SIPA1L3 methylation modifies the benefit of smoking cessation on lung adenocarcinoma survival: An epigenomic–smoking interaction analysis

Ruyang Zhang¹²³, Linjing Lai¹*, Xuesi Dong¹⁴, Jieyu He¹, Dongfang You¹, Chao Chen¹, Lijuan Lin¹, Ying Zhu¹, Hui Huang¹, Sipeng Shen¹²³, Liangmin Wei¹, Xin Chen¹, Yichen Guo²⁵, Liya Liu⁶, Li Su²³, Andrea Shafer⁷, Sebastian Moran⁸, Thomas Fleischer⁹, Maria Moksnes Bjaanaes⁹, Anna Karlsson¹⁰, Maria Planck¹⁰, Johan Staaf¹⁰, Åslaug Helland⁸¹¹, Manel Esteller⁸, Yongyue Wei¹²³**, Feng Chen¹³¹²**, David C. Christiani²³⁶†

¹Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu, China, 211166
2Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA, 02115

3China International Cooperation Center for Environment and Human Health, Nanjing Medical University, Nanjing, Jiangsu, China, 211166

4Department of Epidemiology and Biostatistics, School of Public Health, Southeast University, Nanjing, Jiangsu, China, 210009

5Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA, 02115

6Department of Preventive Medicine, Medical School of Ningbo University, Ningbo, Zhejiang, China, 315211

7Pulmonary and Critical Care Division, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA, 02114

8Bellvitge Biomedical Research Institute and University of Barcelona and Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain, 08908

9Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway, 0424

10Division of Oncology and Pathology, Department of Clinical Sciences Lund and CREATE Health Strategic Center for Translational Cancer Research, Lund University, Lund, Skåne, Sweden, 22381
11 Institute of Clinical Medicine, University of Oslo, Oslo, Norway, 0424

12 Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Cancer Center, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, Nanjing, Jiangsu, China 211166

*Authors contributed equally to the work.

†Senior author who supervised the work.

**Send correspondence to:
Feng Chen, PhD, SPH Building Room 412, 101 Longmian Avenue, Nanjing, Jiangsu 211166, China; E-mail: fengchen@njmu.edu.cn; Tel: +86 25 86868435

Yongyue Wei, PhD, SPH Building Room 418, 101 Longmian Avenue, Nanjing, Jiangsu 211166, China; E-mail: ywei@njmu.edu.cn; Tel: +86 25 86868436

Running title: Epigenomic–smoking interaction analysis of SIPA1L3
List of abbreviations

NSCLC: non-small cell lung cancer
TCGA: The Cancer Genome Atlas
LUAD: lung adenocarcinomas
LUSC: lung squamous cell carcinomas
QC: quality control
SNP: single nucleotide polymorphisms
SD: standard deviation
FDR: false discovery rate
HR: hazard ratio
CI: confidence interval

Abstract

Smoking cessation prolongs survival and decreases mortality of patients with non-small cell lung cancer (NSCLC). In addition, epigenetic alterations of some genes are associated with survival. However, potential interactions between smoking cessation and epigenetics have not been assessed. Here, we conducted an epigenome-wide interaction analysis between DNA methylation and smoking cessation on NSCLC survival. We used a two-stage study design to identify DNA methylation–smoking cessation interactions that affect overall survival for early-stage NSCLC. The discovery phase contained NSCLC patients from USA, Spain, Norway,
and Sweden. A histology-stratified Cox proportional hazards model adjusted for age, sex, clinical stage, and study center was used to test DNA methylation–smoking cessation interaction terms. Interactions with FDR-\(q \leq 0.05\) were further confirmed in a validation phase using The Cancer Genome Atlas (TCGA) database. Histology-specific interactions were identified by stratification analysis in lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) patients. We identified one CpG probe (cg02268510\textsubscript{SIPA1L3}) that significantly and exclusively modified the effect of smoking cessation on survival in LUAD patients (\(HR\text{interaction} = 1.12; 95\% \text{ CI: } 1.07–1.16; \ P = 4.30 \times 10^{-7}\)). Further, the effect of smoking cessation on early-stage LUAD survival varied across patients with different methylation levels of cg02268510\textsubscript{SIPA1L3}. Smoking cessation only benefited LUAD patients with low methylation (\(HR = 0.53; 95\% \text{ CI: } 0.34–0.82; \ P = 4.61 \times 10^{-3}\)) rather than medium or high methylation (\(HR = 1.21; 95\% \text{ CI: } 0.86–1.70; \ P = 0.266\)) of cg02268510\textsubscript{SIPA1L3}. Moreover, there was an antagonistic interaction between elevated methylation of cg02268510\textsubscript{SIPA1L3} and smoking cessation (\(HR\text{interaction} = 2.1835; 95\% \text{ CI: } 1.27–3.74; \ P = 4.46 \times 10^{-3}\)). In summary, smoking cessation benefited survival of LUAD patients with low methylation at cg02268510\textsubscript{SIPA1L3}. The results have implications for not only smoking cessation after diagnosis, but also possible methylation-specific drug targeting.
Keywords: overall survival, smoking cessation, DNA methylation, non-small cell lung cancer, interaction analysis, molecular cancer epidemiology

1. Introduction

Lung cancer is a leading cause of cancer mortality worldwide. In the United States, lung cancer was estimated as likely to account for 154,050 deaths in 2018, or one-fourth of all cancer deaths (Siegel et al., 2017). A large proportion of lung cancer cases are attributed to smoking, a well-known risk factor (Flanders et al., 2003), and smoking cessation prolongs survival and decreases mortality of lung cancer patients (Balduyck et al., 2011; Parsons et al., 2010). However, the underlying mechanisms of these benefits remain largely unclear (Bhatt et al., 2015; Parsons et al., 2010).

DNA methylation, a reversible epigenetic modification, regulates gene expression and provides potential cancer biomarkers and therapeutic targets (Egger et al., 2004; Feinberg and Tycko, 2004), including for non-small cell lung cancer (NSCLC) (Guo et al., 2018; Shen et al., 2018; Wei et al., 2018). Furthermore, as a potential mechanistic link between cigarette smoking and disease, DNA methylation changes can result from various environmental exposures and may explain part of the association between smoking and cancer recurrence or mortality (Lee and Pausova, 2013; Shui et al., 2016).
Progression of complex diseases, such as cancer, results from interactions between clinical, environmental, genetic, and epigenetic factors (Lacombe et al., 2016; Mcnerney et al., 2017). However, most epigenome-wide association studies are designed to identify main effects using a standard marginal test (Karlsson et al., 2014) while ignoring epigenetic–environment interactions. These traditional mining procedures may reduce the power to identify new epigenomic biomarkers (Slade and Kraft, 2016).

In this study, we hypothesized that epigenetic and smoking cessation interactions may affect NSCLC survival. Epigenome-wide DNA methylation data composed of 4 study cohorts containing lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) cases were used for discovery, and the findings were independently validated in The Cancer Genome Atlas (TCGA) data.

2. Materials and methods

2.1. Study population

Early-stage (stage I–II) LUAD and LUSC patients who were former or current smokers were included in the study. Never smokers were defined as those who smoked ≤100 cigarettes over a lifetime. Current smokers were defined as those who were smoking within 1 year of diagnosis. Former smokers were defined as smokers who quit >1 year before diagnosis or interview (Suk et al., 2006). We encoded the
variable smoking cessation as “yes” for former smokers and “no” for current smokers. Data were harmonized from 5 international study centers, which have been previously described (Guo et al., 2018; Shen et al., 2018; Wei et al., 2018; Zhang et al., 2019). All patients provided written informed consent, and the study methodologies conformed to the standards set by the Declaration of Helsinki and received approval by its respective institutional review board.

**Harvard:** The Harvard Lung Cancer Study cohort was described previously (Suk et al., 2006). Briefly, all patients were recruited at Massachusetts General Hospital (MGH) from 1992–present and had newly diagnosed, histologically confirmed primary NSCLC. We included 133 early-stage LUAD and LUSC patients who were former or current smokers for the current study. DNA was extracted from tumor specimens that were evaluated by an MGH pathologist for amount (tumor cellularity >70%) and quality of tumor cells and histologically classified using World Health Organization criteria. The study protocol was approved by the Institutional Review Boards at Harvard School of Public Health and MGH.

**Spain:** The Spain study population was reported previously (Sandoval et al., 2013) and included 196 LUAD and LUSC patients recruited at 8 sub-centers from 1991–2009. In brief, tumor DNA was extracted from fresh-frozen tumor specimens that were collected by surgical resection, and the median clinical follow-up was 7.2
years. The study was approved by the Bellvitge Biomedical Research Institute institutional review board.

**Norway:** The Norway cohort consisted of 116 LUAD patients with operable lung cancer tumors who were seen at Oslo University Hospital, Rikshospitalet, Norway, in 2006–2011 (Bjaanæs et al., 2016). Tumor tissues obtained during surgery were snap frozen in liquid nitrogen and stored at -80 °C until DNA isolation. The project was approved by Oslo University institutional review board and regional ethics committee (S-05307).

**Sweden:** The Sweden cohort included 85 LUAD and LUSC patients. Tumor DNA was collected from early-stage lung cancer patients who underwent an operation at the Skane University Hospital, Lund, Sweden (Karlsson et al., 2014). The study was approved by the Regional Ethical Review Board in Lund, Sweden (Registration no. 2004/762 and 2008/702).

**TCGA:** The TCGA database contains 562 early-stage LUAD and LUSC patients who have full information of survival time and covariates. Level-1 HumanMethylation450 DNA methylation data (image data) for each patient were downloaded on October 1, 2015.
### 2.2. Quality control procedures

DNA methylation was profiled using Infinium HumanMethylation450 BeadChips (Illumina Inc.) for all patients. Raw image data were imported into Genome Studio Methylation ModuleV1.8 (Illumina Inc.) to calculate methylation signals and to perform normalization, background subtraction, and quality control (QC). Beta-values, which range from 0% (unmethylated) to 100% (methylated), were used to measure methylation level of each probe. Unqualified probes were excluded if they met any one of the following QC criteria: (i) failed detection ($P > 0.05$) in $>5\%$ samples; (ii) coefficient of variance of $<5\%$; (iii) methylated or unmethylated in all samples; (iv) common single nucleotide polymorphisms (SNP) located in probe sequence or 10-bp flanking regions; (v) cross-reactive or cross-hybridizing probes (Chen et al., 2013); (vi) or did not pass QC in all centers. Samples with $>5\%$ undetectable probes were excluded. Methylation signals were further processed for quantile normalization, design bias correction for type I and II probes, and batch effects adjustment using *ComBat* correction (Marabita et al., 2013). We performed QC procedures above in each center separately, and then merged all data together before association analysis. Details of QC processes are described in Supplementary Figure S1.
2.3. Gene expression data

Expression and mRNA sequencing data were available for 281 LUAD and 277 LUSC patients of the TCGA dataset (Supplementary Table S1). TCGA mRNA sequencing data processing and QC were done by the TCGA workgroup. Raw counts were normalized using RNA-sequencing by expectation maximization. Level-3 gene quantification data were downloaded from the TCGA data portal (https://tcga-data.nci.nih.gov; now hosted at https://portal.gdc.cancer.gov) and were further checked for quality. Gene expression data were extracted and log2-transformed before analysis.

2.4. Epigenome-wide DNA methylation–smoking cessation interaction analysis

Analysis flow is described in Figure 1. Patients from the first four study centers (USA, Spain, Norway, and Sweden) were assigned into the discovery phase. A histology-stratified Cox proportional hazards model was used to test the interaction item, which was the interaction effect between DNA methylation of each CpG probe and smoking cessation (CpG probe × smoking cessation) on overall survival. The model was adjusted for age, sex, smoking cessation, clinical stage, and study center. Hazard ratio (HR) and 95% confidence interval (CI) were described per 1% methylation increment. Multiple testing corrections were performed using the false discovery rate method (FDR, measured by FDR-q value) by the Benjamini-Hochberg procedure. CpG probes with interaction FDR-q ≤ 0.05 were replicated in the
validation phase using the TCGA dataset. Robustly significant probes were retained if they met all criteria: (i) interaction $P \leq 0.05$ in the validation phase, and (ii) consistent effect direction in both discovery and validation phases. We performed stratified analysis for robustly significant CpG probes in LUAD and LUSC patients. Finally, CpG probes with a significant interaction with smoking cessation in both phases were identified as histology-specific probes.

2.5. Sensitivity analysis for significant CpG probes

Due to the complex tumor microenvironment—including non-cancerous components, which might alter analysis of tumor samples (Aran et al., 2015)—we assessed tumor purity with InfiniumPurify (Zhang et al., 2015) using methylation array data from TCGA samples. Tumor purity was included as an additional covariate in the Cox regression model for sensitivity analysis.

2.6. Genome-wide methylation transcription analysis

For robustly significant histology-specific prognostic CpG probes, we also performed genome-wide methylation transcription analysis using mRNA sequencing data from TCGA. The correlation between DNA methylation and gene expression was tested using a linear regression model adjusted for the same covariates mentioned above. Association with FDR-$q \leq 0.05$ was considered significant. Additionally, we tested the association between gene expression and overall survival using a Cox proportional
hazards model adjusted for the same covariates. Genes involved in significant associations with both methylation and NSCLC survival were filtered.

2.7. Statistical analysis

Continuous variables were summarized as mean ± standard deviation (SD), and categorized variables were described by frequency (n) and proportion (%). Kaplan-Meier survival curves were used to compare survival difference among subgroups. Statistical analyses were performed using R version 3.4.4 (The R Foundation for Statistical Computing).

3. Results

After QC, epigenome-wide DNA methylation data including 311,891 CpG sites from 1,092 tumor samples of early-stage (stage I–II) NSCLC patients were retained. There were 530 patients (\(N_{LUAD} = 413\) and \(N_{LUSC} = 117\)) in the discovery phase and 562 patients (\(N_{LUAD} = 285\) and \(N_{LUSC} = 277\)) in the validation phase. Table 1 details demographic and clinical information for the study population. There were 37% and 27% current smokers in the discovery and validation phases, respectively.

In the discovery phase, 15 methylation–smoking cessation interactions were identified with FDR-\(q \leq 0.05\) (Supplementary Figure S2A) and the Manhattan plot also showed the results for main effect additionally (Supplementary Figure S2B). Only 1
interaction remained statistically significant in the validation phase under the most stringent criteria (Supplementary Table S2). This site, cg02268510, is located in signal induced proliferation associated 1 like 3 (SIPA1L3). Further histology-stratified analysis showed that cg02268510\textsubscript{SIPA1L3} is a LUAD-specific CpG probe that interacts with smoking cessation to affect patient survival in the discovery phase \((HR_{\text{interaction}} = 1.10; \text{95\% CI: 1.05–1.16}; P = 2.95\times10^{-5})\), the validation phase \((HR_{\text{interaction}} = 1.17; \text{95\% CI: 1.02–1.35}; P = 0.0255)\) and the combined data \((HR_{\text{interaction}} = 1.12; \text{95\% CI: 1.07–1.16}; P = 4.30\times10^{-7})\). Moreover, fixed-effect meta-analysis of five centers also remained significant \((HR_{\text{interaction}} = 1.09; \text{95\% CI: 1.05–1.13}; P = 6.66\times10^{-6})\) (Supplementary Figure S3). As presented in Figure 2A, with decreased methylation level of cg02268510\textsubscript{SIPA1L3}, there was an elevated benefit effect size of smoking cessation on LUAD survival. Thus, there was a modification effect of cg02268510\textsubscript{SIPA1L3} on the association between smoking cessation and survival.

After including tumor purity as an additional covariate in sensitivity analysis, DNA methylation at cg02268510\textsubscript{SIPA1L3} retained a significant interaction with smoking cessation on LUAD survival \((HR_{\text{interaction}} = 1.18; \text{95\% CI: 1.02–1.36}; P = 0.024)\). The interaction \(P\)-value was still significant but slightly inflated due to (i) the smaller sample size (51\% of original) of the sensitivity analysis, which was only performed in TCGA, and (ii) low tumor purity (~60\%) for NSCLC samples in TCGA due to mixed cell types (Zheng et al., 2017).
To better illustrate the interaction pattern between DNA methylation and smoking cessation, patients were categorized into low, medium, and high groups based on tertiles of cg02268510\textsubscript{SIPA1L3} methylation. The effect of smoking cessation varied across LUAD patients with different DNA methylation levels. Smoking cessation only benefited LUAD patients with low methylation of cg02268510\textsubscript{SIPA1L3} ($HR_{\text{low}} = 0.53$; 95% CI: 0.34–0.82; $P = 4.61 \times 10^{-3}$). However, there was no significant association between smoking cessation and survival in LUAD patients with medium–high methylation of cg02268510\textsubscript{SIPA1L3} ($HR_{\text{medium}} = 1.12$; 95% CI: 0.67–1.87; $P = 0.665$; $HR_{\text{high}} = 1.29$; 95% CI: 0.80–2.07; $P = 0.293$; $HR_{\text{medium-high}} = 1.21$; 95% CI: 0.86–1.70; $P = 0.266$). We observed significant heterogeneity of smoking cessation effect across the three groups ($P = 0.014$) (Fig. 2B), and Kaplan-Meier curves confirmed these results (Fig. 2C).

These results also indicated that LUAD patients who did not quit smoking (current smokers) had the poorest prognosis if their methylation of cg02268510\textsubscript{SIPA1L3} were in a low level. So we combined the medium and high methylation groups and performed further analysis. Current smokers in the low methylation group had 1.94-times the mortality risk compared with medium or high methylation groups (Fig. 3A), but there was no statistically significant difference between groups for former smokers (Fig. 3B). The results also indicated that smoking cessation was quite urgent for LUAD patients with low methylation of cg02268510\textsubscript{SIPA1L3}. 

Molecular Oncology (2019) © 2019 The Authors. Published by FEBS Press and John Wiley & Sons Ltd.
In addition, we evaluated the joint effect of CpG methylation level (medium–high versus low) and smoking cessation (Yes versus No) on LUAD survival (Table 2). We used the poorest-prognosis group (current smokers with low methylation) as the reference to evaluate effect of elevated methylation level, smoking cessation and their interaction. In the combined dataset, the effect of smoking cessation was $HR = 0.5506$ (95% CI: 0.36–0.84; $P = 5.62 \times 10^{-3}$) and the effect of medium–high methylation of cg02268510$_{SIPA1L3}$ was $HR = 0.5214$ (95% CI: 0.34–0.81; $P = 3.48 \times 10^{-3}$). However, the joint effect was $HR = 0.6268$ (95% CI: 0.43–0.92; $P = 1.84 \times 10^{-2}$), which was greater than the product of the two individual protective effects ($0.5506 \times 0.5214 = 0.2871$). The joint effect of two protective factors was less protective than expected, indicating an antagonistic interaction between elevated methylation of cg02268510$_{SIPA1L3}$ and smoking cessation ($HR_{interaction} = 2.1835$; 95% CI: 1.27–3.74; $P = 4.46 \times 10^{-3}$).

A growing body of research has reported potential associations of DNA methylation with age and smoking (Fraga and Esteller, 2007; Wan et al., 2012; Zaghlool et al., 2015). Therefore, we also tested the association between methylation of cg02268510$_{SIPA1L3}$ and age, as well as smoking related variables: pack-year of smoking, years of smoking and years of smoking cessation using a linear regression model adjusted for age, sex, clinical stage and study centers. Smoking related characteristics of former and current smokers in early-stage LUAD were described in...
Supplementary Table S3. There was no significant association between methylation of cg02268510\textsubscript{SIPA1L3} and age ($\beta$ = –0.01; $P$ = 0.521) or years of smoking ($\beta$ = 0.03; $P$ = 0.210), but pack-year of smoking ($\beta$ = 0.02; $P$ = 3.42×10\textsuperscript{-3}) as well as years of smoking cessation ($\beta$ = –0.06; $P$ = 5.08×10\textsuperscript{-3}) in former smokers LUAD patients (Supplementary Figure S4).

Further, because cg02268510\textsubscript{SIPA1L3} maps to \textit{SIPA1L3}, the association between cg02268510\textsubscript{SIPA1L3} and \textit{SIPA1L3} expression was evaluated using the TCGA dataset. We observed a significant, while moderate, association between cg02268510\textsubscript{SIPA1L3} and \textit{SIPA1L3} expression ($\beta$ = –0.02; $P$ = 0.015) in LUAD patients (Figure 4), indicating that cg02268510\textsubscript{SIPA1L3} cis-regulates gene expression. Moreover, genome-wide methylation transcription analysis revealed that expression of 633 genes was significantly correlated with methylation level of cg02268510\textsubscript{SIPA1L3} (Supplementary Figure S5A). Among them, expression of only 7 genes was significantly associated with overall survival: \textit{GADD45G}, \textit{MTURN}, \textit{TMEM200B}, \textit{RGS20}, \textit{RELL1}, \textit{PGM2}, and \textit{RIPK2} (Supplementary Figure S5B–H).
4. Discussion

In this study, we systematically evaluated all pair-wise DNA methylation–smoking cessation interactions on an epigenome-wide scale and further confirmed these interactions in an independent population. To our knowledge, this is the first study with a large sample size to investigate interactions between DNA methylation and smoking behavior on lung cancer survival, and it provides new evidence to account for the missing heritability of complex diseases (Trerotola et al., 2015). Our results show that the effect of smoking cessation on early-stage LUAD patient survival varies with methylation level of cg02268510_SIPA1L3. Smoking cessation only benefits LUAD patients with low methylation, rather than medium or high methylation, of cg02268510_SIPA1L3. Further, there is an antagonistic interaction between elevated methylation of cg02268510_SIPA1L3 and smoking cessation.

We found that in LUAD patients with low methylation of cg02268510_SIPA1L3, current smokers with more accumulative exposure had worse survival than former smokers. However, for a population with medium–high methylation, the prognosis of current smokers was similar to that of former smokers. The effect of smoking cessation is therefore modified by DNA methylation level, indicating opportunities for epi-drug intervention due to the inherent reversibility of epigenetic events (Wright, 2013).
Up to 50% of lung cancer patients are estimated to keep smoking after diagnosis or to frequently relapse after smoking cessation (Park et al., 2012; Walker et al., 2006a). Our results indicated that smoking cessation was urgent especially for LUAD patients with low methylation of cg02268510\textsubscript{SIPA1L3}. On the other hand, reduced methylation of cg02268510\textsubscript{SIPA1L3} might strengthen the protective effect of smoking cessation on survival.

Many studies have reported significant associations between smoking cessation and overall survival (Koshiaris et al., 2017; Nia et al., 2005), while other studies have reported negative results (Baser et al., 2006; Parsons et al., 2010). Based on our interaction analysis, we suspected that epigenetic modifications might account for this inconsistent phenomenon. Because the effect of smoking cessation varies across populations with different methylation levels of cg02268510\textsubscript{SIPA1L3}, the effect could be neutralized in a population of patients with mixed cg02268510\textsubscript{SIPA1L3} methylation levels. Thus, the traditional marginal test for association between smoking cessation and cancer survival inherently loses statistical power to report significant findings due to complex association patterns.

\textit{SIPA1L3}, the gene in which cg02268510 is located, encodes GTPase-activating proteins (GAPs) specific for the GTP-binding protein \textit{RAP1} (Ras-associated protein-1), which is implicated in regulation of cell adhesion, cell polarity, and cytoskeletal organization (Kooistra et al., 2007). \textit{SIPA1L3} is a member of the SPA1 family of RapGAPs, which play a crucial role in spatiotemporal control of Rap1
activation in cells (Mochizuki et al., 2001). Rap1 plays many roles during cell invasion and metastasis in different cancers (Zhang et al., 2017). Additionally, overexpression of RAP1 may desensitize NSCLC cells to cisplatin, a first-line drug to treat NSCLC (Besse et al., 2014). Our results suggest that low methylation at cg02268510SIPA1L3 might promote SIPA1L3 expression, further leading to Rap1 activation and resulting in poor prognosis (Figure 5).

Many of the deleterious effects of smoking are due to induction of inflammatory responses that contribute to lung cancer progression (Cruz and Balkwill, 2015; Walser et al., 2008). In vitro experiments in human umbilical vein endothelial cells demonstrate that nicotine stimulates cellular inflammatory responses by activating the NF-κB transcription factor axis by a second messenger pathway (Ueno et al., 2006). Activation of NF-κB, one of the most investigated transcription factors, controls multiple cellular processes in cancer, including inflammation, transformation, proliferation, angiogenesis, invasion, metastasis, chemoresistance, and radioresistance (Chaturvedi et al., 2010). Nicotine protects NSCLC cells against chemotherapy-induced apoptosis and serum deprivation-induced apoptosis through NF-κB, and NF-κB activity is also directly stimulated by nicotine (Anto et al., 2002; Tsurutani et al., 2005). Therefore, for current smokers, nicotine in tobacco stimulates activation of NF-κB, induces inflammatory responses, and is relevant to poor patient prognosis (Figure 5).
Moreover, Rap1 is an essential modulator of NF-κB-mediated pathways. NF-κB is induced by ectopic expression of Rap1, whereas its activity is inhibited by Rap1 depletion (Teo et al., 2010). Furthermore, levels of Rap1 are positively regulated by NF-κB, and human breast cancers with NF-κB hyperactivity show elevated levels of cytoplasmic Rap1 (Teo et al., 2010). Thus, positive feedback mechanisms might exist between Rap1 expression and NF-κB activation (Figure 5). In terms of cg02268510\textsubscript{SIPA1L3} and smoking cessation interaction, keeping smoking was associated with poor prognosis only in LUAD patients with low methylation, rather than medium or high methylation, possibly because high activation of both Rap1 and NF-κB may only occur in patients with low methylation.

We also found that methylation level of cg02268510\textsubscript{SIPA1L3} increased along with long pack-year of smoking, but decreased with long years of smoking cessation. As presented at Figure 5, low methylation and keeping smoking resulted in the worst prognosis, which might be due to the positive feedback in Rap1 and NF-κB. However, methylation of cg02268510\textsubscript{SIPA1L3} increased with the cumulative amount of smoking. But, high methylation of cg02268510\textsubscript{SIPA1L3} resulted low SIPA1L3 expression that was hard to active Rap1 and NF-κB, then might weaken the harmful effect of smoking, which also indicated an antagonistic effect. It is implied that there might be a self-protective mechanism in the human body that prevents the body from receiving excessive damage from exposure. As reported, smoking increases reactive oxygen species (ROS) production and is a significant source of oxidative stress (Athanasios et
al., 2013), but in vivo, there is a variety of antioxidant defense mechanisms existed to counteract the detrimental effects of ROS by regulating the production of free radicals and their metabolites (Deponte, 2013; He et al., 2017). It may be an adaptive defense mechanism to counteract the increased ROS production that superoxide dismutase enzyme levels in blood and salivary were increased in smokers (Jenifer et al., 2015). Moreover, previous study has found that activation of Rap1 serves to attenuate ROS production (Remans et al., 2004) and there is a potential inter-relationship between Rap1, ROS and NF-κB activation (Moon et al., 2011). But further functional studies are warranted to elucidate the mechanism of cg02268510 SIPA1L3 and smoking cessation interaction on LUAD survival.

Meanwhile, we observed that some genes trans-regulated by cg02268510 SIPA1L3 are involved in DNA damage response and cell growth [growth arrest and DNA damage inducible gamma (GADD45G)] (Guo et al., 2013), immune cell functions [maturin (MTURN)] (Sun et al., 2014), tumor cell migration [regulator of G protein signaling 20 (RGS20)] (Yang et al., 2016), apoptosis [RELT like 1 (RELL1) and receptor interacting serine/threonine kinase 2 (RIPK2)] (Chin et al., 2002; Cusick et al., 2010), and innate and adaptive immunity (RIPK2) (Jaafar et al., 2018).

GADD45G is a member of the GADD45 family, which plays an essential role in cellular stress response, survival, senescence, and apoptosis regulation (Liebermann et al., 2011). GADD45G has been reported to be a tumor suppressor in multiple cancer types, and can inhibit cell growth and induce apoptosis (Ying et al., 2005).
Patients with high $GADD45G$ expression had a better prognosis in our study. $MTURN$ is a neural progenitor differentiation regulator homolog. 12-O-tetradecanoylphorbol-13-acetate (TPA) is an effective cancer therapeutic reagent for myelocytic leukemia patients (Han et al., 1998), and $MTURN$ is TPA-responsive and may promote both leukemic and normal megakaryocyte differentiation (Sun et al., 2014). Indeed, differentiation therapy by forced differentiation of cancer cells has been successful in curing acute promyeloid leukemia (Chen et al., 2011). Similarly, LUAD patients with high $MTURN$ expression had favorable survival in our study. $RGS20$ is suggested to promote cellular characteristics that contribute to metastasis, including enhanced cell aggregation, motility, and invasion. Selective inhibition of $RGS20$ expression may represent an alternative means to suppress metastasis (Yang et al., 2016). Its high expression is significantly associated with progression and prognosis of triple-negative breast cancer (Li et al., 2017). Additionally, our study showed similar results in LUAD patients. Though there lacks explicit evidence of relevance between these genes and smoking, what we found may inspire functional studies of these potential genes and further help to complete a picture of the mechanism pathway of $cg02268510_{SIPA1L3}$ and smoking cessation interaction on LUAD survival.

Our study has some significant strengths. First, this is the first study to investigate the interaction between DNA methylation and smoking cessation on lung cancer survival on an epigenome-wide scale, which provides new evidence to account for the
missing heritability of complex diseases (Trerotola et al., 2015). Second, the two-stage study design we used to exhaustively search for interactions, as well as the sensitivity analysis, is quite conservative in controlling for false positives. Third, our study included a large sample size to analyze DNA methylation–smoking cessation interactions of early-stage NSCLC prognosis, providing an opportunity to identify complex associations with small–medium effect size.

Despite the strengths of our study, we acknowledge some limitations. First, data measured categorical smoking cessation rather than smoking pack-years, which may render less power in the study. Second, smoking cessation was collected at the time of diagnosis and was not re-assessed during follow-up. Previous studies have found that “former smokers” might more accurately represent a mixed exposure status, since quitters are more likely to relapse (Hughes et al., 2004; Walker et al., 2006b). Thus, we likely underestimated the benefits of smoking cessation. Third, the association between cg02268510SIPA1L3 and expression of several genes requires more biological evidence, though methylation is believed to play a crucial role in regulating gene expression (Bird, 2007) and further influence disease gene function (Schübeler, 2015), cell differentiation, or reprogramming (Khavari et al., 2010). Thus, functional experiments are warranted to confirm these associations, so our findings should be biologically interpreted with caution thus far. In addition, our study consisted mainly of a Caucasian population (89.19%), since TCGA data contained only ~10% non-Caucasian samples. Our results should therefore be translated with
caution for other populations. Lastly, the censored rate of survival time for the TCGA population is relatively high, since early-stage NSCLC patients need longer follow-up time. Thus, the validation phase using TCGA population had low statistical power. However, we still successfully replicated one significant interaction, indicating a quite conservative and robust result (Leung et al., 1997; Watt et al., 1996).

5. Conclusion

This epigenome-wide DNA methylation–smoking cessation interaction analysis of early-stage NSCLC identified one LUAD-specific CpG probe, cg02268510_<sub>SIPA1L3</sub>, which could significantly modify effects of smoking cessation on lung cancer survival. Smoking cessation benefited survival of LUAD patients with low methylation at cg02268510_<sub>SIPA1L3</sub>. These results have implications for not only smoking cessation after diagnosis, but also possible methylation-specific drug targeting.

Acknowledgements

The authors thank all participants enrolled in this study. This study was supported by the National Key Research and Development Program of China (2016YFE0204900 to F.C.); the National Natural Science Foundation of China (81530088 and 81473070 to F.C., 81602940 to L. Liu); the US National Institutes of Health (CA209414, CA092824, and ES000002 to D.C.C.); the Natural Science Foundation of the Jiangsu
Higher Education Institutions of China (18KJB310011 to R.Z. and 14KJA310002 to F.C.); China Postdoctoral Science Foundation (2018M633767 to R.Z.) and the Priority Academic Program Development of Jiangsu Higher Education Institutions. Y.W. and R.Z. were partially supported by the Outstanding Young Teachers Training Program of Nanjing Medical University.

Conflict of interest

The authors have no conflicts to declare.

Author contributions

R.Z., L.Lai, Y.W., and F.C. contributed to the study design. R.Z., S.M., T.F., M.M.B., A.K., M.P., J.S., A.H., M.E., A.S., and L.S. contributed to data collection. R.Z., L.Lai, and Y.W. performed statistical analysis and interpretation and drafted the manuscript. X.D., J.H., D.Y., C.C, L.Lai, Y.Z., H.H., S.S., L.W., X.C., Y.G. and L.Liu revised the manuscript. All authors contributed to critical revision of the manuscript and approved its final version. Financial support and study supervision were provided by R.Z., F.C. and D.C.C.
Consent for publication

All participants or their surrogate care providers gave written informed consent. All authors have reviewed the manuscript and consented for publication.

Web resources

TCGA: https://portal.gdc.cancer.gov

References

Anto, R.J., Mukhopadhyay, A., Shishodia, S., Gairola, C.G., Aggarwal, B.B., 2002. Cigarette smoke condensate activates nuclear transcription factor-kappaB through phosphorylation and degradation of IkappaB(alpha): correlation with induction of cyclooxygenase-2. Carcinogenesis 23, 1511-1518.

Aran, D., Sirota, M., Butte, A.J., 2015. Systematic pan-cancer analysis of tumour purity. Nat. Commun. 6, 8971.

Athanasios, V., Thomais, V., Konstantinos, F., Spyridon, L., 2013. Pulmonary oxidative stress, inflammation and cancer: respirable particulate matter, fibrous dusts and ozone as major causes of lung carcinogenesis through reactive oxygen species mechanisms. Int. J. Environ. Res. Public Health 10, 3886-3907.

Balduyck, B., Sardari, N.P., Cogen, A., Dockx, Y., Lauwers, P., Hendriks, J., Van, S.P., 2011. The effect of smoking cessation on quality of life after lung cancer surgery. Eur. J. Cardiothorac. Surg. 40, 1437-1438.

Baser, S., Shannon, V.R., Eapen, G.A., Jimenez, C.A., Onn, A., Lin, E., Morice, R.C., 2006. Smoking Cessation After Diagnosis of Lung Cancer Is Associated With a Beneficial Effect on Performance Status. Chest 130, 1784-1790.

Besse, B., Adjej, A., Baas, P., Meldgaard, P., Nicolson, M., Paz-Ares, L., Reck, M., Smit, E.F., Syrigos, K., Stahel, R., 2014. 2nd ESMO Consensus Conference on Lung Cancer: non-small cell lung cancer first-line/second and further lines of treatment in advanced disease. Ann. Oncol. 25, 1475.

Bhatt, V.R., Batra, R., Silberstein, P.T., Jr, F.R.L., Ganti, A.K., 2015. Effect of smoking on survival from non-small cell lung cancer: a retrospective Veterans’ Affairs Central Cancer Registry (VACCR) cohort analysis. Med. Oncol. 32, 1-6.

Bird, A., 2007. Perceptions of epigenetics. Nature 447, 396.
Bjaanaes, M.M., Fleischer, T., Halvorsen, A.R., Daunay, A., Busato, F., Solberg, S., Jorgensen, L., Kure, E., Edvardsen, H., Borresen-Dale, A.L., 2016. Genome-wide DNA methylation analyses in lung adenocarcinomas: Association with EGFR, KRAS and TP53 mutation status, gene expression and prognosis. Mol. Oncol. 10, 330.

Chaturvedi, M.M., Sung, B., Yadav, V.R., Kannappan, R., Aggarwal, B.B., 2010. NF-κB addiction and its role in cancer: 'one size does not fit all'. Oncogene 30, 1615-1630.

Chen, S.J., Zhou, G.B., Zhang, X.W., Mao, J.H., De, T.H., Chen, Z., 2011. From an old remedy to a magic bullet: molecular mechanisms underlying the therapeutic effects of arsenic in fighting leukemia. Blood 117, 6425-6437.

Chen, Y.A., Mathieu, L., Sanaa, C., Butcher, D.T., Daria, G., Zanke, B.W., Steven, G., Hudson, T.J., Rosanna, W., 2013. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8, 203-209.

Chin, A.I., Dempsey, P.W., Bruhn, K., Miller, J.F., Xu, Y., Cheng, G., 2002. Involvement of receptor-interacting protein 2 in innate and adaptive immune responses. Nature 416, 190.

Crusz, S.M., Balkwill, F.R., 2015. Inflammation and cancer: advances and new agents. Nat. Rev. Clin. Oncol. 12, 584.

Cusick, J.K., Mustian, A., Goldberg, K., Reyland, M.E., 2010. RELT induces cellular death in HEK 293 epithelial cells. Cell. Immunol. 261, 1.

Deponte, M., 2013. Glutathione catalysis and the reaction mechanisms of glutathione-dependent enzymes. Biochim. Biophys. Acta 1830, 3217-3266.

Egger, G., Liang, G., Aparicio, A., Jones, P.A., 2004. Epigenetics in human disease and prospects for epigenetic therapy. Nature 429, 457.

Feinberg, A.P., Tycko, B., 2004. The history of cancer epigenetics. Nat. Rev. Cancer 4, 143.

Flanders, W.D., Lally, C.A., Zhu, B.P., Henley, S.J., Thun, M.J., 2003. Lung cancer mortality in relation to age, duration of smoking, and daily cigarette consumption: results from Cancer Prevention Study II. Cancer Res. 63, 6556.

Fraga, M.F., Esteller, M., 2007. Epigenetics and aging: the targets and the marks. Trends Genet. 23, 413-418.

Guo, W., Zhu, T., Dong, Z., Cui, L., Zhang, M., Kuang, G., 2013. Decreased expression and aberrant methylation of Gadd45G is associated with tumor progression and poor prognosis in esophageal squamous cell carcinoma. Clin. Exp. Metastasis 30, 977-992.

Guo, Y., Zhang, R., Shen, S., Wei, Y., Salama, S.M., Fleischer, T., Bjaanaes, M.M., Karlsson, A., Planck, M., Su, L., Zhu, Z., Staaf, J., Helland, A., Esteller, M., Christiani, D.C., 2018. DNA Methylation of LRRC3B: A Biomarker for Survival of Early-Stage Non-Small Cell Lung Cancer Patients. Cancer Epidemiol. Biomarkers Prev. 27, 1527-1535.

Han, Z.T., Zhu, X.X., Yang, R.Y., Sun, J.Z., Tian, G.F., Liu, X.J., Cao, G.S., Newmark, H.L., Conney, A.H., Chang, R.L., 1998. Effect of intravenous infusions of
12-O-tetradecanoylphorbol-13-acetate (TPA) in patients with myelocytic leukemia: preliminary studies on therapeutic efficacy and toxicity. Proc. Natl. Acad. Sci. U. S. A. 95, 5357.

He, L., He, T., Farrar, S., Ji, L., Liu, T., Ma, X., 2017. Antioxidants Maintain Cellular Redox Homeostasis by Elimination of Reactive Oxygen Species. Cell. Physiol. Biochem. 44, 532-553.

Hughes, J.R., Keely, J., Naud, S., 2004. Shape of the relapse curve and long-term abstinence among untreated smokers. Addiction 99, 29.

Jaafer, R., Mnich, K., Dolan, S., Hillis, J., Almanza, A., Logue, S.E., Samali, A., Gorman, A.M., 2018. RIP2 enhances cell survival by activation of NF-κB in triple negative breast cancer cells. Biochem. Biophys. Res. Commun. 497.

Jaifer, H.D., Bhola, S., Kalburgi, V., Warad, S., Kokatnur, V.M., 2015. The influence of cigarette smoking on blood and salivary super oxide dismutase enzyme levels among smokers and nonsmokers—A cross sectional study. J. Tradit. Complement. Med. 5, 100-105.

Karlsson, A., Jönsson, M., Lauss, M., Brunnström, H., Jönsson, P., Borg, A., Jönsson, G., Ringnér, M., Planck, M., Staaf, J., 2014. Genome wide DNA Methylation Analysis of Lung Carcinoma Reveals One Neuroendocrine and Four Adenocarcinoma Epitypes Associated with Patient Outcome. Clin. Cancer Res. 20, 6127-6140.

Khavari, D.A., Sen, G.L., Rinn, J.L., 2010. DNA methylation and epigenetic control of cellular differentiation. Cell Cycle 9, 3880-3883.

Kooistra, M.R., Dubé, N., Bos, J.L., 2007. Rap1: a key regulator in cell-cell junction formation. J. Cell Sci. 120, 17.

Koshiaris, C., Aveyard, P., Oke, J., Ryan, R., Szatkowski, L., Stevens, R., Farley, A., 2017. Smoking cessation and survival in lung, upper aero-digestive tract and bladder cancer: cohort study. Br. J. Cancer 117, 1224-1232.

Lacombe, L., Fradet, V., Levesque, E., Pouliot, F., Larue, H., Bergeron, A., Hovington, H., Caron, A., Ngoule-Makao, M., Harvey, M., Fradet, Y., Guillemette, C., 2016. Phase II Drug-Metabolizing Polymorphisms and Smoking Predict Recurrence of Non-Muscle-Invasive Bladder Cancer: A Gene-Smoking Interaction. Cancer Prev. Res. (Phila.) 9, 189-195.

Lee, K.W.K., Pausova, Z., 2013. Cigarette smoking and DNA methylation. Front Genet 4, 132.

Leung, K.M., Elashoff, R.M., Afifi, A.A., 1997. Censoring issues in survival analysis. Annu. Rev. Public Health 18, 83.

Li, Q., Jin, W., Cai, Y., Yang, F., Chen, E., Ye, D., Wang, Q., Guan, X., 2017. Regulator of G protein signaling 20 correlates with clinicopathological features and prognosis in triple-negative breast cancer. Biochem. Biophys. Res. Commun. 485, 693-697.

Liebermann, D.A., Tront, J.S., Sha, X., Mukherjee, K., Mohamedhadley, A., Hoffman, B., 2011. Gadd45 Stress Sensors in Malignancy and Leukemia. Crit. Rev. Oncog. 16, 129.
Marabita, F., Almgren, M., Lindholm, M.E., Ruhrmann, S., Fagerströmbillai, F., Jagodic, M., Sundberg, C.J., Ekström, T.J., Teschendorff, A.E., Tegnér, J., 2013. An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. Epigenetics 8, 333.

Mcnerney, M.E., Godley, L.A., Le, B.M., 2017. Therapy-related myeloid neoplasms: when genetics and environment collide. Nat. Rev. Cancer 17, 513.

Mochizuki, N., Yamashita, S., Kurokawa, K., Ohba, Y., Nagai, T., Miyawaki, A., Matsuda, M., 2001. Spatio-temporal images of growth-factor-induced activation of Ras and Rap1. Nature 411, 1065.

Moon, E.Y., Lee, J.H., Lee, J.W., Song, J.H., Pyo, S., 2011. ROS/Epac1-mediated Rap1/NF-kappaB activation is required for the expression of BAFF in Raw264.7 murine macrophages. Cell. Signal. 23, 1479-1488.

Nia, P.S., Weyler, J., Colpaert, C., Vermeulen, P., Marck, E.V., Schil, P.V., 2005. Prognostic value of smoking status in operated non-small cell lung cancer. Lung Cancer 47, 351-359.

Park, E.R., Japuntich, S.J., Rigotti, N.A., Traeger, L., He, Y., Wallace, R.B., Malin, J.L., Zallen, J.P., Keating, N.L., 2012. A snapshot of smokers after lung and colorectal cancer diagnosis. Cancer 118, 3153.

Parsons, A., Daley, A., Begh, R., Aveyard, P., 2010. Influence of smoking cessation after diagnosis of early stage lung cancer on prognosis: systematic review of observational studies with meta-analysis. BMJ. 340, 251-251.

Remans, P.H.J., Gringhuis, S.I., Laar, J.M., Van, Sanders, M.E., Zwartkruis, F.J.T., Levarht, E.W.N., Marcela, R., Coffer, P.J., Breedveld, F.C., 2004. Rap1 signaling is required for suppression of Ras-generated reactive oxygen species and protection against oxidative stress in T lymphocytes. J. Immunol. 173, 920-931.

Sandoval, J., Mendez-Gonzalez, J., Nadal, E., Chen, G., Carmona, F.J., Sayols, S., Moran, S., Heyn, H., Vizoso, M., Gomez, A., 2013. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. J. Clin. Oncol. 31, 4140.

Schübeler, D., 2015. Function and information content of DNA methylation. Nature 517, 321.

Shen, S., Zhang, R., Guo, Y., Loehrre, E., Wei, Y., Zhu, Y., Yuan, Q., Moran, S., Fleischer, T., Bjaanaes, M.M., Karlsson, A., Planck, M., Staaaf, J., Helland, A., Esteller, M., Su, L., Chen, F., Christiani, D.C., 2018. A multi-omic study reveals BTG2 as a reliable prognostic marker for early-stage non-small-cell lung cancer. Mol. Oncol. 12, 913-924.

Shui, I.M., Wong, C.J., Zhao, S., Kolb, S., Ebot, E.M., Geybels, M.S., Rubicz, R., Wright, J.L., Lin, D.W., Klotzle, B., 2016. Prostate tumor DNA methylation is associated with cigarette smoking and adverse prostate cancer outcomes. Cancer 122, 2168-2177.

Siegel, R.L., Miller, K.D., Jemal, A., 2017. Cancer statistics, 2018. CA Cancer J. Clin. 67, 7.
Slade, E., Kraft, P., 2016. Leveraging Methylome-Environment Interaction to Detect Genetic Determinants of Disease. Hum. Hered. 81, 26-34.

Suk, H.R., Zhou, W., Cogandrew, T., Liu, G., Su, L., Neuberg, D., Lynch, T.J., Wain, J.C., Christiani, D.C., 2006. MDM2 polymorphism and recurrence-free and overall survival in early stage non-small cell lung cancer (NSCLC). J. Clin. Oncol.

Sun, X., Lu, B., Hu, B., Xiao, W., Li, W., Huang, Z., 2014. Novel function of the chromosome 7 open reading frame 41 gene to promote leukemic megakaryocyte differentiation by modulating TPA-induced signaling. Blood Cancer J. 4, e198.

Teo, H., Ghosh, S., Luensch, H., Ghosh, A., Wong, E.T., Malik, N., Orth, A., Jesus, P.D., Perry, A.S., Oliver, J.D., 2010. Telomere-independent Rap1 is an IKK adaptor and regulates NF-κB-dependent gene expression. Nat. Cell Biol. 12, 758-767.

Trerotola, M., Relli, V., Simeone, P., Alberti, S., 2015. Epigenetic inheritance and the missing heritability. Hum. Genomics 9, 17.

Tsurutani, J., Castillo, S.S., Brognard, J., Granville, C.A., Zhang, C., Gills, J.J., Sayyah, J., Dennis, P.A., 2005. Tobacco components stimulate Akt-dependent proliferation and NFκB-dependent survival in lung cancer cells. Carcinogenesis 26, 1182-1195.

Ueno, H., Pradhan, S., Schlessel, D., Hirasawa, H., Sumpio, B.E., 2006. Nicotine enhances human vascular endothelial cell expression of ICAM-1 and VCAM-1 via protein kinase C, p38 mitogen-activated protein kinase, NF-κB, and AP-1. Cardiovasc. Toxicol. 6, 39-50.

Walker, M.S., Vidrine, D.J., Gritz, E.R., Larsen, R.J., Yan, Y., Govindan, R., Fisher, E.B., 2006a. Smoking Relapse during the First Year after Treatment for Early-Stage Non–Small-Cell Lung Cancer. Cancer Epidemiol. Biomarkers Prev. 15, 2370.

Walser, T., Cui, X., Yanagawa, J., Lee, J.M., Heinrich, E., Lee, G., Sharma, S., Dubinett, S.M., 2008. Smoking and lung cancer: the role of inflammation. Proc Am Thorac Soc 5, 811-815.

Wan, E.S., Weiliang, Q., Andrea, B., Carey, V.J., Helene, B., Rennard, S.I., Alvar, A., Wayne, A., Lomas, D.A., Demeo, D.L., 2012. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. Hum. Mol. Genet. 21, 3073.

Watt, D.C., Aitchison, T.C., Mackie, R.M., Sirel, J.M., 1996. Survival analysis: the importance of censored observations. Melanoma Res. 6, 379-385.

Wei, Y., Liang, J., Zhang, R., Guo, Y., Shen, S., Su, L., Lin, X., Moran, S., Helland, Å., Bjaanæs, M.M., 2018. Epigenetic modifications in KDM lysine demethylases associate with survival of early-stage NSCLC. Clin. Epigenetics 10, 41.

Wright, J., 2013. Epigenetics: reversible tags. Nature 498, 10-11.

Yang, L., Lee, M.M., Leung, M.M., Wong, Y.H., 2016. Regulator of G protein signaling 20 enhances cancer cell aggregation, migration, invasion and adhesion. Cell. Signal. 28, 1663-1672.
Ying, J., Srivastava, G., Hsieh, W.-S., Gao, Z., Murray, P., Liao, S.-K., Ambinder, R., Tao, Q., 2005. The stress-responsive gene GADD45G is a functional tumor suppressor, with its response to environmental stresses frequently disrupted epigenetically in multiple tumors. Clin. Cancer Res. 11, 6442-6449. 
Zaghlool, S.B., Mashael, A.S., Muftah, W.A., Al, Pankaj, K., Mario, F., Karsten, S., 2015. Association of DNA methylation with age, gender, and smoking in an Arab population. Clin. Epigenetics 7, 1-12. 
Zhang, N., Wu, H.J., Zhang, W., Wang, J., Wu, H., Zheng, X., 2015. Predicting tumor purity from methylation microarray data. Bioinformatics 31, 3401. 
Zhang, R., Lai, L., He, J., Chen, C., You, D., Duan, W., Dong, X., Zhu, Y., Lin, L., Shen, S., Guo, Y., Su, L., Shafer, A., Moran, S., Fleischer, T., Bjaanaaes, M.M., Karlsson, A., Planck, M., Staaf, J., Helland, A., Esteller, M., Wei, Y., Chen, F., Christiani, D.C., 2019. EGLN2 DNA methylation and expression interact with HIF1A to affect survival of early-stage NSCLC. Epigenetics, 1-12. 
Zhang, Y.L., Wang, R.C., Cheng, K., Ring, B.Z., Li, S., 2017. Roles of Rap1 signaling in tumor cell migration and invasion. Cancer Biol. Med. 14, 90-99. 
Zheng, X., Zhang, N., Wu, H.J., Wu, H., 2017. Estimating and accounting for tumor purity in the analysis of DNA methylation data from cancer studies. Genome Biol. 18, 17.

Supporting information

Table S1. Demographic and clinical characteristics of early-stage NSCLC patients in TCGA dataset

Table S2. Results for 15 methylation–smoking interactions using a two-stage association study

Table S3. Smoking related characteristics of former and current smokers in early-stage LUAD

Figure S1. Quality control processes for DNA methylation chip data.
Figure S2. Manhattan plot of DNA methylation–smoking cessation interaction $P$-values (A) and main effect $P$-values (B) derived from histology-stratified Cox proportional hazards model in the discovery phase.

Figure S3. Fixed-effect meta-analysis of interaction between DNA methylation of cg02268510 and smoking cessation for LUAD patients from five centers.

Figure S4. Linear regression analysis between methylation of cg02268510 and age (A) as well as smoking related variables: pack-year of smoking (B), year of smoking (C), and year of smoking cessation (D), adjusted for age, sex, smoking status, clinical stage, and study center.

Figure S5. Genome-wide methylation transcription analysis results from the TCGA cohort. (A) Circos plot of genome-wide gene expression. For plots in B–H, left panels show correlation of (B) CADD45G, (C) MTURN, (D) TMEM200B, (E) RGS20, (F) REL1, (G) PGM2, and (H) RIPK2 expression (X-axis) with methylation of cg02268510 (Y-axis). Right panels show Kaplan-Meier survival plots of gene expression divided into low and high groups by median value.
Table 1. Demographic and clinical characteristics of former and current smokers with early-stage NSCLC in five study centers.

| Variable                        | Discovery phase | Validation phase | Combined data |
|---------------------------------|-----------------|------------------|---------------|
|                                 | Harvard (N = 133) | Spain (N = 196) | Norway (N = 116) | Sweden (N = 85) | Discovery: All (N = 530) | TCGA (N = 562) | Overall (N = 1092) |
| Age (years), mean ± SD          | 68.04 ± 9.39    | 65.69 ± 10.41    | 64.89 ± 8.94    | 65.65 ± 9.69    | 66.10 ± 9.78       | 66.55 ± 9.39 | 66.33 ± 9.58     |
| Sex, n (%)                      |                 |                  |                |                |                   |               |                 |
| Female                          | 56 (42.11)      | 78 (39.80)       | 59 (50.86)      | 42 (49.41)      | 235 (44.34)        | 217 (38.61)  | 452 (41.39)      |
| Male                            | 77 (57.89)      | 118 (60.20)      | 57 (49.14)      | 43 (50.59)      | 295 (55.66)        | 345 (61.39)  | 640 (58.61)      |
| Smoking cessation, n (%)        |                 |                  |                |                |                   |               |                 |
| No                              | 52 (39.10)      | 71 (36.22)       | 42 (36.21)      | 31 (36.47)      | 196 (36.98)        | 168 (26.89)  | 364 (33.33)      |
| Yes                             | 81 (60.90)      | 120 (61.22)      | 74 (63.79)      | 54 (63.53)      | 329 (62.08)        | 376 (66.90)  | 705 (64.56)      |
| Unknown                         | 0               | 5                | 0              | 0              | 5                 | 18            | 23              |
| Race, n (%)                     |                 |                  |                |                |                   |               |                 |
| White                           | 133 (100)       | 196 (100)        | 116 (100)       | 85 (100)        | 530 (100)          | 444 (79.00)  | 974 (89.19)      |
| Black or African American       | 0               | 0                | 0              | 0              | 0                 | 54 (9.61)    | 54 (4.95)       |
| Asian                           | 0               | 0                | 0              | 0              | 5                 | 5 (0.89)     | 5 (0.46)        |
| Unknown                         | 0               | 0                | 0              | 0              | 0                 | 59            | 59              |
| TNM stage, n (%)                |                 |                  |                |                |                   |               |                 |
| I                               | 87 (65.41)      | 160 (81.63)      | 80 (68.97)      | 79 (92.94)      | 406 (76.60)        | 358 (63.70)  | 764 (69.96)      |
| II                              | 46 (34.59)      | 36 (18.37)       | 36 (31.03)      | 6 (7.06)        | 124 (23.40)        | 204 (36.30)  | 328 (30.04)      |
| Histology, n (%)                |                 |                  |                |                |                   |               |                 |
| LUAD                            | 79 (59.40)      | 155 (79.08)      | 116 (100)       | 63 (74.12)      | 413 (77.92)        | 285 (50.71)  | 698 (63.92)      |
| LUSC                            | 54 (40.60)      | 41 (20.92)       | 0              | 22 (25.88)      | 117 (22.08)        | 277 (49.29)  | 394 (36.08)      |
| Chemotherapy, n (%)             |                 |                  |                |                |                   |               |                 |
| No                              | 125 (93.98)     | 157 (80.10)      | 86 (74.14)      | 51 (60.00)      | 419 (79.06)        | 178 (31.67)  | 597 (54.67)      |
| Yes                             | 8 (6.02)        | 15 (7.65)        | 30 (25.86)      | 6 (7.06)        | 59 (11.13)         | 53 (9.43)    | 112 (10.26)      |
|                  | Radiotherapy, n (%) | Adjuvant therapy, n (%) | Survival year |
|------------------|---------------------|-------------------------|---------------|
|                  | No                  | Yes                     |               |
|                  |                    |                         | Median (95% CI)  |
|                  | 115 (86.47)         | 23 (17.29)              | 6.44 (4.93–7.44) |
|                  | 161 (82.14)         | 24 (12.24)              | 6.53 (5.06–8.70) |
|                  | 115 (99.14)         | 31 (26.72)              | 7.34 (6.70–7.98) |
|                  | 57 (67.06)          | 6 (7.06)                | 5.89 (4.18–8.82) |
|                  | 448 (84.53)         | 84 (15.85)              | 7.12 (6.23–7.95) |
|                  | 220 (39.15)         | 59 (10.50)              | 4.54 (3.69–5.41) |
|                  | 668 (61.17)         | 143 (13.10)             | 6.23 (5.55–7.20) |
|                  |                    |                         |               |
|                  | Yes                 |                         | Censored rateb, %  |
|                  | 18 (13.53)          |                         | 14.29          |
|                  | 11 (5.61)           |                         | 54.59          |
|                  | 1 (0.86)            |                         | 68.97          |
|                  | 0                   |                         | 40.00          |
|                  | 30 (5.66)           |                         | 45.30          |
|                  | 11 (1.96)           |                         | 76.16          |
|                  | 41 (3.75)           |                         | 61.17          |

aSpain is a collaborative study center, containing samples from Spain, Italy, UK, France, and USA.
bCensored rate is the proportion of samples lost to follow-up or alive at the study end.
cRestricted mean survival time is provided because median was not available.

*Statistically significant difference (P ≤ 0.05) was observed between combined discovery set and validation set (TCGA).

LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma; CI: confidence interval; TCGA: The Cancer Genome Atlas.
### Table 2. Joint effect and interaction of elevated methylation and smoking cessation on the prognosis of early-stage LUAD

| Effect type | Medium–high methylation | Smoking cessation | Number | HR (95% CI) | P |
|-------------|-------------------------|-------------------|--------|-------------|---|
| Main effect 1 | No | No | 71 | Ref. | |
| Main effect 2 | Yes | Yes | 157 | 0.5506 (0.3609, 0.8400) | 5.62×10⁻³ |
| Joint effect | Yes | Yes | 158 | 0.5214 (0.3369, 0.8070) | 3.48×10⁻³ |
| Interaction | Yes | Yes | 299 | 0.6268 (0.4251, 0.9243) | 1.84×10⁻² |

*Interaction = Joint effect ÷ (Main effect 1 × Main effect 2). 2.1835 ÷ (0.5506 × 0.5214).*

**Notes:**
- Patients were categorized into two groups (medium–high vs low) by tertiles of cg02268510_SIPA1L3 methylation level.
- Main effects of elevated methylation and smoking cessation, and their joint effect and interaction were derived from the Cox proportional hazards model adjusted for covariates.
Figure legends

Figure 1. Flow chart of study design and statistical analyses.

Figure 2. DNA methylation and smoking cessation interaction on survival of lung adenocarcinoma (LUAD) patients. (A) Hazard ratio (HR) of smoking cessation estimated based on methylation level of cg02268510. The shallow area represents 95% confidence interval, with dark grey areas indicating significance. Histogram on the top shows the distribution of methylation. (B) Forest plots of the effects of smoking cessation among combined LUAD populations with low, medium, or high methylation of cg02268510. $P_{\text{heterogeneity}}$ was used to evaluate heterogeneity of HRs across groups. (C) Kaplan-Meier survival curves of current and former smokers among LUAD patients with varying methylation levels.

Figure 3. Kaplan-Meier survival curves of lung adenocarcinoma (LUAD) patients categorized into low and medium–high methylation groups according to tertiles in different smoking cessation groups (A) current smokers (No) and (B) former smokers (Yes). Hazard ratio (HR), 95% confidence interval, and $P$-value were derived from the Cox proportional hazards model adjusted for age, sex, clinical stage, and study center.
**Figure 4.** Association between DNA methylation of cg02268510 and expression of corresponding gene *SIPA1L3*. The $\beta$ coefficient and $P$ value were based on linear regression analysis adjusted for age, sex, smoking status, clinical stage. Gene expression was log2-transformed before analysis.

**Figure 5.** Diagram for pathway of DNA methylation–smoking cessation interaction effect on survival for lung adenocarcinomas (LUAD) patients.
Discovery phase
530 former or current smokers with early-stage lung cancer from multi-centers
311,391 tumor tissue DNA methylation CpG probes

Histology-stratified multivariate Cox proportional hazards model
FDR-q ≤ 0.05

15 CpG probes passed the threshold

Validation phase
562 early-stage lung cancer patients from TCGA
Interaction P ≤ 0.05
Consistent effect

1 CpG probe retained for the following analyses

Stratified analysis
Reanalysis in LUAD and LUSC patients, respectively

P_{discovery} ≤ 0.05
P_{discrimination} ≤ 0.05

1 LUAD-specific CpG probe:
cg02268510

Sensitivity analysis
Further adjusted for tumor purity in the model

Genome-wide methylation transcription analysis
Methylation
cg02268510 methylation
FDR-q ≤ 0.05

Gene expression
Expression of 698 genes correlated with cg02268510 methylation
FDR-q ≤ 0.05

Survival
Expression of 7 genes associated with overall survival

A
B
C

Molecular Oncology (2019) © 2019 The Authors. Published by FEBS Press and John Wiley & Sons Ltd.
A: \[ HR = 1.94 \ (1.01-3.09) \ P = 4.89 \times 10^{-3} \]

B: \[ HR = 0.97 \ (0.71-1.33) \ P = 0.847 \]

\[ \beta = -0.02 \ P = 0.015 \]
