating B cells: their role and application in

- anti-tumor immunity in lung cancer. *Cell Mol Immunol*. 2019;**16**(1):6–18.
- 63 Hao D, Han G, Sinjab A, Gomez-Bolanos LI, Lazcano R, Serrano A, et al. The single-cell immunogenomic landscape of B and plasma cells in early-stage lung adenocarcinoma. *Cancer Discov*. 2022;**12**(11):2626–45.
  - 64 Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med*. 2018;**24**(7):978–85.
  - 65 Yie SM, Yang H, Ye SR, Li K, Dong DD, Lin XM. Expression of human leucocyte antigen G (HLA-G) is associated with prognosis in non-small cell lung cancer. *Lung Cancer*. 2007;**58**(2):267–74.
  - 66 Yang S, Tang D, Zhao YC, Liu H, Luo S, Stinchcombe TE, et al. Potentially functional variants of ERAP1, PSMF1 and NCF2 in the MHC-I-related pathway predict non-small cell lung cancer survival. *Cancer Immunol Immunother*. 2021;**70**(10):2819–33.
  - 67 Van Dyke AL, Cote ML, Wenzlaff AS, Chen W, Abrams J, Land S, et al. Cytokine and cytokine receptor single-nucleotide polymorphisms predict risk for non-small cell lung cancer among women. *Cancer Epidemiol Biomarkers Prev*. 2009;**18**(6):1829–40.
  - 68 Guillon A, Gueugnon F, Mavridis K, Dalloneau E, Jouan Y, Diot P, et al. Interleukin-22 receptor is overexpressed in nonsmall cell lung cancer and portends a poor prognosis. *Eur Respir J*. 2016;**47**(4):1277–80.

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Fig. S1.** Quality control processes for DNA methylation data.

**Fig. S2.** The functional enrichment analyses of genes significantly associated with CpG probes involving in TRIUNE.

**Fig. S3.** Correlation between *KIAA0226* and *FYTTD1*.

**Table S1.** Demographic and clinical descriptions of early-stage NSCLC patients with gene expression data in four international study centers.

**Table S2.** Association results of four-way interactions in the discovery phase, the validation phase, and the combined data.

**Table S3.** The association results of cg16658473<sub>SHISA9</sub> derived from Cox proportional hazards model adjusted for covariates in NSCLC samples.

**Table S4.** Annotation information for three CpG probes located in the genes *TRIM27*, *KIAA0226*, and *SHISA9*.

**Table S5.** Correlation results between DNA methylation probes and corresponding gene expressions.

**Table S6.** Association results of the four-way interaction among pack-year of smoking, *TRIM27*, *KIAA0226*, and *SHISA9*.

**Table S7.** Association results of the four-way interaction among pack-year of smoking, *TRIM27*, *FYTTD1*, and *SHISA9*.