Received: 1 August 2019 | Revised: 2 October 2019 | Accepted: 22 October 2019
DOI: 10.1002/mc.23132

RESEARCH ARTICLE

Novel genetic variants in *HDAC2* and *PPARGC1A* of the CREB-binding protein pathway predict survival of non-small-cell lung cancer

Dongfang Tang[1,2,3] | Yu Chen Zhao[2,3] | Danwen Qian[2,3] | Hongliang Liu[2,3] | Sheng Luo[4] | Edward F. Patz[2,5] | Patricia G. Moorman[2,6] | Li Su[7] | Sipeng Shen[7] | David C. Christiani[7,8] | Carolyn Glass[7,9] | Wen Gao[1] | Qingyi Wei[2,3,10]

[1]Department of Thoracic Surgery, Shanghai Key Laboratory of Clinical Geriatric Medicine, Huadong Hospital Affiliated to Fudan University, Shanghai, China
[2]Duke Cancer Institute, Duke University Medical Center, Durham, North Carolina
[3]Department of Population Health Sciences, Duke University School of Medicine, Durham, North Carolina
[4]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina
[5]Department of Radiology and Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina
[6]Department of Community and Family Medicine, Duke University Medical Center, North Carolina
[7]Department of Environmental Health and Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts
[8]Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts
[9]Department of Pathology, Duke University School of Medicine, Durham, North Carolina
[10]Department of Medicine, Duke University School of Medicine, Durham, North Carolina

Correspondence
Wen Gao, Department of Thoracic Oncology, Huadong Hospital Affiliated to Fudan University, 200040 Shanghai, China.
Email: gaowen5921@163.com

Qingyi Wei, Duke Cancer Institute, Duke University Medical Center, and Department of Population Health Sciences, Duke University School of Medicine, 905S LaSalle Street, Durham, NC 27710.
Email: qingyi.wei@duke.edu

Funding information
National Institute of Health, Grant/Award Numbers: 5U01CA209414, CA074386, CA092824, CA090578; The short-term international training program for PhD training from Fudan University; Duke Cancer Institute as part of the P30 Cancer Center Support Grant, Grant/Award Number: NIH/NCI CA014236; Fudan University, China; V Foundation for Cancer Research

Abstract

The CREB-binding protein (CBP) pathway plays an important role in transcription and activity of acetyltransferase that acetylates lysine residues of histones and nonhistone proteins. In the present study, we hypothesized that genetic variants in the CBP pathway genes played a role in survival of non-small-cell lung cancer (NSCLC). We tested this hypothesis using the genotyping data from the genome-wide association study (GWAS) dataset from the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. In the single-locus analysis, we evaluated associations between 13 176 (1107 genotyped and 12 069 imputed) single-nucleotide polymorphisms (SNPs) in 72 genes and survival of 1185 patients with NSCLC. The identified 106 significant SNPs in the discovery were further validated in additional genotyping data from another GWAS dataset of 984 patients with NSCLC in the Harvard Lung Cancer Susceptibility Study. The combined results of two datasets showed that two independent, potentially functional SNPs (i.e., *HDAC2* rs13213007G>A and *PPARGC1A* rs60571065T>A) were significantly associated with NSCLC overall survival, with a combined hazards ratio (HR) of 1.26 (95% confidence interval (CI), 1.09-1.45; $P = .002$) and 1.23 (1.04-1.47; $P = .017$), respectively. Furthermore, we performed an expression

Abbreviations: AUC, Area under the curve; BFDP, Bayesian false discovery probability; CBP, CREB-binding protein; CI, Confidence interval; DSS, Disease-specific survival; eQTL, Expression quantitative trait loci; GWAS, Genome-Wide Association Study; HLCS, Harvard Lung Cancer Susceptibility study; HR, Hazards ratio; LD, Linkage disequilibrium; NSCLC, Non-small-cell lung cancer; OS, Overall survival; PLCO, Prostate, Lung, Colorectal, and Ovarian cancer screening trial; ROC, Receiver operating characteristic; SNP, Single-nucleotide polymorphism; TCGA, The Cancer Genome Atlas.

 © 2019 Wiley Periodicals, Inc.

quantitative trait loci analysis and found that the survival-associated *HDAC2* rs13213007A allele (GA+AA), but not *PPARGC1A* rs60571065A allele (TA+AA), was significantly associated with increased messenger RNA expression levels of *HDAC2* in 373 lymphoblastoid cell lines. These results indicate that the *HDAC2* rs13213007A allele is a potential predictor of NSCLC survival, likely by altering the *HDAC2* expression.

KEYWORDS
CREB-binding protein pathway, genetic susceptibility, non-small-cell lung cancer, single-nucleotide polymorphisms

1 | INTRODUCTION

Lung cancer has been the leading cause of cancer-related mortality since 1985, with over a million deaths each year worldwide.[1] In 2018, it is estimated that there were approximately 234 030 new cases and 154 050 lung cancer deaths in the United States (https://seer.cancer.gov/statfacts/html/ lungb.html). Non-small-cell lung cancer (NSCLC) is the most common histological type of lung cancer, accounting for approximately 80% to 85% of all lung cancer cases diagnosed.[2] Over the past few decades, advances in surgical procedures, chemo-radiotherapy, targeted therapy, and more recently immunotherapy have contributed to modest improvements in survival of lung cancer patients[3]; yet, the five-year survival rate is still at an underwhelming 15% to 20%. Previous studies have found that molecular and genetic factors may play an important role in lung cancer progression and outcomes[4]; thus, it is important to search for additional molecular markers that may predict survival of patients with NSCLC, who would benefits from individualized therapies.

Single-nucleotide polymorphisms (SNPs) can affect gene expression and functions and thus may be associated with susceptibility to cancer development and progression.[5] To date, several genome-wide association studies (GWASs) of lung cancer have discovered SNPs that are associated with cancer risk, but few have found markers predictive of outcomes of patients with NSCLC. The vast majority of SNPs under GWAS investigations did not reach the stringent genome-wide level of significance, and most of those identified as significant SNPs are located within noncoding regions of the genome, making such discoveries fundamentally difficult to comprehend and harder to reveal the underlying mechanisms. As a result, it is recommended that research strategies in the post-GWAS era should include "discovery, expansion, and replication," "biological studies," and "epidemiologic studies" (https://epi.grants.cancer.gov/ gameon/ #funded).[6] By using a hypothesis-driven approach of the pathway gene analysis pooling together published GWAS datasets, one may have a better chance with an improved statistical power to identify novel loci with minor yet detectable effects, to further examine functional consequences of such novel loci, and thus to unravel possible mechanisms underlying the observed associations.

The CREB-binding protein (CBP) pathway regulates posttranslational modification with the activity of acetyltransferase that acetylates lysine residues of histones and nonhistone proteins, such as p53.[7] The CBP pathway is also involved in basic cellular functions, including cell growth, differentiation, DNA repair and apoptosis.[8,9] Studies have revealed oncogenic roles of the CBP pathway genes; for example, it has been shown that CBP is highly expressed in lung cancer cells and tumor tissues, upregulating hTERT expression and promoting tumor growth in human lung cancer cells.[10] Furthermore, CBP overexpression caused by members of the activator protein-1 (AP-1) family and down-regulation of the retinoid acid receptor β might promote lung tumor proliferation.[11] Intriguingly, CBP exerts its actions mainly by acetylation of histones and other regulatory proteins[12] and could be fused to MOZ and several other gene products in acute myeloid leukemia by chromosomal translocations.[13] In addition, mice monoallelic for the *CBP* gene may induce multilineage defects in hematopoietic differentiation with an elevated risk of hematologic malignancies.[14]

To date, the roles of SNPs in the CBP pathway genes in and their functionality related to tumor growth and progression are still unknown, and there are few reports about the SNPs in CBP pathway genes and its related gene-set in NSCLC prognosis. In the present study, by using two publicly available GWAS datasets, we performed a CBP pathway gene-set analysis to evaluate association between genetic variants in this gene-set and survival of patients with NSCLC.

2 | MATERIALS AND METHODS

2.1 | Study populations

As shown in the study flowchart (Figure 1), the discovery phase included patients with NSCLC obtained from the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, after our access approval from the National Cancer Institute (NCI). The PLCO is an NCI funded multicenter randomized trial of screening

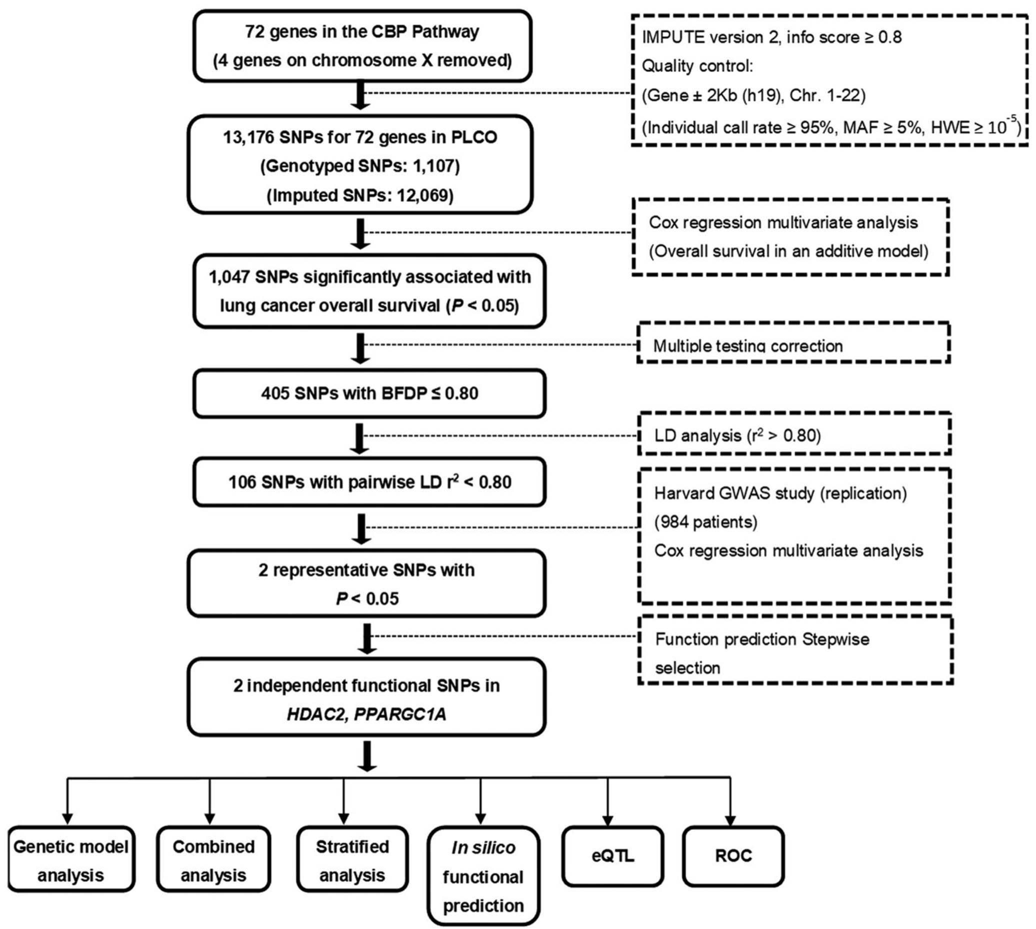

FIGURE 1 Study workflow chart. BFDP, Bayesian false-discovery probability; eQTL, expression quantitative trait loci; Harvard Study, Harvard Lung Cancer Susceptibility Study; PLCO, Prostate, Lung, Colorectal, and Ovarian cancer trial; ROC, receiver operating characteristic; SNP, single-nucleotide polymorphism

for cancer from ten medical centers in the United States between 1993 and 2011.[15] The screening trial enrolled 77 500 men and 77 500 women aged 55 to 74. All individuals were randomized to either the intervention arm with screening or the control arm with standard care.[15] The PLCO trial collected blood specimens from the first screening visit and gathered extensive information about each individual, including demigraphic information, smoking history and family history of cancer.[16] All participants were followed for at least 13 years after the enrollment. Genomic DNA extracted from the blood samples was genotyped in a GWAS with Illumina HumanHap240Sv1.0, Human-Hap300v1.1, and HumanHap550v3.0 (dbGaP accession: phs000093.v2.p2 and phs000336.v1.p1).[17,18] In 1187 Caucasian patients with NSCLC from the PLCO trial, two with missing follow-up information were excluded. Therefore, the eligible subsets of the PLCO lung cancer dataset for survival analysis included 1185 patients with NSCLC, whose clinicopathological variables and genotype data were available. Tumor staging was determined according to the fifth edition American Joint Committee on Cancer staging system. The institutional review boards of each participating institution approved the PLCO trial and the use of biological specimens for further research, and all subjects signed a written informed consent permitting the research represented here.[15]

The validation phase used the genotyping dataset from the Harvard Lung Cancer Susceptibility (HLCS) Study with 984 histology-confirmed Caucasian patients with NSCLC. The histological classification of the tumors were recorded by two staff pulmonary pathologists at the Massachusetts General Hospital. The time of blood collection was within one to four weeks of the diagnosis for each patient. DNA was extracted from blood samples by using the Auto Pure Large Sample Nucleic Acid Purification System (QIAGEN Company, Venlo, Limburg, Netherlands). Genotype data were obtained by using Illumina Humanhap610-Quad arrays, and imputation was performed by using MaCH based on the 1000 Genomes project.[19] Details of the participants in the HLCS study were described previously.[20] The comparison of the characteristics between the PLCO trial and the HLCS study is presented in Table S1.

2.2 | Gene and SNP selection

The genes or gene-set involved in the CBP pathway were selected through the Molecular Signatures Database (http://software.

broadinstitute.org/gsea/msigdb/index.jsp), by the keyword "CBP." After removal of duplicated genes and excluding genes in the X chromosome, 72 genes remained as candidate genes for further analysis (Table S2). We first performed imputation for the 72 genes plus the 500-kb flanking buffer regions by using IMPUTE2 and the 1000 Genomes Project data (phase 3).[19] After imputation, we extracted all the SNPs in these genes and within their ± 2 kb flanking regions according to the following criteria: a minor allele frequency ≥ 0.05, a genotyping rate ≥ 95%, and Hardy-Weinberg equilibrium $P \geq 1 \times 10^{-5}$. As a result, 1107 genotyped SNPs were chosen from the PLCO GWAS dataset with an additional 12 069 SNPs as a result of imputation.

2.3 | Statistical analyses

The follow-up time in both PLCO and HLCS datasets was from the diagnosis of lung cancer to the last follow-up or time of death. Overall survival (OS) was the primary endpoint of the present study, and disease-specific survival (DSS) of lung cancer was also examined. In the single-locus analysis, multivariate Cox proportional hazards regression analysis was used to evaluate associations between each of the SNPs and survival (in an additive genetic model) with adjustment for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery, and the top four principal components of the PLCO dataset using the GenABEL package of R software.[21] We first used the false discovery rate (FDR) to correct multiple testing with a threshold of 0.2. Since the majority of SNPs were imputed and there was a high level of linkage disequilibrium (LD) among the SNPs, we used the Bayesian false discovery probability (BFDP) with a cutoff value of 0.8 for multiple test corrections as recommended.[22] We assigned a prior probability of 0.10 to detect an hazards ratio (HR) of 2.0 for an association with variant genotypes or minor alleles of the SNPs with $P < .05$. Then, we chose SNPs for the HLCS GWAS dataset validation, which satisfied the following conditions: SNPs passed the threshold of BFDP ≤ 0.8 and tagging SNPs based on their LD. To identify independent SNPs, we included the validated SNPs in a multivariate stepwise Cox model with adjustment for demographic characteristics, previously published SNPs, clinical variables and the top four principal components of the genotyping data in the PLCO dataset. Combined analysis of discovery and validation datasets was also performed to provide summary results. A fixed-effects model was applied if the Cochran's Q-test $P > .100$ and the heterogeneity statistic (I^2) < 50%; otherwise, a random-effects model was employed. Kaplan-Meier curves were used to estimate the survival associated with genotypes. The combination of risk or unfavorable genotypes was used to estimate the cumulative effects of the identified SNPs.

Functions of the validated SNPs were further predicted by HaploReg,[23] SNPinfo,[24] and RegulomeDB.[25] The criteria for functional SNPs were: (a) the SNPs were associated with survival in both the PLCO trial and HLCS study (both $P < .05$) and (b) genotypes of the SNPs were associated with messenger RNA (mRNA) expression of their genes as shown in one of these databases. Expression quantitative trait loci (eQTL) analysis was further performed to assess correlations between SNPs and mRNA expression levels by using linear regression analysis with the R (version 3.5.0) software. mRNA expression data of genes were obtained from lymphoblastoid cell lines derived from the 373 European descendants included in the 1000 Genomes Project[19] and from the whole blood and normal lung tissues in the genotype-tissue expression (GTEx) project.[26] Using the data from The Cancer Genome Atlas (TCGA) database (dbGaP Study Accession: phs000178.v9.p8), we examined the differences in mRNA expression levels between paired tumor and adjacent normal tissues by paired *t* test.[27] Next, we detected the association of mRNA expression levels and OS through Kaplan-Meier (KM) analysis (n = 1926) (http://kmplot.com/analysis/index.php?p=service&cancer=lung). Finally, receiver operating characteristic (ROC) curves were constructed and time-dependent ROC analysis were performed to examine the prediction accuracy of models integrating both clinical and genetic variables on NSCLC survival with the "timeROC" package in R (version 3.5.0).[28] Unless specified, all statistical analyses were performed using the SAS software (version 9.4; SAS Institute, Cary, NC).

3 | RESULTS

3.1 | Associations between SNPs in the CBP pathway gene-set and NSCLC survival in both PLCO and HLCS data setsr

The study flowchart is shown in Figure 1, and basic characteristics of 1185 patients with NSCLC have been described previously[29] (also see Table S3). In the PLCO discovery dataset with an additive genetic model, the multivariate Cox model with adjustment for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery, and the first four principal components (Table S4) identified 405 SNPs that were significantly associated with NSCLC survival after multiple test correction by BFDP ≤ 0.8 but not by false discovery rate ≤ 0.2 – the results are summarized in a Manhattan plot (Figure 2A). Potentially functional SNPs were further validated by the HLCS dataset. As a result, two SNPs in two different genes (i.e., rs13213007 in *HDAC2* and rs60571065 in *PPARGC1A*) were validated (Table 1), both of which were imputed. Further combined analysis of these SNPs of the two datasets showed a worse survival associated with the rs13213007 A and rs60571065 A alleles ($P_{adjusted}$ = .0003 and .002, respectively), and no heterogeneity between the two studies were observed (Table 1).

3.2 | Identification of independent SNPs associated with survival of NSCLC in the PLCO dataset

Because the HLCS study provided only the summary data, we used only the PLCO dataset to identify independent SNPs after full adjustment for other available covariables. To identify potential functional SNPs associated with NSCLC survival with three online bioinformatics tools (ie, RegulomeDB,[25] SNPinfo,[24] and HaploReg[23]), we found that both of the validated SNPs were located in the intron

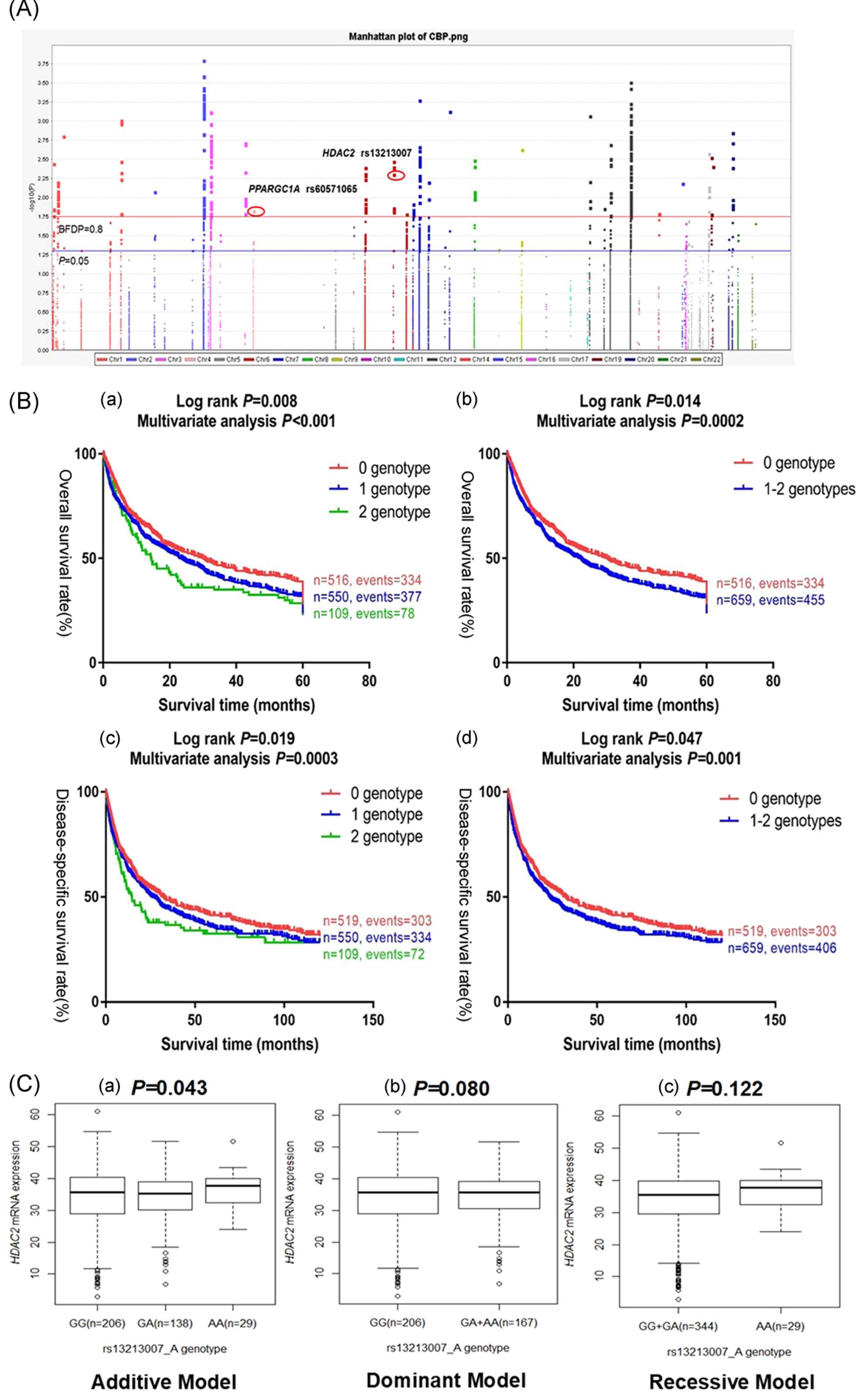

FIGURE 2 Continued.

TABLE 1 Combined analysis of the two unfavorable SNPs in PLCO and HLCS datasets

			PLCO (n = 1185)				HLCS (n = 984)			Combined analysis			
SNP	Allele[a]	Gene	EAF	HR (95% CI)[b]	P^b	BFDP	EAF	HR (95% CI)[c]	P^c	HR (95% CI)[d]	P^d	P_{het}^e	I^2
rs13213007	G>A	*HDAC2*	0.26	1.18 (1.05-1.32)	.004	0.58	0.29	1.14 (1.02-1.29)	.022	1.16 (1.07-1.26)	<.001	.758	0
rs60571065	T>A	*PPARGC1A*	0.11	1.22 (1.04-1.43)	.015	0.77	0.10	1.21 (1.00-1.46)	.050	1.22 (1.08-1.37)	.002	.948	0

Abbreviations: BFDP, Bayesian false-discovery probability; EAF, effect allele frequency; HLCS, Harvard Lung Cancer Susceptibility Study; HR, hazards ratio; PLCO, Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; SNPs, single-nucleotide polymorphisms.
[a]Major>minor allele.
[b]Adjusted for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3 and PC4.
[c]Adjusted for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2 and PC3.
[d]In fixed-effect model.
[e]*P* value for heterogeneity test by Cochrane's Q test.

regions with some functions. In the RegulomeDB,[25] rs60571065 had a score of 5, indicating a potential transcriptional factor binding or DNase peak, while rs13213007 had no data (Table S5); both of the validated SNPs had no obvious function based on SNPinfo[24] or HaploReg.[23]

The two validated SNPs (i.e., *HDAC2* rs13213007 and *PPARGC1A* rs60571065) were also included into a multivariate stepwise Cox model with adjustments for demographic and clinical variables as well as previously published SNPs and the first four principal components available in the PLCO dataset. As a result, both of the two SNPs remained independently associated with NSCLC survival (Table 2, also see Table S6); a regional association plot of each SNP is shown in Figure S1. In the PLCO dataset, patients with rs13213007 A or rs60571065 A alleles had an increased risk of death (P_{trend} = .004 or .015, Table 3). Compared with the reference genotype in a dominant genetic model, *HDAC2* rs13213007 GA+AA and *PPARGC1A* rs60571065 TA+AA were associated with a significantly increased risk of death (HR = 1.26, 95% CI, 1.09-1.45 and *P* = .002 for rs13213007 GA+AA; and HR = 1.23, 95% CI, 1.04-1.47 and *P* = .017 for rs60571065 TA+AA).

3.3 | Combined and stratified analysis of the two independent and functional SNPs in the PLCO dataset

To provide a better estimation of the hazards for survival, we combined the risk genotypes (i.e., rs13213007 GA+AA and rs60571065 TA+AA) into a genetic score, which was used to divide all patients with NSCLC into three groups. As shown in Table 3, in the multivariate analysis, an increased genetic score was associated with a worse survival (trend test: $P < .001$). After dichotomizing the genetic score, we regrouped all the patients into a low-risk group (0 risk score) and a high-risk group (1-2 risk scores); patients in the high-risk group had a significantly poorer survival compared with the low-risk group (HR = 1.31, 95% CI, 1.14-1.52, and *P* = .0002). Kaplan-Meier survival curves were generated to depict the associations between risk genotypes and NSCLC OS (Figure 2B(a) and 2B(b)) and DSS (Figure 2B(c) and 2B(d)).

To assess the ability of risk genotypes to predict outcomes, we compared the areas under the ROC curves (area under the curve [AUC]) from the model with clinical variables to that from the model including both clinical variables and risk genotypes. The addition of risk genotypes to the prediction model of five-year OS increased the AUC from 87.0% to 87.27% (*P* = .375; Figure S2a); similarly, the addition of risk genotypes to the prediction model of five-year DSS increased the AUC from 86.7% to 87.01% (*P* = .391; Figure S2b). The add-on of risk genotypes and previously published SNPs to the prediction model of five-year OS increase the AUC from 88.03% to 89.06% (*P* = .066; Figure S2e); furthermore, the prediction model of five-year DSS also increased the AUC from 88.2% to 89.4% (*P* = .030; Figure S2f). Finally, the time-dependent AUC curve was generated to quantify the ability of risk genotypes to predict NSCLC survival through the entire follow-up period (Figure S2c, S2d, S2g, and S2h).

We further analyzed the effect of combined risk genotypes on NSCLC OS and whether it was modified by age, sex, smoking

FIGURE 2 Functional and survival-associated SNPs. A, Manhattan plot of 13 176 SNPs of the CBP Pathway genes in the PLCO dataset. The statistical values across the autosomes for associations between 13 176 SNPs and overall survival of patients with NSCLC are plotted as −log10 *P* values. The blue horizontal line indicates *P* = .05 and the red line indicates BFDP = 0.80. B, Kaplan-Meier (KM) survival curves for patients with NSCLC of two validated SNPs and their combined unfavorable genotypes in the PLCO dataset. (a) By 0, 1, and 2 unfavorable genotypes (log-rank test for trend and multivariate analysis: *P*) in OS, (b) by 0 and 1-2 unfavorable genotypes (log-rank test and multivariate analysis: *P*) in OS from the PLCO dataset, (c) by 0, 1, and 2 unfavorable genotypes (log-rank test for trend and multivariate analysis: *P*) in DSS, and (d) by 0 and 1-2 unfavorable genotypes (log-rank test and multivariate analysis: *P*) in DSS from the PLCO dataset. C, eQTL analysis of *HDAC2* rs13213007 and *PPARGC1A* rs60571065 genotypes and corresponding gene mRNA expression. All the data were from the 1000 Genomes Project dataset. (a) rs13213007, additive model (*P* = .043); (b) rs13213007, dominant model (*P* = .080); and (c) rs13213007, recessive model (*P* = .122) but no significant results for rs60571065 (*P* = .556, .743, and 0.339, figures not shown). BFDP, Bayesian false-discovery probability; CBP, CREB-binding protein; DSS, disease-specific survival; eQTL, expression quantitative trait loci; NSCLC, non-small-cell lung cancer; OS, overall survival; PLCO, Prostate, Lung, Colorectal, and Ovarian cancer screening trial; SNPs, single-nucleotide polymorphisms [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Predictors of OS obtained from stepwise Cox regression analysis in the PLCO data set

Variables[a]	Category	Frequency[b]	HR (95% CI)	*P*
Age	Continuous	1174	1.04 (1.02-1.05)	<.001
Sex	Male	694	1.00	
	Female	480	0.82 (0.70-0.95)	.010
Smoking status	Never	114	1.00	
	Current	416	1.78 (1.32-2.39)	<.001
	Former	644	1.78 (1.35-2.35)	<.001
Histology	Adenocarcinoma	575	1.00	
	Squamous cell	283	1.29 (1.07-1.56)	.008
	others	316	1.34 (1.13-1.59)	<.001
Tumor stage	I-IIIA	654	1.00	
	IIIB-IV	520	2.97 (2.45-3.61)	<.001
Chemotherapy	No	638	1.00	
	Yes	536	0.56 (0.46-0.66)	<.001
Radiotherapy	No	760	1.00	
	Yes	414	0.96 (0.86-1.13)	.617
Surgery	No	634	1.00	
	Yes	540	0.19 (0.15-0.25)	<.001
***HDAC2 rs*13213007 G>A**[c]	**GG/GA/AA**	**648/450/76**	**1.16 (1.03-1.30)**	**.012**
***PPARGC1A rs*60571065 T>A**[c]	**TT/TA/AA**	**925/231/18**	**1.21 (1.04-1.42)**	**.017**

Abbreviations: CI, confidence interval; HR, hazards ratio; PLCO, Prostate, Lung, Colorectal, and Ovarian cancer trial; OS, overall survival; SNPs, single-nucleotide polymorphisms.

[a]Stepwise analysis included age, sex, smoking status, tumor stage, tumor histology, chemotherapy, radiotherapy, surgery, top four principal components, and eight published SNPs and two new validated SNPs (*HDAC2* rs13213007 and *PPARGC1A rs*60571065 in an additive model).

[b]Ten observations with missing clinical variables (two of tumor stage and eight of chemotherapy/radiotherapy/surgery); one missing for *TLE1* rs199731120 C>CA (a published SNP); 1174 patients remained for the stepwise analysis.

[c]Two SNPs identified in the present study.

status, histology, tumor stage, chemotherapy, radiotherapy and surgery (Table 4). The results showed no significant interactions ($P > .05$) except for smoking status ($P = .002$). Hence, as shown in Tables S7 and S8, we further assessed the associations of *HDAC2* rs13213007 and *PPARGC1A* rs60571065 with OS of NSCLC in three subgroups of never, current, and former smokers and found that cigarette smoking and *HDAC2* rs13213007 had an interactive effect on NSCLC OS, where the highest risk was observed amongst current smokers. In addition, never smokers were at a higher risk than former smokers.

3.4 | In silico functional validation

Experimental data from the ENCODE project[25] (Figure S3a-b) revealed that the two SNPs (i.e., *HDAC2* rs13213007 and *PPARGC1A* rs60571065) are located in a DNase I hypersensitive site, where the DNase hypersensitivity and histone modification H3K4 acetylation indicated some strong signals for active enhancer and promoter functions. The evidence from the DNase cluster and transcription factor CHIP-seq data[25] predicted that rs60571065 is located in the Barhl2/Msx-3/Sox1 motif as shown by the position weight matrix (Figure S3c-e) and that the minor allele may affect the binding activity to have an impact on the transcription factors.

In addition, we performed the eQTL analysis to correlate genotypes of the SNPs and mRNA expression levels using data of the 373 European descendants available from the 1000 Genomes Project.[19] Only the *HDAC2* rs13213007 A allele (GA+AA) showed a significant correlation with increased mRNA expression levels of the gene ($P = 0.043$; Figure 2C(a)), while this was not the case for the *PPARGC1A* rs60571065 A allele; however, in the whole blood and lung tissue data of the GTEx project, both SNPs were not associated with expression levels of *HDAC2* and *PPARGC1A* (Table S9). Taken together, these findings suggest that *HDAC2* rs13213007A allele, but not *PPARGC1A* rs60571065A allele, may influence *HDAC2* gene expression at the transcription level.

Finally, to assess the expression of these genes in NSCLC, we compared the data on mRNA expression levels of these genes in 111 paired NSCLC tumor and adjacent normal tissues obtained from the TCGA database. Expression levels of *HDAC2* were higher in the tumor tissues ($P < .001$), compared with the adjacent normal tissues (Figure S4a-c), and the higher expression levels were associated with a poorer NSCLC OS (Figure S5a). On the other hand, the expression levels of *PPARGC1A* were lower in the tumor tissues ($P < .001$), compared with the adjacent normal tissues (Figure S4d-f), and the higher expression levels were associated with a better NSCLC OS[30] (Figure S5b).

TABLE 3 Associations of the two validated SNPs in the *CBP* pathway with OS of NSCLC in the PLCO dataset

	OS univariate analysis				OS multivariable analysis[a]			
Genotype	Frequency (n = 1185)	Death (%)	HR (95% CI)	P	Frequency (n = 1175)	Death (%)	HR (95% CI)	P
***HDAC2* rs13213007 G>A**								
GG	656	431 (65.7)	1.00		648	424 (65.4)	1.00	
GA	453	311 (68.7)	1.19 (1.03-1.38)	.018	451	309 (68.5)	1.26 (1.09-1.47)	.002
AA	76	56 (73.7)	1.28 (0.97-1.70)	.079	76	56 (73.7)	1.24 (0.93-1.64)	.145
Trend				.008				.004
GA+AA	529	367 (69.4)	1.21 (1.05-1.39)	.008	527	365 (69.3)	1.26 (1.09-1.45)	.002
***PPARGC1A rs*60571065 T>A**								
TT	942	629 (66.8)	1.00		934	621 (66.5)	1.00	
TA	225	159 (70.7)	1.11 (0.93-1.32)	.235	223	158 (70.9)	1.22 (1.03-1.46)	.025
AA	18	10 (55.6)	0.78 (0.42-1.46)	.441	18	10 (55.6)	1.42 (0.76-2.67)	.276
Trend				.555				.015
TA+AA	243	169 (69.6)	1.08 (0.91-1.29)	.351	241	168 (69.7)	1.23 (1.04-1.47)	.017
NUG[b]								
0	522	340	1.00		516	334	1.00	
1	554	380 (68.6)	1.17 (1.01-1.35)	.039	550	377 (68.6)	1.28 (1.10-1.48)	.002
2	109	78 (71.6)	1.32 (1.03-1.69)	.027	109	78 (71.6)	1.50 (1.17-1.93)	.001
Trend				.008				<.001
0	522	340 (65.1)	1.00		516	334 (64.7)	1.00	
1–2	663	458 (69.1)	1.19 (1.03-1.37)	.015	659	455 (69.0)	1.31 (1.14-1.52)	<.001
Phenotype missing[c]	N/A				10			

Abbreviations: CI, confidence interval; HR, hazards ratio; NSCLC, non-small-cell lung cancer; NUG, number of unfavorable genotypes; OS, overall survival; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Trial; SNPs, single-nucleotide polymorphisms.

[a]Multivariate Cox hazards regression analyses were adjusted for age, sex, smoking, stage, histology, chemotherapy, radiotherapy, surgery and top four principal components.

[b]NUG were *HDAC2* rs11606640 GA+AA and *PPARGC1A* rs62053220 TA+AA.

[c]Two observations with missing tumor stage and eight observations with missing chemotherapy/radiotherapy/surgery in the PLCO dataset.

3.5 | Mutation analysis

We also assessed the mutation status of *HDAC2* and *PPARGC1A* in lung tumor tissues by using the public database of the cBioPortal for Cancer Genomics. As shown in Figure S6a-b, *HDAC2* had a low somatic mutation rate in NSCLC (1.05%, 12/1144) in the 2016 TCGA study,[31] LUAD (1.64%, 3/183% and 1.24%, 7/566) in the Broad[32] and TCGA PanCan studies,[27] respectively; and LUSC (0.41%, 2/487) in the TCGA PanCan study. In contrast, *PPARGC1A* had a relative higher somatic mutation rate in NSCLC (3.06%, 35/1144) in the TCGA 2016 study,[31] LUAD (3.28%, 6/183; 3.89%, 22/566; and 3.89%, 22/566) in the Broad[32] and TCGA PanCan studies,[27] respectively; and LUSC (3.37%, 6/178; 2.26%, 11/487; and 0.98%, 5/511) in the TCGA pub, TCGA PanCan, and TCGA[27] studies, respectively. Given the rarity of mutations in these two genes, our results suggest that the functional SNPs in *HDAC2* may play a relatively important role in the dysregulation of mRNA expression in tumor tissues, whereas the mutations might more likely play a role in altered functions and expression of *PPARGC1A*.

4 | DISCUSSION

In the present study, we found two genetic variants (i.e., *HDAC2* rs13213007 G>A and *PPARGC1A* rs60571065 T>A) in the CBP pathway gene-set to be significantly associated with survival of patients with NSCLC in both PLCO and HLCS GWAS datasets. The results suggest that patients with risk genotypes (i.e., *HDAC2* rs13213007 GA+AA or *PPARGC1A* rs60571065 TA+AA) had a worse prognosis. Furthermore, *HDAC2* rs13213007 G>A appeared to have an effect on *HDAC2* mRNA expression, which makes this SNP-associated risk of death biologically plausible.

Histone deacetylases (HDACs) function to deacetylate the ε-amino group of lysine residues of both histone and nonhistone substrates.[7] HDACs are grouped into class I–IV enzymes,[33] that is, Class I contains HDAC1, 2, 3 and 8, class II HDAC4, 5, 6, 7, 9 and 10, and class IV HDAC, whereas Class III enzymes contain the *SIRT* deacetylases.[34] Studies suggest that *HDAC2* is crucial for embryonic development, growth regulation and cytokine signaling relevant for immune responses,[35] often overexpressed in lung cancer and other

TABLE 4 Stratified analyses for association between unfavorable genotypes and OS in patients with NSCLC in the PLCO dataset

	No. of 0 risk genotypes		No. of 1-2 risk genotypes		Multivariate analysis[a]		
Characteristics	All	Death (%)	All	Death (%)	HR (95% CI)	P	P_{inter}[b]
Age (y)							
≤ 71	277	168 (60.7)	357	251 (64.7)	1.31 (1.07-1.61)	.009	
> 71	239	166 (69.5)	302	224 (74.2)	1.35 (1.10-1.65)	.005	.697
Sex							
Male	273	190 (69.6)	422	315 (74.6)	1.33 (1.11-1.60)	.002	
Female	243	144 (59.3)	237	140 (59.1)	1.32 (1.03-1.69)	.029	.906
Smoking status							
Never	57	28 (49.1)	57	34 (59.7)	1.44 (0.81-2.57)	.212	
Current	182	110 (60.4)	235	156 (66.4)	1.72 (1.33-2.23)	<.001	
Former	277	196 (70.8)	367	265 (72.2)	1.08 (0.90-1.31)	.409	.002*
Histology							
Adenocarcinoma	259	149 (57.5)	316	197 (62.3)	1.36 (1.09-1.70)	.007	
Squamous	129	87 (67.4)	155	104 (67.1)	1.32 (0.98-1.79)	.069	
Others	128	98 (76.6)	188	154 (81.9)	1.43 (1.10-1.87)	.009	.780
Tumor stage							
I-IIB	287	133 (46.3)	367	181 (49.3)	1.23 (0.98-1.55)	.077	
IIIA-IV	229	201 (87.8)	292	274 (93.8)	1.36 (1.12-1.64)	.002	.524
Chemotherapy							
No	272	147 (54.0)	366	219 (59.8)	1.33 (1.07-1.65)	.009	
Yes	244	187 (76.6)	293	236 (80.6)	1.22 (1.00-1.50)	.049	.847
Radiotherapy							
No	323	179 (55.4)	438	270 (61.6)	1.35 (1.11-1.64)	.002	
Yes	193	155 (80.3)	221	185 (83.7)	1.25 (1.00-1.56)	.055	.397
Surgery							
No	283	242 (85.5)	352	323 (91.8)	1.34 (1.13-1.59)	<.001	
Yes	233	92 (39.5)	307	132 (43.0)	1.33 (1.01-1.76)	.041	.771
Phenotype missing[c]	10						

Abbreviations: CI, confidence interval; HR, hazards ratio; NSCLC, non-small-cell lung cancer; OS, overall survival; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Trial.

[a]Adjusted for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3 and PC4.

[b]P_{inter}: *P*-value for interaction analysis between characteristic and number of unfavorable genotypes.

[c]Two observations with missing tumor stage and eight observations with missing chemotherapy/radiotherapy/surgery in the PLCO dataset.

tumors.[36] Both published in vitro and in vivo experiments have indicated that the aberrant regulation of *HDAC2* may confer an oncogenic potential for lung cancer cells and hepatocellular carcinoma cells by deregulating apoptosis and expression of cell cycle proteins,[37] while overexpression of *HDAC2* has been found to be associated with a shorter relapse-free survival after radical prostatectomy in prostate cancer.[38] Furthermore, the upregulation of *HDAC2* may enhance proliferation of gastric cancer cells by deregulating cell cycle proteins,[39] and *HDAC2* may also play a fundamental role in the Myc-mediated oncogenesis.[40] These are consistent with our findings that *HDAC2* expression was increased in lung cancer tissues, compared with normal tissues in the TCGA dataset. However, the detailed mechanisms of *HDAC2* SNPs underlying the observed death-risk associations warrant further investigations.

Few studies have reported the role of *HDAC2* SNPs in cancer outcomes. One study found that the rs11391 SNP in the 3′-untranslated region of *HDAC2* was a potential genetic marker of poor prognosis for NSCLC,[41] while another study identified several functional SNPs in *HDAC2* and evaluated their associations with clinical outcomes in hepatocellular carcinoma, but no significant associations were found.[42] In the present study, we showed that the *HDAC2* rs13213007 A allele was associated with a poorer survival of NSCLC, likely due to its association with an increase in mRNA expression levels of the gene; moreover, *HDAC2* mRNA expression levels were found to be higher in tumor tissues, and higher expression levels were also associated with a worse survival in NSCLC. According to the ENCODE project database,[25] *HDAC2* rs13213007 is located in a DNase I hypersensitive site with considerable levels of histone modification H3K4 acetylation, which may lead to an enhanced transcriptional activity. Therefore, it is likely that that *HDAC2* might act as an oncogene in NSCLC.

Similarly, there are also few studies that have reported the role of *PPARGC1A* SNPs in cancer outcomes. Previous studies have reported that abnormalities in cell metabolism are associated with tumorigenesis where mitochondria are key regulators.[43] *PPARGC1A* is a known regulator of mitochondrial biogenesis and is a multiple-function transcriptional coactivator that has been noted to be associated with

various human diseases, including type II diabetes mellitus[44] and coronary disease.[45] One study utilized ChIP-seq to obtain *PPARGC1A* binding sites across the genome in hepatoma cells HepG2, and the subsequent conserved-motif analysis showed that the majority of the *PPARGC1A* binding sites were located in multiple regulatory factor binding regions.[46] To date, however, the studies on the expression and function as well as oncogenic roles of *PPARGC1A* in lung cancer are inconsistent. For example, one report found that *PPARGC1A* might be an oncogene,[47] while the other demonstrated that it might be a suppressor gene.[48]

In the TCGA database, we found that mRNA expression levels of *PPARGC1A* were higher in normal lung tissues than in tumor tissues (n = 111), while its lower expression levels were associated with a worse survival in NSCLC. However, *PPARGC1A* was also reported to act as an oncogene, because its expression levels were upregulated in tumor tissues (n = 3), facilitating lung cancer metastasis.[49] It is possible that the expression levels of *PPARGC1A* in lung cancer tissues may be affected by other factors, such as an imbalanced activation by risk genotypes, which may cause abnormal expression of *PPARGC1A*, in addition to mutations. Based on the ENCODE Project data, *PPARGC1A* rs60571065 is located in a DNase I hypersensitive site, indicating its potential function to modify gene expression. Clearly, additional studies are warranted to investigate possible mechanisms underlying *PPARGC1A* rs60571065-associated death risk of NSCLC.

In the clinical setting for patients with NSCLC having *EGFR* mutations, treatment with EGFR-tyrosine kinase inhibitors (EGFR-TKIs) has resulted in an increased survival rate in comparison with the standard chemotherapy.[50] Because EGFR-TKIs have been used as the first-line treatment in the application of concurrent or sequential chemo-radiotherapy regimens, different treatment options should be incorporated into the phenotype analysis.[51] Recently, the 3rd-generation EGFR-TKIs (EGFR-TKI3rd) have displayed remarkable efficacy against the T790M-related resistance in mutation-positive NSCLC.[52] Despite dramatic initial responses, a genetically driven resistance to the targeted therapy eventually emerged; thus, limiting their effectiveness in prolonged treatments.[52] Some reports suggest that it is due in part to differences in co-occurring genetic alterations, such as mutations of *TP53* or *PTEN*,[53] but few reports have addressed the roles of SNPs in other oncogenes. In the present study, we found that risk genotypes of *HDAC2* rs13213007 enhanced mRNA expression levels, suggesting that the risk genotypes may have an impact on T790M-related resistance in NSCLC, which, once validated, may provide a clue for individualized therapy.[54]

While tobacco smoking is a major risk factor for lung cancer, smoking behavior could enhance the adverse effect of the variants by increasing c-Jun expression in lung cancer.[55] In the present study, we found that smoking might increase the risk of death associated with *HDAC2* rs13213007 SNP in patients with NSCLC. It is likely that smoking could affect the function of epigenetics or the H3K4 acetylation activity associated with *HDAC2*, a similar phenomenon seen on other cancers, including cancers of the pancreas, stomach, bladder, head and neck and the colorectum.[56]

Although we observed associations between genetic variants in two CBP pathway genes and NSCLS survival along with some functional evidence, the exact molecular mechanisms are still unclear. Further functional experiments are required to corroborate the present findings. In addition, since both the discovery and validation datasets used in the present study were from Caucasian populations, the results may not be generalizable to other ethnic populations. Although the sample size of PLCO was relatively large, the number of patients in each subgroup were still relatively small, which might have reduced the statistical power to detect a small effect in one particular subpopulation. Furthermore, since only a few clinical factors were available for additional analysis, other covariates, such as performance status and treatments, were not available for further adjustments. In addition, it is worth noting that the detailed genotype and phenotype data from the HLCS study were not accessible for us to perform additional stratified analysis in a larger sample size.

In conclusion, two independent functional SNPs (i.e., *HDAC2* rs13213007 G>A and *PPARGC1A* rs60571065 T>A) were found to be significantly associated with NSCLC survival in both PLCO and HLCS GWAS datasets. The combined analysis revealed that the two SNPs had a significant association with survival and that patients with more risk genotypes had a worse prognosis, possibly due to the progression acceleration and metastatic effects of overexpressed *HDAC2*. Knowing the characteristic overexpression of *HDAC2* in cancer, the discovery of the novel genetic variant of *HDAC2* rs13213007 may serve to showcase additional insights for potential therapeutic capabilities of HDAC inhibitors and in cancer treatment, once validated by additional studies. The findings in the present study also provide a solid foundation for further functional studies to identify molecular mechanisms underlying the observed death-risk associations in the progression of NSCLC.

ACKNOWLEDGMENTS

Carolyn Glass, Yu Chen Zhao, Danwen Qian, Hongliang Liu, and Sipeng Shen participated in the data collection and analysis. Dongfang Tang, Sheng Luo, David C. Christiani, Wen Gao, and Qingyi Wei participated in the design of the study. Dongfang Tang, Sheng Luo, Edward F. Patz, Patricia G. Moorman, Li Su, David C. Christiani, Wen Gao, and Qingyi Wei participated in the writing of the manuscript and the interpretation of the data. All authors read and approved the final manuscript. The authors thank all the participants of the PLCO Cancer Screening Trial. The authors also thank the National Cancer Institute for providing the access to the data collected by the PLCO trial. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by the National Cancer Institute. The authors would also like to acknowledge dbGaP repository for providing cancer genotyping data sets. The accession numbers for the data sets of lung cancer are phs000336.v1.p1 and phs000093.v2.p2. A list of contributing investigators and funding

agencies for those studies can be found in the Supplemental data. Qingyi Wei was supported by the V Foundation for Cancer Research (D2017-19) and partly supported by the Duke Cancer Institute as part of the P30 Cancer Center Support Grant (Grant ID: NIH/NCI CA014236). Dongfang Tang was supported by the short-term international training program for PhD from Fudan University, China. The Harvard Lung Cancer Susceptibility Study was supported by NIH grants 5U01CA209414, CA092824, CA074386, and CA090578 to David C. Christiani.

CONFILICT OF INTERESTS

The authors declare that there are no conflict of interests.

FUNDING INFORMATION

This study was supported by the National Institute of Health (CA090578, CA074386, CA092824, and 5U01CA209414); the short-term international training program for PhD training from Fudan University, China; the Duke Cancer Institute as part of the P30 Cancer Center Support Grant (NIH/NCI CA014236); the V Foundation for Cancer Research (D2017-19).

DATA AVAILABILITY STATEMENT

All data generated or analyzed during the present study are included in this article.

ORCID

Sipeng Shen http://orcid.org/0000-0003-0436-4736
Qingyi Wei http://orcid.org/0000-0002-5265-3930

REFERENCES

1. Zarredar H, Ansarin K, Baradaran B, Ahdi Khosroshahi S, Farajnia S. Potential molecular targets in the treatment of lung cancer using siRNA technology. *Cancer Invest*. 2018;36:37-58.
2. Houston KA, Mitchell KA, King J, White A, Ryan BM. Histologic lung cancer incidence rates and trends vary by race/ethnicity and residential county. *J Thorac Oncol*. 2018;13:497-509.
3. Barnfield PC, Ellis PM. Second-line treatment of non-small cell lung cancer: new developments for tumours not harbouring targetable oncogenic driver mutations. *Drugs*. 2016;76:1321-1336.
4. Li Y, Sheu CC, Ye Y, et al Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. *Lancet Oncol*. 2010;11:321-330.
5. Rosell R, Wei J. Single nucleotide polymorphisms (SNPs) in non-small cell lung cancer (NSCLC) patients. *Oncologist*. 2012;17:1484-1485.
6. Li MJ, Sham PC, Wang J. Genetic variant representation, annotation and prioritization in the post-GWAS era. *Cell Res*. 2012;22:1505-1508.
7. Scolnick DM, Chehab NH, Stavridi ES, et al CREB-binding protein and p300/CBP-associated factor are transcriptional coactivators of the p53 tumor suppressor protein. *Cancer Res*. 1997;57:3693-3696.
8. Du Y, Teng X, Wang N, et al NF-kappaB and enhancer-binding CREB protein scaffolded by CREB-binding protein (CBP)/p300 proteins regulate CD59 protein expression to protect cells from complement attack. *J Biol Chem*. 2014;289:2711-2724.
9. Lin Z, Feng R, Li J, et al Nuclear translocation of CBP/p300-interacting protein CITED1 induced by parathyroid hormone requires serine phosphorylation at position 79 in its 63-84 domain. *Cell Signal*. 2014;26:2436-2445.
10. Guo W, Lu J, Dai M, et al Transcriptional coactivator CBP upregulates hTERT expression and tumor growth and predicts poor prognosis in human lung cancers. *Oncotarget*. 2014;5:9349-9361.
11. Hsu FF, Chiang MT, Li FA, Yeh CT, Lee WH, Chau LY. Acetylation is essential for nuclear heme oxygenase-1-enhanced tumor growth and invasiveness. *Oncogene*. 2017;36:6805-6814.
12. Levine AA, Guan Z, Barco A, Xu S, Kandel ER, Schwartz JH. CREB-binding protein controls response to cocaine by acetylating histones at the fosB promoter in the mouse striatum. *Proc Natl Acad Sci U S A*. 2005;102:19186-19191.
13. Serravalle S, Melchionda F, Astolfi A, Libri V, Masetti R, Pession A. A novel specific signature of pediatric MOZ-CBP acute myeloid leukemia. *Leuk Res*. 2010;34:e292-e293.
14. Kung AL, Rebel VI, Bronson RT, et al Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP. *Genes & development*. 2000;14:272-277.
15. Weissfeld JL, Schoen RE, Pinsky PF, et al Flexible sigmoidoscopy in the randomized prostate, lung, colorectal, and ovarian (PLCO) cancer screening trial: added yield from a second screening examination. *J Natl Cancer Inst*. 2012;104:280-289.
16. Oken MM, Marcus PM, Hu P, et al Baseline chest radiograph for lung cancer detection in the randomized Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. *J Natl Cancer Inst*. 2005;97:1832-1839.
17. Tryka KA, Hao L, Sturcke A, et al NCBI's database of genotypes and phenotypes: dbGaP. *Nucleic Acids Res*. 2014;42:D975-D979.
18. Mailman MD, Feolo M, Jin Y, et al The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*. 2007;39:1181-1186.
19. Lappalainen T, Sammeth M, Friedlander MR, et al Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013;501:506-511.
20. Zhai R, Yu X, Wei Y, Su L, Christiani DC. Smoking and smoking cessation in relation to the development of co-existing non-small cell lung cancer with chronic obstructive pulmonary disease. *Int J Cancer*. 2014;134:961-970.
21. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 2007;23:1294-1296.
22. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst*. 2004;96:434-442.
23. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. 2012;40:D930-D934.
24. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res*. 2009;37:W600-W605.
25. Boyle AP, Hong EL, Hariharan M, et al Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22:1790-1797.
26. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348:648-660.
27. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511:543-550.
28. Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Stat Med*. 2006;25:3474-3486.

29. Wang Y, Liu H, Ready NE, et al Genetic variants in ABCG1 are associated with survival of nonsmall-cell lung cancer patients. *Int J Cancer*. 2016;138:2592-2601.
30. Gyorffy B, Surowiak P, Budczies J, Lanczky A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLOS One*. 2013;8:e82241.
31. Campbell JD, Alexandrov A, Kim J, et al Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet*. 2016;48:607-616.
32. Imielinski M, Berger AH, Hammerman PS, et al Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. 2012;150:1107-1120.
33. Neugebauer RC, Sippl W, Jung M. Inhibitors of NAD+ dependent histone deacetylases (sirtuins). *Curr Pharm Des*. 2008;14:562-573.
34. Peserico A, Simone C. Physical and functional HAT/HDAC interplay regulates protein acetylation balance. *J Biomed Biotechnol*. 2011;2011:371832.
35. Krämer OH. HDAC2: a critical factor in health and disease. *Trends Pharmacol Sci*. 2009;30:647-655.
36. Noh JH, Jung KH, Kim JK, et al. Aberrant regulation of HDAC2 mediates proliferation of hepatocellular carcinoma cells by deregulating expression of G1/S cell cycle proteins. *PLOS One*. 2011;6:e28103.
37. Jung KH, Noh JH, Kim JK, et al HDAC2 overexpression confers oncogenic potential to human lung cancer cells by deregulating expression of apoptosis and cell cycle proteins. *J Cell Biochem*. 2012;113:2167-2177.
38. Weichert W, Roske A, Gekeler V, et al Histone deacetylases 1, 2 and 3 are highly expressed in prostate cancer and HDAC2 expression is associated with shorter PSA relapse time after radical prostatectomy. *Br J Cancer*. 2008;98:604-610.
39. Wisnieski F, Calcagno DQ, Leal MF, et al Differential expression of histone deacetylase and acetyltransferase genes in gastric cancer and their modulation by trichostatin A. *Tumour Biol*. 2014;35:6373-6381.
40. Stojanovic N, Hassan Z, Wirth M, et al HDAC1 and HDAC2 integrate the expression of p53 mutants in pancreatic cancer. *Oncogene*. 2017;36:1804-1815.
41. Xu J, Tian S, Yin Z, et al MicroRNA-binding site SNPs in deregulated genes are associated with clinical outcome of non-small cell lung cancer. *Lung Cancer*. 2014;85:442-448.
42. Yang Z, Zhou L, Wu LM, Xie HY, Zhang F, Zheng SS. Combination of polymorphisms within the HDAC1 and HDAC3 gene predict tumor recurrence in hepatocellular carcinoma patients that have undergone transplant therapy. *Clin Chem Lab Med*. 2010;48:1785-1791.
43. Dang CV. MYC, metabolism, cell growth, and tumorigenesis. *Cold Spring Harbor Perspect Med*. 2013;3(8):pii: a014217.
44. Ling C, Del Guerra S, Lupi R, et al Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion. *Diabetologia*. 2008;51:615-622.
45. Zhang Y, Xu W, Li X, et al. Association between PPARGC1A gene polymorphisms and coronary artery disease in a Chinese population. *Clin Exp Pharmacol Physiol*. 2008;35:1172-1177.
46. Charos AE, Reed BD, Raha D, Szekely AM, Weissman SM, Snyder M. A highly integrated and complex PPARGC1A transcription factor binding network in HepG2 cells. *Genome Res*. 2012;22:1668-1679.
47. Li JD, Feng QC, Qi Y, Cui G, Zhao S. PPARGC1A is upregulated and facilitates lung cancer metastasis. *Exp Cell Res*. 2017;359:S0014482717304275.
48. Luo C, Lim JH, Lee Y, et al A PGC1alpha-mediated transcriptional axis suppresses melanoma metastasis. *Nature*. 2016;537:422-426.
49. Li JD, Feng QC, Qi Y, Cui G, Zhao S. PPARGC1A is upregulated and facilitates lung cancer metastasis. *Exp Cell Res*. 2017;359:356-360.
50. Zhai H, Zhong W, Yang X, Wu YL. Neoadjuvant and adjuvant epidermal growth factor receptor tyrosine kinase inhibitor (EGFR-TKI) therapy for lung cancer. *Transl Lung Cancer Res*. 2015;4:82-93.
51. Xu C, Zhou Q, Wu YL. Can EGFR-TKIs be used in first line treatment for advanced non-small cell lung cancer based on selection according to clinical factors? - A literature-based meta-analysis. *J Hematol Oncol*. 2012;5:62.
52. Mountzios G. Making progress in epidermal growth factor receptor (EGFR)-mutant non-small cell lung cancer by surpassing resistance: third-generation EGFR tyrosine kinase inhibitors (EGFR-TKIs). *Ann Transl Med*. 2018;6:140.
53. Kim Y, Lee B, Shim JH, et al. Concurrent genetic alterations predict the progression to target therapy in EGFR-mutated advanced non-small cell lung cancer. *J Thorac Oncol*. 2018;14(2):193-202.
54. Yu Z, Boggon TJ, Kobayashi S, et al Resistance to an irreversible epidermal growth factor receptor (EGFR) inhibitor in EGFR-mutant lung cancer reveals novel treatment strategies. *Cancer Res*. 2007;67:10417-10427.
55. Huang B, Liu B, Yang L, et al Functional genetic variants of c-Jun and their interaction with smoking and drinking increase the susceptibility to lung cancer in southern and eastern Chinese. *Int J Cancer*. 2012;131:E744-E758.
56. Cleary SP, Cotterchio M, Shi E, Gallinger S, Harper P. Cigarette smoking, genetic variants in carcinogen-metabolizing enzymes, and colorectal cancer risk. *Am J Epidemiol*. 2010;172:1000-1014.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Tang D, Zhao YC, Qian D, et al. Novel genetic variants in *HDAC2* and *PPARGC1A* of the CREB-binding protein pathway predict survival of non-small-cell lung cancer. *Molecular Carcinogenesis*. 2019;1–12. https://doi.org/10.1002/mc.23132