HHS Public Access

Author manuscript

Int J Cancer. Author manuscript; available in PMC 2019 November 04.

Published in final edited form as:

Int J Cancer. 2019 August 01; 145(3): 621–631. doi:10.1002/ijc.32128.

Genetic Variants in *RUNX3*, *AMD1* and *MSRA* in the Methionine Metabolic Pathway and Survival in Non-small Cell Lung Cancer Patients

Ka Chen[1,2,3,*], **Hongliang Liu**[2,3,*], **Zhensheng Liu**[2,3], **Sheng Luo**[4], **Edward F. Patz Jr.**[2,5], **Patricia G. Moorman**[2,6], **Li Su**[6], **Sipeng Shen**[7], **David C. Christiani**[7,8], **Qingyi Wei**[2,3,9,**]

[1]Research Center for Nutrition and Food Safety, Institute of Military Preventive Medicine, Third Military Medical University, Chongqing 400038, P. R. China

[2]Duke Cancer Institute, Duke University Medical Center, Durham, NC 27710, USA

[3]Department of Population Health Sciences, Duke University School of Medicine, Durham, NC 27710, USA

[4]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC 27710, USA

[5]Department of Radiology, Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, NC 27710, USA

[6]Department of Community and Family Medicine, Duke University Medical Center, Durham, NC, USA

[7]Departments of Environmental Health and Department of Epidemiology, Harvard School of Public Health, Boston, MA, 02115 USA

[8]Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA

[9]Department of Medicine, Duke University School of Medicine, Durham, NC 27710, USA

Abstract

Abnormal methionine dependence in cancer cells has led to methionine restriction as a potential therapeutic strategy. We hypothesized that genetic variants involved in methionine-metabolic genes are associated with survival in non-small cell lung cancer (NSCLC) patients. Therefore, we investigated associations of 16,378 common single-nucleotide polymorphisms (SNPs) in 97 methionine-metabolic pathway genes with overall survival (OS) in NSCLC patients using genotyping data from two published genome-wide association study (GWAS) datasets. In the single-locus analysis, 1,005 SNPs were significantly associated with NSCLC OS ($P < 0.05$ and false-positive report probability < 0.2) in the discovery dataset. Three SNPs (*RUNX3*

**Correspondence author: Qingyi Wei, M.D., Ph.D., Duke Cancer Institute, Duke University Medical Center and Department of Population Health Sciences, Duke University School of Medicine, 905 S LaSalle Street, Durham, NC 27710, USA, Tel.: (919) 660-0562, qingyi.wei@duke.edu.

*Ka Chen and Hongliang Liu contributed equally to this work

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

rs7553295G>T, *AMD1* rs1279590G>A and *MSRA* rs73534533C>A) were replicated in the validation dataset and their meta-analysis showed that adjusted hazards ratio [HR] of 0.82 [95% confidence interval (CI) =0.75-0.89] and P_{meta}=2.86 x 10^{-6}, 0.81 (0.73-0.91) and P_{meta}=4.63 x 10^{-4}, and 0.77 (0.68-0.89) and P_{meta}=2.07 x 10^{-4}, respectively). A genetics score of protective genotypes of these three SNPs revealed an increased OS in a dose-response manner (P_{trend} <. 0001). Further expression quantitative trait loci (eQTL) analysis showed significant associations between these genotypes and gene mRNA expression levels. Moreover, differential expression analysis further supported a tumor-suppressive effect of *MSRA*, with lower mRNA levels in both lung squamous carcinoma and adenocarcinoma (P <.0001 and <.0001, respectively) than in adjacent normal tissues. Additionally, low mutation rates of these three genes indicated the critical roles of these functional SNPs in cancer progression. Taken together, these genetic variants of methionine-metabolic pathway genes may be promising predictors of survival in NSCLC patients.

Keywords

Methionine; Non-small cell lung cancer; Genome-wide association study; Single-nucleotide polymorphism; Survival

Introduction

Lung cancer is the leading cause of cancer deaths worldwide, and non–small cell lung cancers (NSCLC) accounts for almost 80% of lung cancer deaths[1,2]. Therefore, NSCLC represents the most frequent type of bronchogenic carcinomas, consisting primarily of adenocarcinoma (LUAD), squamous cell carcinoma (LUSC), and to a lesser extent large-cell lung cancer. Despite improvements in the treatments in recent years, NSCLC prognosis has only marginally improved with a 5-year survival rate of only around 18% in the United States between 2007 and 2013 [2]. Although epidemiological studies have identified several risk factors for NSCLC, such as smoking and radon exposure [3], there are reports about altered mRNA and protein expression among thousands of genes, such as *p53, K-ras, PTEN and FHIT*, that also contribute to lung carcinogenesis [1]. Recently, some pathway-based hypothesis-driven studies using published genome-wide association study (GWAS) datasets have identified a number of genetic variants, i.e., single nucleotide polymorphisms (SNPs), with moderate but detectable effects on clinical outcomes of NSCLC, following by studying potential biological functions in some biological pathways, which have shed some light on prognosis prediction and possible individualized therapeutics [1,3–5].

Methionine is an essential amino acid with multiple roles in mammalian growth and development, including protein synthesis, methylation of DNA and polyamine synthesis [6–8]. Methionine dependence is a unique metabolic defect found only in transformed and malignant cells [6–8]. This defect is defined as the inability of cells to grow in vitro when methionine is replaced with its immediate precursor homocysteine [8]. In contrast, normal cells are relatively resistant to exogenous methionine restriction. It has been hypothesized that this defect in cancerous, as compared to normal, cells could be used for therapeutic purpose [9–11]. A methionine-free diet or methionine-deprived total parenteral nutrition causes regression of a variety of animal tumors [12]. It has also been suggested that vegan diets,

containing relatively lower methionine, may be a useful nutritional strategy controlling for cancer growth [13]. In addition, the methioninase, which depletes circulating levels of methionine, has also been identified as another useful strategy in limiting cancer growth in preclinical models [14]. Currently, methionine restriction in combination with chemotherapy has become a new focus in cancer treatment, and further investigation into fundamental mechanisms of methionine dependency is clearly needed.

Methionine dependence in cancer may be due to either genetic variants or alterations in expression of genes in the methionine *de novo* and *salvage* pathways. Previous studies have revealed that genetic variants of methionine metabolism genes, including the methionine synthase (*MTR*) and methionine synthase reductase (*MTRR*), may affect enzyme activities and thereby affect cancer risk in Turkish population and non-Hispanic whites [15, 16]. In addition, functional SNPs in *MTR, MTRR* and other genes related to the methionine metabolism have also been found to be associated with lung cancer prognosis in two small studies [17, 18]. In the present study, we used two large published GWAS datasets to determine whether common genetic variants in genes involved in the methionine metabolism pathway are associated with overall survival (OS) of NSCLC patients. Identification of promising prognostic biomarkers may scientific foundation for the metabolism-based therapeutics.

Methods

Study populations

Two independently published GWAS datasets were used in this study. The discovery dataset was obtained from the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, which enrolled patients between 1993 and 2011 from ten medical centers in the United States [19]. The validation dataset was obtained from the Harvard Lung Cancer Susceptibility (HLCS) Study. The study protocols were approved by institutional review boards at both the PLCO trial and the HLCS study with a written informed consent obtained from each of the subjects.

The PLCO trial enrolled 77,500 men and 77,500 women aged 55 to 74, who were randomized to either the intervention arm with screening or the control arm with standard care. All participants were followed for at least 13 years after enrollment. At enrollment, their blood samples and personal information including demographic characteristics, family history of cancer, smoking history and personal medical history were collected [20]. The PLCO dataset identified 1,185 NSCLC patients.

Genomic DNA extracted from the blood samples of the PLCO participants was genotyped with Illumina HumanHap240Sv1.0 and HumanHap550v3.0 (dbGaP accession: phs000093.v2.p2 and phs000336.v1.p1) [21, 22]. There were 1,185 Caucasian NSCLC patients with genotyping data and complete follow-up information for survival analysis, which were available in the subset of the PLCO lung cancer database. Tumor staging was determined according to the 5th edition American Joint Committee on Cancer (AJCC) staging system (5^{th} edition). The follow-up time was defined from lung cancer diagnosis to the last follow-up or time of death. OS was the primary endpoint of the current study, and disease-specific survival (DSS) of lung cancer was also examined

The validation dataset from the HLCS GWAS study included 984 histologically-confirmed Caucasian NSCLC patients as described previously [23]. The tumor histological classification was determined by two staff pulmonary pathologists at the Massachusetts General Hospital. Blood was collected from each patient was within 1–4 weeks of their diagnosis, and DNA was extracted from the blood samples with the Auto Pure Large Sample Nucleic Acid Purification System (QIAGEN Company, Venlo, Limburg, Netherlands). Genotyping data was obtained by using Illumina Humanhap610-Quad arrays, and imputation was performed by using MaCH1.0 based on the 1000 Genomes project.

Gene and SNP selection

Based on the databases of the Molecular Signatures Databases and GeneCards, 100 autosome genes related to the methionine metabolism were selected for further investigation (Supplementary Table S1). After excluding three genes in the X chromosome, 16,378 SNPs within the remaining 97 genes and their 2-kb flanking regions (genotyping rate ≥ 95%, minor allelic frequency (MAF) ≥ 0.05 and Hardy-Weinberg equilibrium (HWE) ≥ 1×10^{-6}) were extracted from and imputed for the PLCO dataset (dbGaP accession: phs000093.v2.p2 and phs000336.v1.p1) [21]. In total, 16,378 SNPs (1,680 genotyped and 14,698 imputed SNPs) were selected and used first in a single locus analysis. Significant SNPs were subjected to multiple test correction by a false-positive report probability (FPRP) method, followed by pairwise linkage disequilibrium (LD) analyses to select representative SNPs in each gene in high LD (r^2>0.8) and functional SNPs according to functional annotation based on RegulomeDB [24] and *P* value. Finally, significant, representative and potentially functional SNPs were further subjected to validation by the HLCS dataset.

Statistical analysis

Cox proportional hazards regression models were used to estimate the hazards ratio (HR) and 95% confidence interval (CI) for the associations of demographic and clinical characteristics with OS. Imputation was implemented with IMPUTE2 according to the 1000 Genomes CEU data (phase 1 release V3). The associations between SNPs and OS (in an additive genetic model) were analyzed by both univariate and multivariate Cox regression models using the GenABEL package of R software, with adjustment for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy and surgery, where appropriate or available from the original GWAS datasets [25]. For multiple testing corrections, the FPRP approach was used with a cut-off value of 0.2 to lower the probability of false positive findings, because the vast majority (near 90%) of the SNPs to be tested were imputed. Pairwise LD was estimated by using the data from the 1000 Genomes Project of 373 European individuals. Inverse variance weighted meta-analysis was performed to combine the results of discovery and validation studies. Cochran's Q statistics and I^2 were carried out to assess an inter-study heterogeneity. A fixed-effects model was used when no heterogeneity was found between the two studies (Q > 0.10 and I^2 < 25.0%); otherwise, a random-effects model was used. The meta-analysis of the two studies was performed by PLINK 1.07.

The number of protective genotypes (NPGs) was used as a genetic score to assess the combined effect of all independent and significant SNPs. Kaplan-Meier curves and log-rank

tests were used to estimate the effects of genotypes or NPGs on the cumulative probability of OS. The heterogeneity test of associations between subgroups in stratified analyses was performed by using the Chi-square-based Q-test to assess possible interactions. The receiver operating characteristic (ROC) curves and time-dependent area under the curve (AUC) were constructed from the logistic regression model with the survival ROC package of R software. Statistical significance of the improvement in AUC was analyzed by the Delong's test.

For those SNPs identified as significant, we first performed bioinformatics functional prediction by using two online tools: RegulomeDB (http://www.regulomedb.org) and HaploReg (http://archive.broadinstitute.org/mammals/haploreg/haploreg.php). Then, we performed the eQTL analysis using linear regression analysis between genotypes of SNPs and corresponding gene expression levels with the R software. Gene expression data were obtained from multiple sources: lymphoblastoid cell data of 373 European individuals from Genetic European Variation in Health and Disease Consortium (GEUVADIS) and the 1000 Genomes Project (phase I integrated release 3, March 2012) [26]; the whole blood and lung tissues data from the genotype-tissue expression (GTEx) project; tumor tissues and adjacent normal tissue data from the Cancer Genome Atlas (TCGA) database. In the TCGA database, associations between gene expression levels and lung cancer OS were accessible for 408 lung cancer patients of European descent with follow-up information; differences in mRNA expression levels between 109 paired lung cancer tissues and adjacent normal tissues (51 cases of LUSC and 58 cases of LUAD) were examined by the Student *t* test [27, 28]. The TCGA level 3 RNAseq data (LUSC_rnaseqv2_Level_3_RSEM_genes_normalized_data.2016012800.0.0.tar.gz and LUAD_Level_3_RSEM_genes_normalized_data_2016012800.0.0.tar.gz) were downloaded from the Broad TCGA GDAC site (http://gdac.broadinstitute.org). The mutation data of those identified genes in lung tumor tissues were also publically available from the database of the cBioPortal for Cancer Genomics (http://www.cbioportal.org).

Results

Basic characteristics of the two GWAS study populations

The overall workflow chart is shown in Supporting information Fig. 1A. Basic demographics and clinical characteristics of 1,185 NSCLC patients from the PLCO trial have been described previously [5], and seven clinical variables (i.e., age at diagnosis, sex, smoking status, histology, stage, chemotherapy and surgery) were found to be significantly associated with NSCLC OS and DSS. In the HLCS study, the basic characteristics of 984 NSCLC patients, including age, sex, smoking status, and histology and stage of lung cancer, were also previously described [5]. Although both studies included a Caucasian population, there had some differences in the distribution of age, sex, tumor histology and stage, and each of these factors were adjusted in the multivariate Cox models for survival analyses.

Multivariate analyses of association between SNPs and NSCLC OS in the PLCO trial

In the discovery PLCO dataset, multivariate Cox regression analysis was firstly performed to assess associations between 16,378 common SNPs (i.e., 1,680 genotyped and 14,698

imputed SNPs) of the methionine metabolism pathway genes and NSCLC OS with adjustment for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy and surgery. The Manhattan plot of associations between these variants and NSCLC OS in PLCO is shown in Supporting information Fig. 1B. The QQ plot of the observed *P* values showed a fairly uniform distribution (Supporting information Fig. 1C), although the actual *P* values deviated from the expected values at the early stage. In the single locus analysis, 1,390 SNPs were found to be significantly associated with NSCLC OS at $P < 0.05$ in an additive genetic model, of which 1,005 SNPs in 33 genes were still considered noteworthy with an FPRP value < 0.2. After that, we performed pairwise LD analyses ($r^2 > 0.6$) of these SNPs in each gene, and the tag SNPs with the lowest RegulomeDB scores (functional prediction) were chosen. As a result, there were 101 representative, potentially functional, SNPs in these 33 genes with $P < 0.05$ and FPRP <0.2, which were subjected to validation.

Validation analysis with Harvard dataset and meta-analysis of two studies

To confirm the findings from the PLCO dataset, the 101 representative SNPs were further subjected to validation in the HLCS dataset. As shown in Table 1, three SNPs in three genes identified in the discovery phase remained statistically significant ($P < 0.05$): there were rs7553295 in runt-related transcription factor 3 (*RUNX3*), rs1279590 in S-adenosylmethionine decarboxylase 1 (*AMD1*) and rs73534533 in methionine sulfoxide reductase (*MSRA*), all of which were associated with an improved survival in both datasets. Meta-analysis of these three SNPs from both datasets showed that these associations remained statistically significant, without evidence for heterogeneity across the two GWAS datasets (Table 1).

Three independent SNPs as NSCLC OS predictors

We further performed functional prediction with RegulomeDB and Haploreg for these three validated SNPs (Supporting information Table 2). As indicated by RegulomeDB, the scores of *RUNX3* rs7553295, *AMD1* rs1279590 and *MSRA* rs73534533 were 4, 5 and 6, respectively. Functional annotation of these SNPs in HaploReg demonstrated that *RUNX3* rs7553295 overlaps with a promoter and an enhancer, potentially disrupting the motif of Zfp691 and affecting the mRNA expression; *AMD1* rs1279590 overlaps with an enhancer in 18 tissues (e.g., lung, IMR90 fetal lung fibroblasts) and may disrupt three motifs, including Nanog, POU class 2 homeobox 2 (Pou2f2/Oct-2) and POU class 5 homeobox 1 (Pou5f1/Oct-4) and thus affect mRNA expression of the corresponding gene; similarly, *MSRA* rs73534533 may disrupt seven motifs. As the online functional prediction tools suggested that these SNPs were biologically functional, then we used initial stepwise Cox regression analyses to identify whether these three SNPs were independent predictors of OS. The results suggested that these three validated representative SNPs were statistically significant independent predictors of NSCLC OS (Table 2).

For each of the three independent SNPs, univariate and multivariable Cox regression analysis were further performed to evaluate their effects on risk of death with adjustment for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy and surgery (Table 3). In the PLCO dataset, the risk of death was significantly decreased with the increasing number of rs7553295T, rs1279590 A and rs73534533 A alleles ($P_{trend} <0.0001$,

0.011 and 0.005, respectively). Similarly, consistent trends were observed in the risk of DSS of in the PLCO dataset (P_{trend} = 0.0002, 0.004 and 0.013, respectively) (Supporting information Table 3). For the illustrative purposes, Kaplan-Meier survival curves of these associations of these SNPs with NSCLC OS and DSS are depicted in Figs. 1A and 1C, respectively. In addition, regional association plots for variants in *RUNX3, AMD1* and *MSRA*, including the 250-kb regions flanking the neighborhoods of these genes, are shown in Supporting information Fig. 2.

Combined effects of the three independent SNPs

To evaluate the joint effect of the three independent SNPs on OS, the protective genotypes (i.e., *RUNX3* rs7553295 GT+TT, *AMD1* rs1279590 GA+AA *and MSRA* rs73534533 CA +AA) were combined into a genetic score as the number of protective genotypes (NPGs) (Table 3). The trend test indicated that an increased number of NPGs was associated with a decreased risk of death in the PLCO dataset (P_{trend} < .0001). We next dichotomized all patients into a low- protective group (0-1 NPGs) and a high-protective group (2-3 NPGs). We observed that the high-protective group had a HR of 0.72 (95% CI =0.60-0.86, P = 0.0003), compared with the low-protective group. For the illustrative purposes, Kaplan-Meier survival curves of these associations of the NPGs with NSCLC OS are depicted in Fig. 1B. The analysis of lung cancer DSS showed the results similar to that of OS and that the high-protective group had a significantly better prognosis (HR = 0.67, 95%CI = 0.55-0.82 and P <.0001) (Supporting information Table 3). Kaplan-Meier survival curves of these associations of the NPGs with NSCLC DSS are shown in Fig. 1D.

Stratified analyses for associations between NPGs and NSCLC OS

Stratified analysis was performed to investigate whether the combined effect of protective genotypes on NSCLC OS was modified by covariates in the PLCO dataset. Compared with those with 0-1 NPGs, individuals with 2-3 NPGs showed better survival, with no statistically significant differences in strata defined by age, sex, smoking status, histology, tumor stage or type of treatment (Supporting information Table 4). No statistically significant heterogeneity or interactions were observed among these subgroups.

The ROC curves and time dependent AUC

We further estimated the predictive value of the NPGs with time-dependent AUC and ROC curves in PLCO. As shown in Supporting information Fig. 3, the time-dependent AUC plot indicated an improved prediction performance with the addition of NPGs to the model with covariates compared to the model with covariates only. When evaluating the five-year and ten-year NSCLC OS, adding NPGs into the model with covariates increased the AUCs from 86.71% to 86.95% (P = 0.154) and from 84.51% to 84.95% (P = 0.067) respectively, which were marginally significant. A similar trend was observed for the predictive value of the NPGs in DSS with significantly increased AUC for ten-year DSS (from 83.29% to 83.96%, P = 0.049).

The eQTL analyses

To evaluate correlations between SNPs and their corresponding mRNA expression levels, we primarily used the RNA-Seq data of lymphoblastoid cell lines from 373 European descendants in the 1000 Genomes Project. As shown in Fig. 2, rs1279590 GA+AA (or the A allele) was significantly associated with a decreased mRNA expression level of *AMD1* (a trend test in a dominant model: $P = 0.017$) and rs73534533 CA+AA (or the A allele) was significantly associated with an increased mRNA level of *MSRA* (a trend test in a dominant model: $P = 0.027$). However, there was no significant association between rs7553295 GT +TT and *RUNX3* mRNA expression levels (Supporting information Fig. 4A).

In Westra's 2013 study [29], we also found that rs7553295 GT+TT (or the T allele) and rs1279590 GA+AA (or the A allele) were associated with decreased mRNA expression levels of *RUNX3* and *AMD1* in the whole blood ($P = 1.25 \times 10^{-5}$ and 5.45×10^{-6}, respectively).

Next, we performed the eQTL analysis using data from the GTEx project. In lung normal tissues from the donors, the rs7553295 T allele (i.e., GT+TT genotypes and rs1279590 A allele (i.e., GA+AA genotypes) had a non-significant trend in correlation with decreased mRNA expression levels of *RUNX3* ($P = 0.140$, Supporting information Fig. 4B) and *AMD1* ($P = 0.390$, Supporting information Fig. 4C), but this non-significant trend was not observed between the rs73534533 A allele (i.e., CA+AA genotypes) and mRNA expression levels of *MSRA* ($P = 0.62$, Supporting information Fig. 4D).

Finally, we performed an analysis of SNP and mRNA expression correlation using the expression data in tumor tissues from 408 NSCLC patients (237 lung squamous cell carcinomas and 171 lung adenocarcinomas) from the TCGA database. Only rs7553295 GT +TT (or the T allele) showed a non-significant trend in correlation with decreased mRNA expression levels of *RUNX3* in all patients with NSCLC, LUSC or LUAD ($P = 0.174$, 0.597 and 0.197, respectively) (Supporting information Figs. 4E–4G). The data for other two SNPs were not available in this database.

Differential expression analysis in the TCGA dataset

Using the TCGA dataset, we evaluated mRNA expressions levels of 109 paired tumor and adjacent normal tissue samples in NSCLC. As shown in Fig. 3A, lung cancer tissues had a lower mRNA expression level of *MSRA*, compared with that in the adjacent normal tissues (135.77 ± 90.57 in tumor vs. 269.65 ± 57.63 in normal, $P<.0001$). Moreover, the mRNA expression levels of *MSRA* were also lower in tumor tissues either in the 51 pairs of LUSC tissues (91.02 ± 53.48 in LUSC vs. 269.23 ± 46.67 in normal, $P<.0001$) and the 58 pairs of LUAD tissues (175.12 ± 98.31 in LUAD vs. 270.02 ± 66.21 in normal, $P<.0001$). However, *AMD1* or *RUNX3* mRNA expression levels in paired tumor and normal adjacent tissues in either LUSC or LUAD were not significantly different in either LUSC or LUAD.

Mutation analyses

Finally, we investigated the mutation status of *RUNX3, AMD1 and MSRA* in lung tumor tissues by using the public database of the cBioPortal for Cancer Genomics. As shown in

Fig. 3B, *RUNX3* had a low somatic mutation rate in NSCLC (mutation rate = 0.79% [9/1144] in the TCGA 2016 [30]study) and LUAD (mutation rate = 8.84% [3/34], 1.09% [2/183] and 0.87% [2/230] in the MSKCC [31], Broad [32] and TCGA[28] studies, respectively); *AMD1* had a low somatic mutation rate in NSCLC (mutation rate = 0.35% [4/1144] in the TCGA 2016 study[30]), LUAD (mutation rate = 0.87% [2/230], in the TCGA study [28]) and LUSC (mutation rate = 1.1% [1/178]) in the TCGA study [27]; and *MSRA* also had a low somatic mutation rate in NSCLC (mutation rate = 0.79% [9/1144] in the TCGA 2016 study [30]) and LUAD (mutation rate = 0.43% [1/230] in the TCGA study [28]). These results suggested that given the low mutation rates, the functional SNPs in *RUNX3, AMD1* and *MSRA* may play a more important role in the dysregulation of mRNA expression in tumor tissues than mutations had.

Discussion

Recent findings in cancer metabolism have resulted in new translational opportunities for drug development, dietary intervention in cancer prevention, and biomarkers for the efficacy of the anti-metabolic chemotherapeutic drugs. Methionine dependence is a unique metabolic defect commonly seen in various cancers, and methionine restriction in combination with chemotherapy has become a new focus in cancer treatment [33]. In the present study, we first identified rs7553295 G>T, rs1279590 C>A and rs73534533 C>A as predictors of NSCLC OS. Remarkably, functional relevance of these SNPs with their corresponding mRNA expression levels was further confirmed by assessing publicly available datasets. These findings suggested that genetic variants in the methionine metabolism pathway genes might have biological roles in NSCLC progression, possibly through a mechanism of modulating expression of these genes, which provides new scientific insights into metabolism-based therapeutics.

In the present study, it appears that the independent effects of genetic variants in *RUNX3, AMD1* and *MSRA* on OS and DSS in NSCLC patients can add to a much strong effect through a genetic sore, which represents the combined effects of the three genetic variants. Remarkably, the combined effect was consistent across analyses of different datasets and through stratified analyses, in the presence of other covariates, such as age, sex, smoking status, histology and different treatment strategies. Furthermore, based on the genotype-phenotype correlation analysis and *in silico* functional prediction, we believe that our results are biologically plausible.

RUNX3, located on chromosome 1p36.1, encodes a protein that belongs to the runt domain family of transcription factors acting as master regulators of gene expression during normal tissue development [34, 35]. *RUNX3* has been identified as a tumor suppressor in a variety of human malignancies, including lung cancer [36–38]. Due to aberrant hypermethylation of its CpG islands [39], *RUNX3* expression has been reported to be lost in the range of 19~95% of lung cancer cell lines and tissue samples, and *RUNX3* inactivation has been causally linked to the preneoplastic stage of lung adenocarcimoma [38]. In the present study, we found that rs7553295, located in the 2-kb of the 3' of *RUNX3*, was associated with a better survival in NSCLC patients. According to the Haploreg data [40, 41], rs7553295 overlaps with an enhancer activity cluster, which is classified as a genic enhancer by the 15-state core model

and as a transcribed 3' enhancer by the 25-state model. Meanwhile, rs7553295 overlaps with histone modification markers H3K4me1, H3K4me4 and H3K27ac, which all contribute to the chromatin state assignment at this SNP location. Additional data in Haploreg show that this SNP changes the match to a regulatory motif Zfp691 (zinc finger protein 691). In line with that, we found the *RUNX3* rs7553295 T allele was associated with a significant decrease in mRNA expression levels of *RUNX3* in whole blood. Therefore, rs7553295 probably affects gene expression levels by modifying the accessibility of chromatin during transcription.

AMD1, located on chromosome 6, encodes the S-adenosylmethionine decarboxylase 1, an enzyme that catalyzes the conversion of S-adenosyl methionine to S-adenosylmethioninamine, controlling the second rate-limiting step in the polyamine biosynthetic pathway [42]. Polyamines are highly regulated essential cations that are elevated in rapidly proliferating tissues of diverse cancers. Reduced levels of intracellular polyamines activate checkpoints that constrain proliferation, as seen in senescent and post-mitotic cells, while enhanced polyamine synthesis accompanies oncogenic proliferation [42, 43]. Recent studies demonstrated that rapamycin complex 1 (mTORC1)-dependent AMD1 upregulation sustains polyamine metabolism in prostate cancer [44]. In the present study, we found that NSCLC patients with genotypes of *AMD1* rs1279590 GA+AA had better survival. According to the Haploreg, this SNP overlaps with an enhancer in many tissues including that of the lung. In line with that, we found the *AMD1* rs1279590 A allele was associated with a significant decrease in mRNA expression levels of *AMD1* in lymphobpastoid cell lines and whole blood, which supports an oncogenic effect of *AMD1*. Meanwhile, rs1279590 might have some effects on three motifs including Nanog [45], Pou2f2 [46] and Pou5f1 [47], which are all associated with tumorigenesis and metastasis. Taken together, the evidence may possibly explain the mechanism underlying the association between rs1279590 and NSCLC OS, but further functional investigation is needed.

MSRA, located on chromosome 8p23.1, encodes the methionine sulfoxide reductase, which is known to protect proteins from oxidation and acts as a reactive oxygen species (ROS) scavenger [48]. In the present study, our findings suggested that *MSRA* rs73534533 is associated with a better survival in NSCLC patients. Consistent with that, rs73534533 CA + AA genotypes were associated with a significant increase in mRNA expression levels of *MSRA* in lymphobplastoid cell lines. Various studies have shown that MSRA is down-regulated in several human tumors including lung carcinoma and that the reduction of MSRA levels results in increased cell proliferation, extracellular matrix degradation and up-regulation of VEGF, consequently leading to a more aggressive cellular phenotype, both *in vivo* and *in vitro* [49, 50]. Consistently, the evidence from differential expression analyses supports a tumor-suppressive effect of *MSRA* on NSCLC, with lower mRNA levels in both lung squamous carcinoma and adenocarcinoma. According to Haploreg, this SNP influences genomic instability (GI). Consistent with that, rs73534533 CA + AA genotypes are associated with a significant increase in mRNA expression levels of *MSRA* in lymphobplastoid cell lines, leading to suppressive genic effects on NSCLC. Moreover, rs73534533 might have effects on tumor progression through disruption of seven transcription regulators motifs, including homeobox protein Hox-B13, which might influence angiogenesis and epithelial cell maturation [51]; paired box protein Pax-4, which is

correlated with cell differentiation and apoptotic process [52]; and THAP Domain Containing 1, which regulates endothelial cell proliferation and G1/S cell-cycle progression [53]. Taken together, this evidence may partly explain the biological and molecular mechanisms underlying the observed associations.

There are several limitations in the present study. First, although some clinical factors were included in the analyses, the information about methionine intake, nutritional status, or nutrition-based treatment received by the patients was not available. Second, some top SNPs from the PLCO trial, including genetic variants of *MTRR*, which were identified in some other previous studies [17, 18], were not validated in the Harvard study. Different distributions of the basic characteristics between the two study populations might partially explain the reason of non-validated SNPs. Additional validation by studies with larger sample sizes are needed to confirm these findings. Third, we used a less stringent FPRP method to control for multiple comparisons in the discovery dataset. Although this may lead to some false positive findings, it is noteworthy the effects of the identified SNPs on NSCLC OS were consistently observed in both the discovery and replication datasets and that these three SNPs all have potential functions in mRNA expression regulation. Lastly, although independent SNPs were discovered one one dataset and validated in another dataset and their combined effects on NSCLC OS were demonstrated, no direct biological experiments were conducted *in vitro* or *in vivo* for additional validations. Therefore, additional functional studies are needed to explore the exact biological mechanisms of these SNPs or genes underlying NSCLC progression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the National Cancer Institute for access to NCI's data collected by the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI. The author would also like to acknowledge dbGaP repository for providing the cancer genotyping datasets. The accession numbers for the datasets of lung cancer are phs000336.v1.p1 and phs000093.v2.p2. Qingyi Wei was supported by a start-up funds from Duke Cancer Institute, Duke University Medical Center and support from the Duke Cancer Institute as part of the P30 Cancer Center Support Grant (Grant ID: NIH CA014236). Ka Chen is sponsored by the China Scholarship Council for studying at the Duke University. The Harvard Lung Cancer Susceptibility Study was supported by NIH grants CA092824, CA074386, and CA090578 to David C. Christiani. We thank all individuals who participated in this project.

References

1. Potti A, Mukherjee S, Petersen R, Dressman HK, Bild A, Koontz J, Kratzke R, Watson MA, Kelley M, Ginsburg GS, West M, Harpole DH Jr., et al. Retraction: A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. N Engl J Med 2006;355:570–80. [PubMed: 16899777] N Engl J Med 2011;364: 1176. [PubMed: 21366430]
2. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA Cancer J Clin 2017;67: 7–30. [PubMed: 28055103]
3. Liu H, Liu Z, Wang Y, Stinchcombe TE, Owzar K, Han Y, Hung RJ, Brhane Y, McLaughlin J, Brennan P, Bickeboller H, Rosenberger A, et al. Functional variants in DCAF4 associated with lung cancer risk in European populations. Carcinogenesis 2017;38: 541–51. [PubMed: 28383684]

4. Xu YWY, Liu H, Kang X, Li W, Wei Q. Genetic variants of genes in the Notch signaling pathway predict overall survival of non-small cell lung cancer patients in the PLCO study. Oncotarget 2016; 7: 6176–87.

5. Wang Y, Liu H, Ready NE, Su L, Wei Y, Christiani DC, Wei Q. Genetic variants in ABCG1 are associated with survival of nonsmall-cell lung cancer patients. Int J Cancer 2016;138: 2592–601. [PubMed: 26757251]

6. Finkelstein JD. Methionine metabolism in mammals. J Nutr Biochem 1990;1: 228–37. [PubMed: 15539209]

7. Wallace PHSCD; Hoffman RM. Altered Methionine Metabolism Occurs in All Members of a Set of Diverse Human Tumor Cell Lines. Journal of cellular physiology 1984;119: 29–34. [PubMed: 6707100]

8. Hoffman RM. Altered methionine metabolism, DNA methylation and oncogene expression in carcinogenesis. Biochimica et Biophysica Acta 1984;738: 49–87. [PubMed: 6204687]

9. Cellarier E, Durando X, Vasson MP, Farges MC, Demiden A, Maurizis JC, Madelmont JC, Chollet P. Methionine dependency and cancer treatment. Cancer Treat Rev 2003;29: 489–99. [PubMed: 14585259]

10. Hoshiya Y, Guo H, Kubota T, Inada T, Asanuma F, Yamada Y, Koh J, Kitajima M, Hoffman RM. Human tumors are methionine dependent in vivo. Anticancer research 1995;15: 717–8. [PubMed: 7645948]

11. Breillout F, Antoine E, Poupon MF. Methionine dependency of malignant tumors: a possible approach for therapy. Journal of the National Cancer Institute 1990;82: 1628–32. [PubMed: 2213904]

12. Newman AC, Maddocks ODK. One-carbon metabolism in cancer. Br J Cancer 2017;116: 1499–504. [PubMed: 28472819]

13. McCarty MF, Barroso-Aranda J, Contreras F. The low-methionine content of vegan diets may make methionine restriction feasible as a life extension strategy. Med Hypotheses 2009;72: 125–8. [PubMed: 18789600]

14. Epner DE. Can dietary methionine restriction increase the effectiveness of chemotherapy in treatment of advanced cancer? Journal of the American College of Nutrition 2001;20: 443S–9S; discussion 73S-75S. [PubMed: 11603655]

15. Aksoy-Sagirli P, Erdenay A, Kaytan-Saglam E, Kizir A. Association of Three Single Nucleotide Polymorphisms in MTR and MTRR Genes with Lung Cancer in a Turkish Population. Genetic testing and molecular biomarkers 2017;21: 428–32. [PubMed: 28537809]

16. Shi Q, Zhang Z, Li G, Pillow PC, Hernandez LM, Spitz MR, Wei Q. Polymorphisms of methionine synthase and methionine synthase reductase and risk of lung cancer: a case-control analysis. Pharmacogenetics and genomics 2005;15: 547–55. [PubMed: 16006998]

17. Matakidou A, El Galta R, Rudd MF, Webb EL, Bridle H, Eisen T, Houlston RS. Prognostic significance of folate metabolism polymorphisms for lung cancer. Br J Cancer 2007;97: 247–52. [PubMed: 17533396]

18. Jin G, Huang J, Hu Z, Dai J, Tang R, Chen Y, Xu L, Huang X, Shu Y, Shen H. Genetic variants in one-carbon metabolism-related genes contribute to NSCLC prognosis in a Chinese population. Cancer 2010;116: 5700–9. [PubMed: 20737570]

19. Hocking WG, Hu P, Oken MM, Winslow SD, Kvale PA, Prorok PC, Ragard LR, Commins J, Lynch DA, Andriole GL, Buys SS, Fouad MN, et al. Lung cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. Journal of the National Cancer Institute 2010;102: 722–31. [PubMed: 20442215]

20. Oken MM, Marcus PM, Hu P, Beck TM, Hocking W, Kvale PA, Cordes J, Riley TL, Winslow SD, Peace S, Levin DL, Prorok PC, et al. Baseline chest radiograph for lung cancer detection in the randomized Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. Journal of the National Cancer Institute 2005;97: 1832–9. [PubMed: 16368945]

21. Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, Lee M, Popova N, Sharopova N, Kimura M, Feolo M. NCBI's Database of Genotypes and Phenotypes: dbGaP. Nucleic Acids Res 2014;42: D975–9. [PubMed: 24297256]

22. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet 2007;39: 1181–6. [PubMed: 17898773]

23. 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: 961–70. [PubMed: 23921845]

24. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. Genome research 2012;22: 1790–7. [PubMed: 22955989]

25. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics 2007;23: 1294–6. [PubMed: 17384015]

26. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T, Ferreira PG, Barann M, Wieland T, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013;501: 506–11. [PubMed: 24037378]

27. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. Nature 2012;489: 519–25. [PubMed: 22960745]

28. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014;511: 543–50. [PubMed: 25079552]

29. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, Zhernakova A, Zhernakova DV, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nature Genetics 2013;45: 1238. [PubMed: 24013639]

30. Campbell JD, Alexandrov A, Kim J, Wala J, Berger AH, Pedamallu CS, Shukla SA, Guo G, Brooks AN, Murray BA, Imielinski M, Hu X, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat Genet 2016;48: 607–16. [PubMed: 27158780]

31. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS, Miller ML, Rekhtman N, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 2015;348: 124–8. [PubMed: 25765070]

32. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, Sougnez C, Auclair D, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell 2012;150: 1107–20. [PubMed: 22980975]

33. Cavuoto P, Fenech MF. A review of methionine dependency and the role of methionine restriction in cancer growth control and life-span extension. Cancer Treat Rev 2012;38: 726–36. [PubMed: 22342103]

34. Ito K, Liu Q, Salto-Tellez M, Yano T, Tada K, Ida H, Huang C, Shah N, Inoue M, Rajnakova A, Hiong KC, Peh BK, et al. RUNX3, a novel tumor suppressor, is frequently inactivated in gastric cancer by protein mislocalization. Cancer Res 2005;65: 7743–50. [PubMed: 16140942]

35. Bae SC, Choi JK. Tumor suppressor activity of RUNX3. Oncogene 2004;23: 4336–40. [PubMed: 15156190]

36. Araki K, Osaki M, Nagahama Y, Hiramatsu T, Nakamura H, Ohgi S, Ito H. Expression of RUNX3 protein in human lung adenocarcinoma: implications for tumor progression and prognosis. Cancer Sci 2005;96: 227–31. [PubMed: 15819721]

37. Lee KS, Lee YS, Lee JM, Ito K, Cinghu S, Kim JH, Jang JW, Li YH, Goh YM, Chi XZ, Wee H, Lee HW, et al. Runx3 is required for the differentiation of lung epithelial cells and suppression of lung cancer. Oncogene 2010;29: 3349–61. [PubMed: 20228843]

38. Lee YS, Lee JW, Jang JW, Chi XZ, Kim JH, Li YH, Kim MK, Kim DM, Choi BS, Kim EG, Chung JH, Lee OJ, et al. Runx3 inactivation is a crucial early event in the development of lung adenocarcinoma. Cancer Cell 2013;24: 603–16. [PubMed: 24229708]

39. Li Q-L, Kim H-R, Kim W-J, Choi J-K, Hee Lee Y, Kim H-M, Shan Li L, Kim H, Chang J, Ito Y, Youl Lee K, Bae S-C. Transcriptional silencing of the RUNX3 gene by CpG hypermethylation is

associated with lung cancer. Biochemical and Biophysical Research Communications 2004;314: 223–8. [PubMed: 14715269]

40. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res 2016;44: D877–81. [PubMed: 26657631]
41. 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: D930–4. [PubMed: 22064851]
42. Gamble LD, Hogarty MD, Liu X, Ziegler DS, Marshall G, Norris MD, Haber M. Polyamine pathway inhibition as a novel therapeutic approach to treating neuroblastoma. Front Oncol 2012;2: 162. [PubMed: 23181218]
43. Evageliou NF, Haber M, Vu A, Laetsch TW, Murray J, Gamble LD, Cheng NC, Liu K, Reese M, Corrigan KA, Ziegler DS, Webber H, et al. Polyamine Antagonist Therapies Inhibit Neuroblastoma Initiation and Progression. Clin Cancer Res 2016;22: 4391–404. [PubMed: 27012811]
44. Zabala-Letona A, Arruabarrena-Aristorena A, Martin-Martin N, Fernandez-Ruiz S, Sutherland JD, Clasquin M, Tomas-Cortazar J, Jimenez J, Torres I, Quang P, Ximenez-Embun P, Bago R, et al. mTORC1-dependent AMD1 regulation sustains polyamine metabolism in prostate cancer. Nature 2017;547: 109–13. [PubMed: 28658205]
45. Iv Santaliz-Ruiz LE, Xie X, Old M, Teknos TN, Pan Q. Emerging role of nanog in tumorigenesis and cancer stem cells. Int J Cancer 2014;135: 2741–8. [PubMed: 24375318]
46. Wang SM, Tie J, Wang WL, Hu SJ, Yin JP, Yi XF, Tian ZH, Zhang XY, Li MB, Li ZS, Nie YZ, Wu KC, et al. POU2F2-oriented network promotes human gastric cancer metastasis. Gut 2016;65: 1427–38. [PubMed: 26019213]
47. Miyoshi N, Fujino S, Ohue M, Yasui M, Takahashi Y, Sugimura K, Tomokuni A, Akita H, Kobayashi S, Takahashi H, Omori T, Miyata H, et al. The POU5F1 gene expression in colorectal cancer: a novel prognostic marker. Surgery today 2018.
48. Moskovitz J, Bar-Noy S, Williams WM, Requena J, Berlett BS, Stadtman ER. Methionine sulfoxide reductase (MsrA) is a regulator of antioxidant defense and lifespan in mammals. Proc Natl Acad Sci U S A 2001;98: 12920–5. [PubMed: 11606777]
49. Methionine sulfoxide reductase A down-regulation in human breast cancer cells results in a more aggressive phenotype.
50. Hansel A, Kuschel L, Hehl S, Lemke C, Agricola HJ, Hoshi T, Heinemann SH. Mitochondrial targeting of the human peptide methionine sulfoxide reductase (MSRA), an enzyme involved in the repair of oxidized proteins. FASEB journal: official publication of the Federation of American Societies for Experimental Biology 2002;16: 911–3. [PubMed: 12039877]
51. Kristiansen I, Stephan C, Jung K, Dietel M, Rieger A, Tolkach Y, Kristiansen G. Sensitivity of HOXB13 as a Diagnostic Immunohistochemical Marker of Prostatic Origin in Prostate Cancer Metastases: Comparison to PSA, Prostein, Androgen Receptor, ERG, NKX3.1, PSAP, and PSMA. International journal of molecular sciences 2017;18.
52. Kalousova A, Benes V, Paces J, Paces V, Kozmik Z. DNA binding and transactivating properties of the paired and homeobox protein Pax4. Biochem Biophys Res Commun 1999;259: 510–8. [PubMed: 10364449]
53. Aguilo F, Zakirova Z, Nolan K, Wagner R, Sharma R, Hogan M, Wei C, Sun Y, Walsh MJ, Kelley K, Zhang W, Ozelius LJ, et al. THAP1: Role in Mouse Embryonic Stem Cell Survival and Differentiation. Stem cell reports 2017;9: 92–107. [PubMed: 28579396]

What's new?

Methionine dependence is a unique metabolic defect seen in various cancers including non-small cell lung cancer (NSCLC), and thus genetic variants involved in methionine-metabolic genes may have an impact on clinical outcomes of NSCLC. By using publically available genome-wide association study datasets, we identified *RUNX3* rs7553295 T, *AMD1* rs1279590 A and *MSRA* rs73534533 A variant genotypes that are associated with survival of NSCLC patients. Once validated by other investigators, these variants could possibly be used in personalized clinical management of NSCLC.

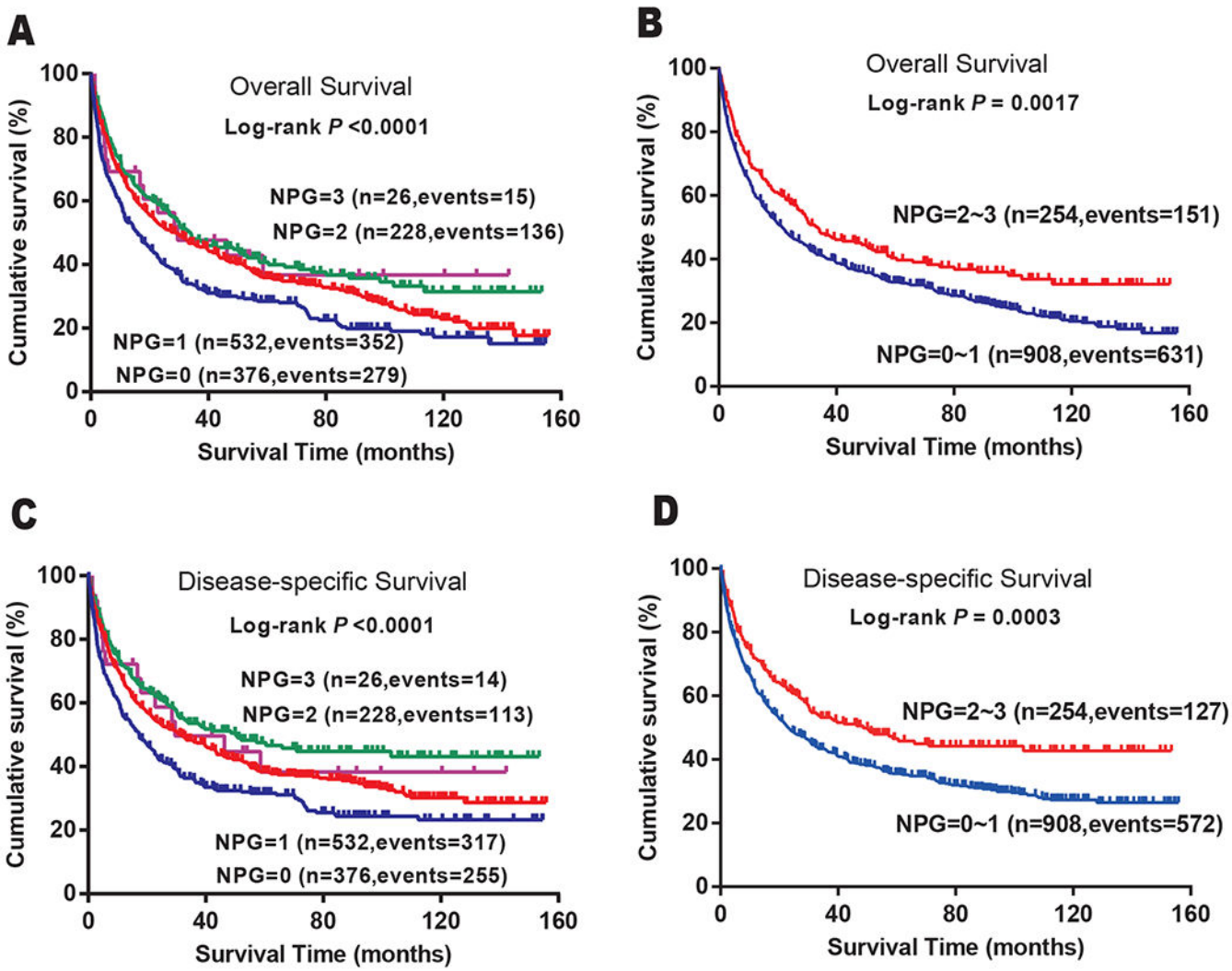

Figure 1.

The combined risk genotypes and survival prediction. Kaplan-Meier survival curves for the overall survival of the combined risk genotypes (A) and dichotomized groups of the NPG (B) in the PLCO dataset; Kaplan-Meier survival curves for the disease-specific survival of the combined risk genotypes (C) and dichotomized groups of the NPG (D) in the PLCO dataset. NPG, number of protective genotypes; PLCO, PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening trial.

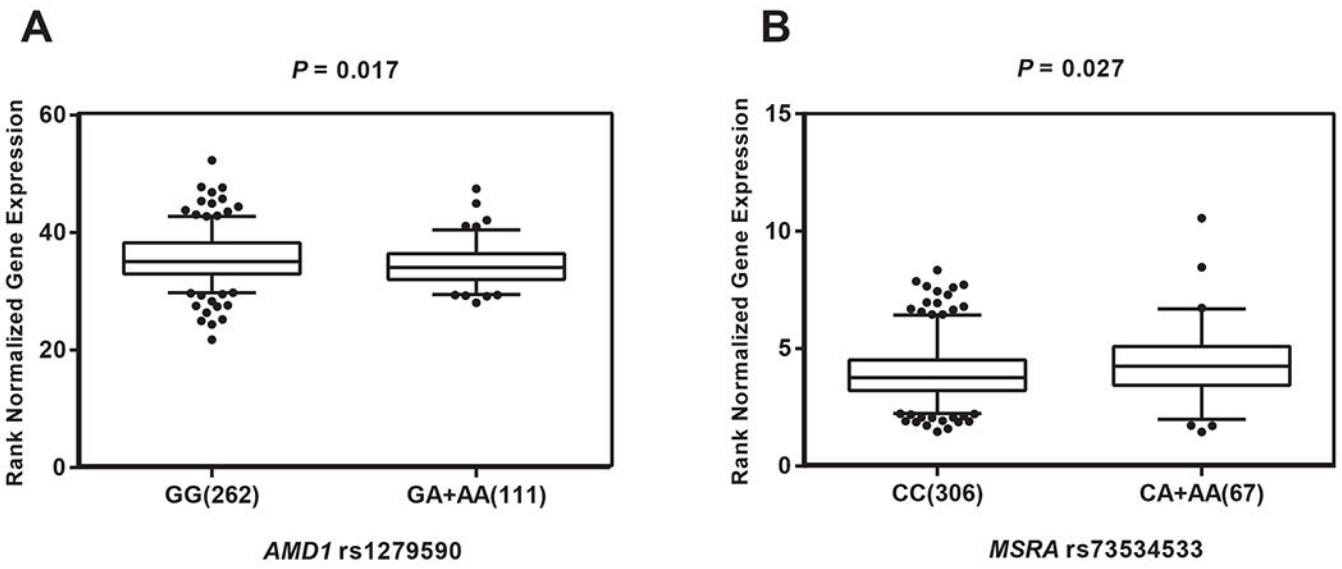

Figure 2.

Associations between the protective genotypes and their corresponding mRNA expression levels. The eQTL for *AMD1* rs1279590 (A) and *MSRA* rs73534533 (B) in 373 Europeans from the 1000 Genomes Project in the dominant model. eQTL, expression quantitative trait loci analysis.

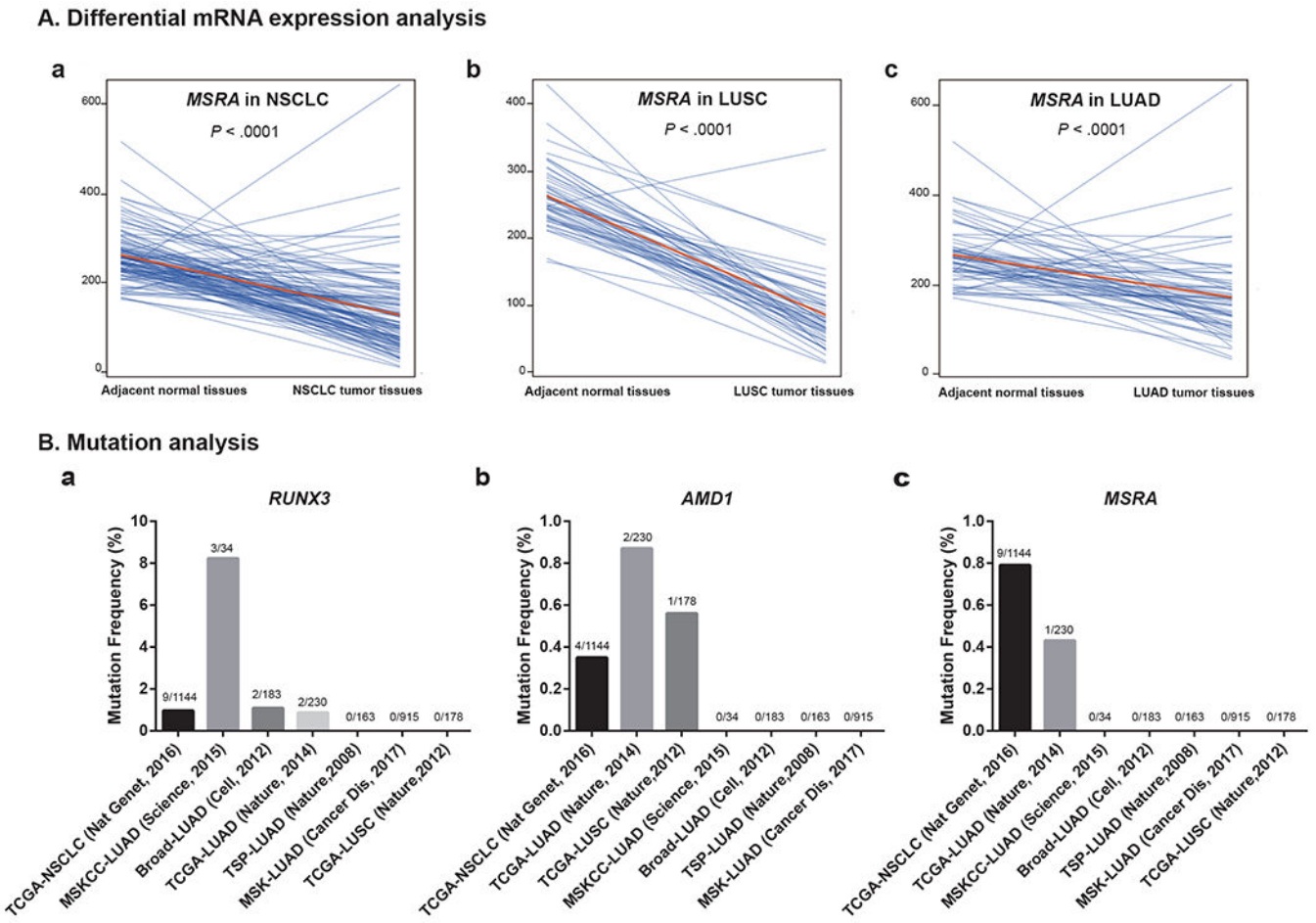

Figure 3.

Differential mRNA expression analysis and mutation analysis of the three genes. (A) Differential mRNA expression analysis of *MSRA* by using the data generated by The Cancer Genome Atlas (TCGA). Higher *MSRA* mRNA expression were found in the tumor tissues of 109 NSCLC (a), 51 LUSC (b) and 58 LUAD (c) than in the adjacent normal tissues ($P < .0001$, $P < .0001$ and $P < .0001$, respectively). (B) Mutation analysis of *RUNX3, AMD1 and MSRA* gene in non-small cell lung tumor tissues by using public available data in the database of the cBioportal for Cancer Genomics **(**http://www.cbioportal.org). *RUNX3* (a), *AMD1* (b) and *MSRA* (b) had low mutation frequency in NSCLC, LUAD and LUSC.

Table 1.

Meta-analysis of three validated SNPs of genes in the methionine metabolism pathway using two published GWAS datasets

				PLCO (n = 1,185)				Harvard (n = 984)			Meta-analysis			
SNP	**Allele**[1]	**Gene**	**Chr**	**EAF**	**HR (95% CI)**[2]	***p***[2]	**FPRP**	**EAF**	**HR (95% CI)**[3]	***p***[3]	p_{het}[6]	I^2	**HR (95% CI)**[5]	***p***[7]
rs7553295[4]	G/T	*RUNX3*	1	0.28	0.79 (0.70–0.89)	8.0×10^{-5}	0.001	0.28	0.85 (0.75–0.96)	0.007	0.424	0	0.82 (0.75–0.89)	2.86×10^{-6}
rs1279590[5]	G/A	*AMD1*	6	0.14	0.83 (0.72–0.96)	0.011	0.088	0.12	0.79 (0.66–0.95)	0.013	0.693	0	0.81 (0.73–0.91)	4.63×10^{-4}
rs73534533[5]	C/A	*MSRA*	8	0.09	0.76 (0.63–0.92)	0.005	0.049	0.10	0.79 (0.65–0.96)	0.016	0.998	0	0.77 (0.68–0.89)	2.07×10^{-4}

Abbreviations: SNP: single nucleotide polymorphism; GWAS: genome-wide association study; PLCO: Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening trial; EAF: effect allele frequency; HR: hazards ratio; CI: confidence interval; p_{het}: *p* value for heterogeneity by Cochrane's Q test.

[1] Reference/effect allele; EAF, effect allele frequency;

[2] Adjusted for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3 and PC4 (PC = principal component);

[3] Adjusted for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery. PC1, PC2, PC3;

[4] Genotyped SNP in the PLCO trial.

[5] Imputed SNP in the PLCO trial.

[6] P_{het}: *P* value for heterogeneity by Cochrane's Q test;

[7] Meta-analysis in the fix-effect model;

Table 2.

Independent predictors of OS obtained from stepwise cox regression analysis of selected variables in the PLCO trial

Parameter[1]	**Category**[2]	**Frequency**	**HR (95% CI)**	***p***
Age	≤71/>71	636/549	1.03 (1.01–1.04)	<0.0001
Sex	Male/Female	698/487	0.78 (0.67–0.91)	0.001
Smoking status	Never/Current/Former	115/423/647	1.18 (1.06–1.31)	0.003
Histology	AD/SC/others	577/285/323	1.17 (1.08–1.28)	0.0003
Stage	I-IIIA/IIIB-IV	655/528/2	2.79 (2.30–3.39)	<0.0001
Chemotherapy	No/Yes	639/538/8	0.58 (0.49–0.70)	<0.0001
Radiotherapy	No/Yes	762/415/8	0.95 (0.80–1.11)	0.498
Surgery	No/Yes	637/540/8	0.21 (0.17–0.27)	<0.0001
RUNX3 rs7553295	GG/GT/TT	612/483/90	0.78 (0.69–0.88)	<0.0001
AMD1 rs1279590	GG/GA/AA	868/290/25	0.81 (0.70–0.94)	0.005
MSRA rs73534533	CC/CA/AA	967/187/10	0.76 (0.63–0.92)	0.004

Abbreviations: OS, overall survival; HR, hazards ratio; CI, confidence interval.

[1]Stepwise analysis included age, sex, smoking, stage, histology, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3, PC4 and 3 SNPs (rs7553295, rs1279590 and rs73534533).

[2]The "category" was used as the reference.

Table 3.

Associations of the three validated SNPs in the methionine metabolism process with OS of NSCLC in the PLCO trial

	Frequency		Univariate analysis		Multivariate analysis[1]	
Genotype	**All**	**Death (%)**	**HR (95% CI)**	***p***	**HR (95% CI)**	***p***
RUNX3 rs7553295 G > T						
GG	612	439 (71.73)				
GT	483	305 (63.15)	0.76 (0.66–0.87)	0.0003	0.74 (0.64–0.86)	<0.0001
TT	90	54 (60.00)	0.77 (0.58–1.02)	0.071	0.71 (0.53–0.95)	0.022
Trend test				0.0007		<0.0001
GT + TT	573	359 (62.65)	0.77 (0.67–0.88)	0.0002	0.73 (0.64–0.85)	<0.0001
AMD1 rS1279590 G > A						
GG	868	597 (68.78)				
GA	290	183 (63.10)	0.82 (0.70–0.98)	0.025	O.80 (0.67–0.94)	0.008
AA	25	17 (68.00)	1.02 (0.63–1.65)	0.948	0.03 (0.50–1.37)	0.461
Trend test				0.066		0.011
GA + AA	315	200 (63.49)	0.84 (0.72–0.99)	0.034	0.80 (0.60–0.94)	0.007
MSRA rs7353A533 C > A						
CC	967	660 (68.25)				
CA	187	117 (62.57)	0.85 (0.70–1.04)	0.106	0.72 (0.59–0.89)	0.002
AA	10	6 (60.00)	0.90 (0.40–2.01)	0.800	1.13 (0.50–2.53)	0.774
Trend test				0.119		0.005
CA + AA	197	123 (62.44)	0.85 (0.70–1.03)	0.104	0.74 (0.60–0.90)	0.003
Number of protective genotypes[2]						
0	376	279 (74.20)				
1	532	352 (66.17)	0.75 (0.64–0.88)	0.0004	0.63 (0.53–0.74)	<0.0001
2	228	136 (59.65)	0.64 (0.52–0.78)	<0.0001	0.54 (0.43–0.66)	<0.0001
3	26	15 (57.69)	0.63 (0.37–1.06)	0.080	0.55 (0.33–0.93)	0.026
Trend test				<0.0001		<0.0001

	Frequency		Univariate analysis		Multivariate analysis[1]	
Genotype	**All**	**Death (%)**	**HR (95% CI)**	***p***	**HR (95% CI)**	***p***
0–1	908	631 (69.49)				
2–3	254	151 (59.45)	0.75 (0.36–0.90)	0.002	0.72 (0.60–0.86)	0.0003

Abbreviations: SNPs, single nucleotide polymorphisms; OS, overall survival; NSCLC, non-small cell lung cancer; PLCO: Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening trial; HR, hazards ratio.

[1]Adjusted for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3 and PC4;

[2]Protective genotypes were *RUNX3* rs7553295 GT + TT, *AMD1* rs1279590 GA + AA and *MSRA* rs73534533 CA + AA.