HHS Public Access

Author manuscript

Int J Cancer. Author manuscript; available in PMC 2021 April 27.

Published in final edited form as:

Int J Cancer. 2020 September 15; 147(6): 1559–1570. doi:10.1002/ijc.32932.

Novel genetic variants in *PLIN2, SULT2A1* and *UGT1A9* genes of the Ketone metabolism pathway are associated with survival of non-small cell lung cancer

Dongfang Tang[#1,2,3], **Yu Chen Zhao**[#2,3], **Hongliang Liu**[2,3], **Sheng Luo**[4], **Jeffrey M. Clarke**[2,5], **Carolyn Glass**[2,6], **Li Su**[7], **Sipeng Shen**[7], **David C. Christiani**[7,8], **Wen Gao**[1,**], **Qingyi Wei**[2,3,4,**]

[1]Department of Thoracic Oncology, Huadong Hospital, Fudan University, Shanghai 200040, 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 Medicine, Duke University School of Medicine, Durham, NC 27710, USA

[6]Department of Pathology, Duke University School of Medicine, Durham, NC 27710, 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

[#] These authors contributed equally to this work.

Abstract

The ketone metabolic pathway is a principle procedure in physiological homeostasis and induces cancer cells to switch between glycolysis and oxidative phosphorylation as their main energy production. We conducted a two-phase analysis for associations between genetic variants in the ketone metabolism pathway genes and survival of non-small cell lung cancer (NSCLC) by using genotyping data from published genome-wide association studies (GWASs). The discovery used genotyping dataset from the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial in the multivariate Cox proportional hazards regression analysis with Bayesian false discovery probability (≤0.80) for multiple testing correction to evaluate associations between 27,322 (2,176 genotyped and 25,146 imputed) single-nucleotide polymorphisms (SNPs) in 162 genes and survival of 1,185 NSCLC patients. Subsequently, significant SNPs were further validated with 984

[**]Correspondence to: Wen Gao, Department of Thoracic Oncology, Huadong Hospital affiliated to Fudan University, Shanghai 200040, China; gaowen5921@163.com; or Qingyi Wei, 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.

Conflicts of interest: The authors declare no conflict of interest.

NSCLC patients in another dataset from the Harvard Lung Cancer Susceptibility GWAS study. Finally, three independent and potentially functional SNPs in three different genes (i.e., *PLIN2* rs7867814 G>A, *SULT2A1* rs2547235 C>T and *UGT1A9* rs2011404 C>T and) were independently associated with NSCLC overall survival, with a combined hazards ratio of 1.22 [95% confidence interval = 1.09–1.36 and *P*=0.0003], 0.82 (0.74–0.91 and *P*=0.0002) and 1.21 (1.10–1.33 and *P*=0.0001), respectively. Additional expression quantitative trait loci analysis found that the survival-associated *PLIN2* rs7867814 GA+AA genotypes, but not *UGT1A9* rs2011404 CT+TT genotypes and *SULT2A1* rs2547235 CT+TT genotypes, were significantly associated with increased mRNA expression levels in 373 lymphoblastoid cell lines. These results indicated that *PLIN2* variants may be potential predictors of NSCLC survival through regulating the *PLIN2* expression.

Keywords

Non-small cell lung cancer; Ketone metabolism pathway; genetic susceptibility; single-nucleotide polymorphism; Survival

Introduction

Lung cancer ranked the top for cancer-related mortality, with over a million deaths each year worldwide [1]. Non-small-cell lung cancer (NSCLC) is the most common histological type, accounting for approximately 85% of lung cancer patients [2]. Although about 80% of lung cancer cases are attributable to smoking, lung cancer patients accounts for about only 15% of smokers [3]. Chemo-radiotherapy has long been a standard treatment for unresectable advanced NSCLC; however, median survival of NSCLC patients remained only 15–20% [4]. Recent consolidative immunotherapy for stage III lung cancer has improved the survival, but the impact on the long-term outcomes remain unknown [5]. Studies have shown that genetic variants are responsible for individual variation in response to treatment outcomes [6]. Thus, it is important to identify genetic factors in pivotal genes and pathways that may influence progression, metastasis and outcomes of NSCLC.

A number of genome-wide association studies (GWASs) on susceptibility to lung cancer have identified multiple genetic loci at chromosomal regions of 3q28, 5p15.33, 6p21.33, 6p22.1, 13q13.1, 15q25.1 and 22q12.1 in European populations [7–12], but few GWASs on clinical outcome of lung cancer were reported. However, most of the published GWASs had mainly focused on single nucleotide polymorphism (SNPs), few of which reached the genome-wide significance and often did not have a clear biological function [13]. In the post-GWAS era, identification of genetic variants with moderate but detectable effects and potential biological functions may provide additional insights into the complex mechanisms of cancer development and tumor progression [13].

The difference in energy production between normal tissue and cancer cells has long been considered one of the available targets in the therapy of cancer [14]. For example, the efficacy of many current chemotherapeutic agents and radiation is at least partially dependent on the production of reactive oxygen species (ROS) [15]. Studies have indicated that ketone bodies and medium-chain fatty acids (MCFAs) may be used as a tool to induce

cancer cells to switch between glycolysis and oxidative phosphorylation during energy production, indicating that ketogenic and MCFA-enriched diets may be beneficial to cancer patients who received chemoradiotherapies [16]. Ketogenic diets rich in medium-chain triglycerides have demonstrated its inhibitory effects on cancer growth. For example, it was observed that prostate cancer cells had a lower ability to utilize dietary fatty acids, compared to normal cells, indicating possible therapeutic potential [16].

To date, the roles of SNPs in the ketone metabolism pathway genes in cancer development and their functionality related to tumor progression and metastasis are still unknown. In the present study, by using publicly available GWAS datasets, we performed a ketone metabolism pathway gene-set analysis to evaluate the associations between genetic variants in this gene-set and survival of NSCLC patients.

Materials and Methods

Study populations

In the present study, as shown in the study flowchart (Figure 1), the discovery phase used the dataset from the GWAS of 1,185 NSCLC patients from the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, with the approval from Duke Internal Review Board (#Pro00054575) and data access approval of the dbGAP database (#6404) from the National Center for Biotechnology Information (NCBI). The PLCO is an NCI funded multicenter randomized trial of screening for cancer from ten medical centers in the United States between 1993 and 2011 [17]. The screening trial enrolled 77,500 men and 77,500 women aged 55–74. All individuals were randomized to either the intervention arm with screening or the control arm with standard care [17]. The PLCO trial collected blood specimens from the first screening visit and gathered extensive information about each individual, including smoking history, family history of cancer and demographic information [18]. All participants were followed-up for at least 13 years after the enrollment. Genomic DNA extracted from the blood samples was genotyped in a genome-wide association study (GWAS) with Illumina HumanHap240Sv1.0, HumanHap300v1.1 and HumanHap550v3.0 (dbGaP accession: phs000093.v2.p2 and phs000336.v1.p1) [19, 20]. In 1,187 Caucasian NSCLC patients 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 1,185 NSCLC patients, 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 all subjects signed a written informed consent permitting the use of biospecimens for further research [17].

The validation phase used the dataset from the Harvard Lung Cancer Susceptibility (HLCS) GWAS study with 984 Caucasian patients with histology-confirmed NSCLC. The histological classification of the tumors was recorded by two staff pulmonary pathologists at the Massachusetts General Hospital. The time of blood collection was within 1–4 weeks of the diagnosis for each patient, and DNA was extracted from the 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 1,000 Genomes project [21]. Details of the participants in the HLCS study were described previously [22]. The comparison of the characteristics between the PLCO trial and the HLCS study is presented in Supplemental Table 1.

Gene and SNP selection

We selected the genes or a gene-set involved in the ketone metabolism pathway through the Molecular Signatures Database (http://software.broadinstitute.org/gsea/msigdb/index.jsp), by the keyword "Ketone" and "Metabolism". After removal of duplicated genes and exclusion of genes in the X chromosome, 162 genes remained as candidate genes for further analysis (Supplemental Table 2). We first performed imputation for the 162 genes plus the 500-kb flanking buffer regions by using IMPUTE2 and the 1,000 Genomes Project data (phase 3) [21]. 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 a Hardy-Weinberg equilibrium P value ≥ 1×10^{-5}. As a result, 2,176 genotyped SNPs were chosen from the PLCO GWAS dataset with an additional 25,146 SNPs.

Statistical analysis

The follow-up time in both PLCO and HLCS datasets were from the diagnosis of lung cancer to the last follow-up time or time of death. Overall survival (OS) of lung cancer was the primary endpoint, and the disease-specific survival (DSS) was also examined. In the single-locus analysis, multivariate Cox proportional hazards regression analysis was used to evaluate the association between each of the SNPs and OS (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 genotyping dataset using the GenABEL package of R software [23]. Since the majority of SNPs were imputed by using a high level of linkage disequilibrium (LD), we used the Bayesian false discovery probability (BFDP) with a cutoff value of 0.80 for multiple test corrections as recommended [24]. We assigned a prior probability of 0.10 to detect an HR of 3.0 for an association with variant genotypes or minor alleles of the SNPs with $P < 0.05$. These significant SNPs identified in the single-locus analysis were summarized in a Manhattan plot. Then, we validated the remaining significant SNPs by using the HLCS dataset with the following criteria: SNPs passed the threshold of BFDP ≤ 0.8, potentially functional SNPs predicted by HaploReg [25], SNPinfo [26] and RegulomeDB [27], and tagging SNPs based on the LD analysis. To identify independent SNPs, we included the validated SNPs in a multivariate stepwise Cox model with adjustment for demographic characteristics,, clinical variables and the top four principal components of the genotyping data in the PLCO dataset as well as previously published SNPs from the same PLCO dataset. The combined analysis of discovery and validation datasets was also performed to provide a summary of the results. The fixed-effects model was applied, if the Cochran's Q-test P value > 0.100 and the heterogeneity statistic (I^2) < 50%; otherwise, the random-effects model was employed. The detailed LD information with independent SNPs were shown in regional association plots. The combination of unfavorable genotypes was also used to estimate the cumulative effects

of the identified SNPs in survival. Kaplan-Meier survival curves were used to visualized the survival associated with the genotypes.

Expression quantitative trait loci (eQTL) analysis was further performed to assess correlations between SNPs and mRNA expression levels of their associated genes by using linear regression analysis with the R (version 3.5.0) software. The mRNA expression data were obtained from lymphoblastoid cell lines derived from the 373 European descendants included in the 1,000 Genomes Project [21] in addition to the whole blood and normal lung tissues in the genotype-tissue expression (GTEx) project [28]. Additional data from the Cancer Genome Atlas (TCGA) database (dbGaP Study Accession: phs000178.v9.p8) were also used to assess the differences in mRNA expression levels between paired tumor tissues and adjacent normal tissues by the paired *t* test [29] as well as the association between mRNA expression levels and OS through the Kaplan-Meier (KM) analysis (n=1926) (http://kmplot.com/analysis/index.php?p=service&cancer=lung). Finally, we constructed receiver operating characteristic (ROC) curves and performed time-dependent ROC analysis to assess prediction accuracy of models integrating both clinical and genetic variables on NSCLC survival with the "timeROC" package in R (version 3.5.0) [30]. Unless specified otherwise, all statistical analyses were performed using the SAS software (version 9.4; SAS Institute, Cary, NC, USA).

Results

Associations between SNPs in the Ketone metabolism pathway gene-set and NSCLC OS in both PLCO and HLCS datasets

The discovery phase used the data from 1,185 NSCLC patients whose basic characteristics of have been described previously [31] (Supplemental Table 1, also see Supplemental Table 3). In the PLCO discovery dataset with an additive genetic model, the single-locus multivariate Cox models with adjustment for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery and first four of the 10 principal components (Supplemental Table 4), 673 SNPs were significantly associated with NSCLC OS after multiple test correction by BFDP of ≤ 0.8, which were summarized in a Manhattan plot (Figure 2A), which were further validated by the HLCS dataset. As a result, we identified ten potentially functional SNPs. Further combined analysis showed that these ten SNPs of the two datasets were all associated with OS ($P \leq 0.001$ for all), and no heterogeneity between the two studies was observed (Table 1).

Identification of independent SNPs associated with OS of NSCLC in the PLCO dataset

Because the HLCS study only provided the summary data without detailed genotyping data, we had to use the PLCO dataset to identify independent SNPs with adjustment for other co-variables. In a multivariate stepwise Cox model with adjustment for available demographics and clinical variables and the first four principal components as well as previously published SNPs from the PLCO GWAS dataset, three imputed SNPs in three different genes (i.e., rs7867814 in *PLIN2*, rs2547235 in *SULT2A1* and rs2011404 in *UGT1A9* with a p value of 0.003, 0.004 and 0.025, respectively) remained independently associated with NSCLC OS. Regional association plots for these three SNPs are shown in Supplemental Figure 1. To

identify potentially functional SNPs associated with NSCLC OS, we used three online bioinformatics tools (i.e., SNPinfo [26], RegulomeDB [27] and HaploReg [25]). According to RegulomeDB [27], rs7867814 had a score of 4 and rs73158145 had a score of 6, indicating minimal binding evidence (Supplemental Table 5). In HaploReg [25], these SNPs were predicted to cause either three or more motifs changed: rs7867814 had enhancer histone marks and rs2011404 had promoter histone marks; however, all SNPs did not had obvious functions based on SNPinfo [26].

As shown in Table 3, patients with *PLIN2* rs7867814 A and *SULT2A1* rs2547235 T alleles had a worse OS (P_{trend} = 0.002 or 0.0008) and DSS (P_{trend} = 0.003 or 0.002), while patients with *UGT1A9* rs2011404 T alleles had a better OS (P_{trend} = 0.005) and DSS (P_{trend} = 0.006). In comparison with the reference genotype in a dominant genetic model, *PLIN2* rs7867814 GA+AA and *SULT2A1* rs2547235 CT+TT were associated with a significantly worse OS (HR=1.30, 95% CI=1.10–1.53 and *P*=0.003 for rs7867814 GA+AA; and HR=1.22, 95% CI=1.05–1.42 and *P*=0.011 for rs2547235 CT+TT) and DSS (HR=1.28, 95% CI=1.07–1.53 and *P*=0.006 for rs7867814 GA+AA; and HR=1.20, 95% CI=1.02–1.41 and *P*=0.025 for rs73158145 CT+TT), whereas *UGT1A9* rs2011404 CT+TT were associated with a better OS (HR=0.82, 95% CI=0.70–0.96 and *P*=0.015 for rs2011404 CT+TT) and DSS (HR=0. 82, 95% CI=0.69–0.96 and *P*=0.016 for rs2011404 CT+TT).

Combined and stratified analysis of the three independent SNPs in the PLCO dataset

To provide a better accumulative effect of the three SNPs on survival, we combined their unfavorable genotypes (i.e., rs7867814 GA+AA, rs2547235 CT+TT and rs2011404 CC +CT) into a genetic score to divide all NSCLC patients into four groups. As shown in Table 3, in the multivariate analysis, an increased genetic score was associated with a worse survival (trend test: *P*<0.001 for both OS and DSS). After dichotomizing the genetic score, we re-grouped all the patients into a low-risk group (0– 1 risk score) and a high-risk group (2– 3 risk scores). Patients in the high-risk group had both poorer OS (HR=1.32, 95% CI=1.13–1.53 and *P*=0.0003) and DSS (1.32, 1.13–1.54 and 0.0006), compared with those in the low-risk group. Kaplan-Meier survival curves were presented to depict the associations between unfavorable genotypes and NSCLC OS and DSS (Figure 2B).

To assess the ability of the unfavorable genotypes to predict NSCLC OS and DSS, we compared the area under the receiver operating characteristic curve (AUC) from the model for clinical variables and previously published SNPs with or without unfavorable genotypes. The addition of unfavorable genotypes to the prediction model (not including published SNPs) of five-year survival rate non-significantly increased the AUC from 87.97% to 88.30% (*P*=0.160, Supplemental Figure 2a); similarly, the addition of unfavorable genotypes to the prediction model of five-year disease-specific survival non-significantly increased the AUC from 88.13% to 88.54% (*P*=0.275, Supplemental Figure 2b). The time-dependent AUC curve was provided to quantify the ability of unfavorable genotypes to predict NSCLC survival through the entire follow-up period (Supplemental Figure 2c and 2d). We then performed stratified analysis to evaluate whether the effects of combined unfavorable genotypes on NSCLC OS and DSS was modified by age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy and surgery (Supplemental Table 6 and

Supplemental Table 7). The results showed that no significant interactions were found (P>0.05).

in silico functional validation

Experimental data from the ENCODE project [27] (Supplemental Figure 3) showed that the *PLIN2* rs7867814 to be located in a DNase I hypersensitive site, where the DNase hypersensitivity and histone modification H3K4Me1 acetylation indicated strong signals for active enhancer and promoter functions. To further explore potential functions of these SNPs, we performed the eQTL analysis and found that only the *PLIN2* rs7867814 A allele showed a significant correlation with increased mRNA expression levels of the gene in both additive (P=0.003, Figure 2C–a) and dominant (P=0.005, Figure 2C–b) models. Furthermore, the *PLIN2* rs7867814 A allele also showed a significant correlation with increased mRNA expression levels of the gene in additive model (P=0.01, Supplemental Figure 4a) and recessive model (P=3.31E-05, **Figure 4c**) in lung adenocarcinoma tissues in the TCGA dataset. Taken together, these findings suggest that the *PLIN2* rs7867814 A, but not the *UGT1A9* rs2011404 T and *SULT2A1* rs2547235 T alleles, may influence its gene expression at the transcriptional level.

Additionally, in 111 paired NSCLC tumor and adjacent normal tissue samples obtained from the TCGA database, we found that expression levels of *PLIN2* were higher in the adjacent normal tissues than in the tumor tissues (P<0.001, Figure 3a, 3b and 3c), while the higher expression levels were not associated with a better NSCLC OS (Supplemental Figure 5a). On the other hand, the expression levels of *UGT1A9* were also higher in the adjacent normal tissues than in the tumor tissues (P<0.001, Figure 3d, 3e and 3f), while the higher expression levels were associated with a worse NSCLC OS, and lower expression of *SULT2A* was associated with a better NSCLC OS [32] (Supplemental Figure 5b).

Mutation analysis

It is likely that the effects of gene mutations in the tumor tissues may overwhelmed the effects of germline SNPs. Therefore, we investigated the mutation status of *PLIN2*, *UGT1A9* and *SULT2A1* in lung tumor tissues by using the public database of the cBioPortal for Cancer Genomics. As shown in Supplemental Figure 6a, *PLIN2* had a low somatic mutation rate in NSCLC (0.79% 9/1,144) in the TCGA 2016 study [33], LUAD (2.19% 4/183 and 0.88% 5/566) in the Broad [34] and TCGA PanCan studies [29], respectively; and LUSC (1.12% 2/178 and 0.21% 1/487) in the TCGA pub and TCGA PanCan studies, respectively. These results suggest that the functional SNPs in *PLIN2* may play a rather important role in the dysregulation of mRNA expression in tumor tissues, considering a low mutation rate of the *PLIN2* gene in tumor tissues.

Additionally, *UGT1A9* also had a low somatic mutation rate in NSCLC (0.96% 11/1,144) in the TCGA 2016 study [33] (Supplemental Figure 6b); a relatively higher mutation rate in LUAD (2.19% 4/183 and 1.06% 4/566), in the Broad [34] and TCGA PanCan studies [29], respectively; and LUSC (2.25% 4/178 and 0.62% 3/487) in the TCGA pub and TCGA PanCan studies, respectively. Finally, *SULT2A1* also had a low somatic mutation rate in NSCLC (1.14% 13/1,144) in the TCGA 2016 study [33] (Supplemental Figure 6c), LUAD

(1.74% 4/230 and 1.06% 6/566), in the TCGA Pub [34] and TCGA PanCan studies [29], respectively; and LUSC (1.69% 3/178 and 1.03% 5/487) in the TCGA pub and TCGA PanCan studies, respectively. Therefore, the roles of SNPs in *UGT1A9* and *SULT2A1* in regulating gene expression and NSCLC survival remain to be investigated.

Discussion

In the present study, we identified three novel genetic variants in the ketone metabolism gene-set (i.e., *PLIN2* rs7867814 G>A, *UGT1A9* rs2011404 C>T and *SULT2A1* rs2547235 C>T) that were significantly associated with NSCLC survival. Additionally, *PLIN2* rs7867814 G>A appeared to have an effect on *PLIN2* mRNA expression, which makes this SNP-associated risk of death biologically plausible.

The ketone metabolism is one of the central nodes in physiological homeostasis, because ketone bodies are regarded as vital metabolic mediators, when carbohydrates are abundant [35]. Studies have implicated ketone formation in the biological function of cancer cells. For instance, treatment of pancreatic cancer cells with ketone bodies could effectively inhibit cancer cell growth, proliferation and the cells' glycolysis pathway [36]; furthermore, ketogenic diet may reduce tumor weight and glycemia in the animals with implanted cancer [37]. Another study found that the splice variants might block two main enzymes (*HMGCS2* and *HMGCL*) for ketone-body synthesis in specific human tissues, such as lung and thymus [38]. Taken together, the ketone metabolism is one critical physiological pathway that has been considered an available target for cancer treatment, but the role of SNPs in the ketone metabolism gene-set in tumorigenesis was not fully understood.

The PAT family members, including *PLIN1* and *PLIN2/ADRP*, played an essential role in the formation or degradation of lipid droplets. *PLIN2* was originally found in fat and steroid-generated cells, and lipid droplets were the major consumables during the ketone metabolism [39]. The expression of *PLIN2* showed an increasing trend during embryo development, suggesting its involvement in the maintenance of lipid stocks in cells [39]. For example, it has been shown that that *PLIN2* might play an important role in lipid formation 3T3-L1 murine adipocytes [40], but few studies of the effect of *PLIN2* on cancer cells have been reported. One study reported that higher expression of *PLIN2* was an independent prognostic factor of clear cell renal cell carcinoma and that knockdown of *PLIN2* could promote the proliferation of carcinoma cells and enhance cell invasion and migration [41]. These indicated that *PLIN2* could act as a suppressor gene in cancer biology, consistent with our finding that the expression of *PLIN2* was elevated in normal tissues than in lung cancer tissues in the TCGA dataset. Another study found that the minor allele of the missense polymorphism Ser251Pro in *PLIN2* disrupted lipolysis and was associated with reduced plasma triglyceride concentration, suggesting an effect on the metabolic progress [42]. We also showed that the *PLIN2* rs7867814 A allele was correlated with an increased mRNA expression of *PLIN2* and thus likely plays a role in the lipid metabolism of NSCLC, which was associated with a poorer survival of NSCLC; furthermore, *PLIN2* rs7867814 is predicted to be located in DNase I hypersensitive site with considerable levels of DNase hypersensitivity and histone modification H3K4Me1 acetylation. Considering the lower

somatic mutation rate of *PLIN2* in NSCLC, therefore, we believed that the rs7867814 A allele may also lead to an enhanced transcriptional activity in the tumors.

Human cytosolic sulfotransferase (SULT) 2A1 catalyzed dehydroepiandrosterone (DHEA) sulfation in the adrenal cortex. One study found that three *SULT2A1* nonsynonymous coding SNPs were associated with decreased levels of both expression and activity, compared to the wild-type cDNA, when expressed in COS cells (one prostates cancer cell-line), but no significant associations of various *SULT2A1* alleles with prostates cancer risk [43]. In the present study, we showed that the *SULT2A1* rs2547235 T allele was associated with a poorer survival of NSCLC, but we did not find any evidence for the *SULT2A1* rs2547235 T allele to have an effect on mRNA expression levels in the tissues tested in public databases, although higher expression levels in the NSCLC were shown to be associated with a poorer NSCLC survival; therefore, it is possible that other molecular mechanisms may be involved in the abnormal expression levels of *SULT2A1* in the tumors and need to be further investigated.

Glucuronidation by the UDP-glucuronosyltransferase enzymes (UGTs) was one of the primary detoxification pathways of dietary heterocyclic amines (HCAs) and polycyclic aromatic hydrocarbons (PAHs) [44]. *UGT1A9* was one of the most active UGT towards the hydroxy metabolites of benzo(a)pyrene (BaP) [45,46] and also has the capacity to conjugate *N*-OH-PhIP particularly at the N^3-position [47,48]. This suggests that any genetic alterations reducing the *UGT1A9* activity or expression could influence the elimination of HCAs or PAHs. Previous studies have shown that the *UGT1A9*-275 AT genotype were associated with a higher expression level of *UGT1A9 in vitro* [49, 50]. In the present study, we did not find evidence for an effect of the *UGT1A9* rs2011404 T allele on mRNA expression levels, although this allele was associated with a better survival in NSCLC; Additionally, *UGT1A9* mRNA expression levels were higher in normal tissues than in tumor tissues in the TCGA dataset, but higher expression levels in NSCLC were associated with a poorer survival; therefore, we could not determine whether *UGT1A9* was an oncogene or a suppressor in NSCLC. Because the mutation analysis showed a relatively higher mutation rate of *UGT1A9* in both LUAD and LUSC, we speculated that the possible reason for the observed abnormal expression levels of *UGT1A9* in tumors may be affected by the somatic mutations.

In conclusion, we demonstrated that three independent functional SNPs (i.e., *PLIN2* rs7867814 G>A, *UGT1A9* rs2011404 C>T and *SULT2A1* rs2547235 C>T) were significantly associated with NSCLC survival in both the PLCO trial and HLCS GWAS datasets. The eQTL analysis found that the survival-associated *PLIN2* rs7867814 GA+AA genotypes were correlated with significantly increased mRNA expression levels of *PLIN2*, but such a correlation was not seen for *UGT1A9* rs2011404 CT+TT and *SULT2A1* rs2547235 CT+TTT genotypes. Our findings provide some new clues for further functional studies of these three genes to understand the molecular mechanisms underlying the observed associations with outcomes the patients with NSCLC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

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 datasets. The accession numbers for the datasets 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 also partly supported by the Duke Cancer Institute as part of the P30 Cancer Center Support Grant (Grant ID: NIH/NCI CA014236). The Harvard Lung Cancer Susceptibility Study was supported by NIH grants 5U01CA209414, CA092824, CA074386 and CA090578 to David C. Christiani.

Grant sponsor:

NIH; Grant numbers: CA090578, CA074386, CA092824, 5U01CA209414;

Grant sponsor:

Duke Cancer Institute as part of the P30 Cancer Center Support Grant; Grant numbers: NIH/NCI CA014236;

Grant sponsor:

V Foundation for Cancer Research; Grant numbers: D2017-19.

Abbreviations:

NSCLC	Non-small cell lung cancer
SNP	single nucleotide polymorphism
GWAS	Genome-Wide Association Study
PLCO	the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial
HLCS	Harvard Lung Cancer Susceptibility study
OS	overall survival
LD	linkage disequilibrium
BFDP	Bayesian false discovery probability
eQTL	expression quantitative trait loci
TCGA	the Cancer Genome Atlas
ROC	receiver operating characteristic
EAF	effect allele frequency
HR	hazards ratio
CI	confidence interval
AUC	area under the curve

References

1. Hu Q, Li B, Garfield D, et al.Prognostic factors for survival in a Chinese population presenting with advanced non-small cell lung cancer with an emphasis on smoking status: A regional, single-institution, retrospective analysis of 4552 patients. Thorac Cancer. 2012; 3: 162–8. [PubMed: 28920299]
2. Houston KA, Mitchell KA, King J, et al.Histologic Lung Cancer Incidence Rates and Trends Vary by Race/Ethnicity and Residential County. J Thorac Oncol. 2018; 13: 497–509. [PubMed: 29360512]
3. Wilson LF, Antonsson A, Green AC, et al.How many cancer cases and deaths are potentially preventable? Estimates for Australia in 2013. Int J Cancer. 2018; 142: 691–701. [PubMed: 28983918]
4. Vrankar M, Stanic K.Long-term survival of locally advanced stage III non-small cell lung cancer patients treated with chemoradiotherapy and perspectives for the treatment with immunotherapy. Radiol Oncol. 2018; 52: 281–8. [PubMed: 30210037]
5. Antonia SJ, Villegas A, Daniel D, et al.Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC. N Engl J Med. 2018; 379: 2342–50. [PubMed: 30280658]
6. Schrodi SJ, Mukherjee S, Shan Y, et al.Genetic-based prediction of disease traits: prediction is very difficult, especially about the future. Front Genet. 2014; 5: 162. [PubMed: 24917882]
7. Amos CI, Wu X, Broderick P, et al.Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet. 2008; 40: 616–22. [PubMed: 18385676]
8. Hung RJ, McKay JD, Gaborieau V, et al.A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature. 2008; 452: 633–7. [PubMed: 18385738]
9. Landi MT, Chatterjee N, Yu K, et al.A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am J Hum Genet. 2009; 85: 679–91. [PubMed: 19836008]
10. McKay JD, Hung RJ, Gaborieau V, et al.Lung cancer susceptibility locus at 5p15.33. Nat Genet. 2008; 40: 1404–6. [PubMed: 18978790]
11. Wang Y, Broderick P, Webb E, et al.Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat Genet. 2008; 40: 1407–9. [PubMed: 18978787]
12. Wang Y, McKay JD, Rafnar T, et al.Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet. 2014; 46: 736–41. [PubMed: 24880342]
13. Freedman ML, Monteiro AN, Gayther SA, et al.Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet. 2011; 43: 513–8. [PubMed: 21614091]
14. Coller HA.Is cancer a metabolic disease? Am J Pathol. 2014; 184: 4–17. [PubMed: 24139946]
15. Hess JA, Khasawneh MK.Cancer metabolism and oxidative stress: Insights into carcinogenesis and chemotherapy via the non-dihydrofolate reductase effects of methotrexate. BBA Clin. 2015; 3: 152–61. [PubMed: 26674389]
16. Kadochi Y, Mori S, Fujiwara-Tani R, et al.Remodeling of energy metabolism by a ketone body and medium-chain fatty acid suppressed the proliferation of CT26 mouse colon cancer cells. Oncol Lett. 2017; 14: 673–80. [PubMed: 28693220]
17. 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: 280–9. [PubMed: 22298838]
18. 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: 1832–9. [PubMed: 16368945]
19. Mailman MD, Feolo M, Jin Y, et al.The NCBI dbGaP database of genotypes and phenotypes. Nat Genet. 2007; 39: 1181–6. [PubMed: 17898773]
20. Tryka KA, Hao L, Sturcke A, et al.NCBI's Database of Genotypes and Phenotypes: dbGaP. Nucleic Acids Res. 2014; 42: D975–9. [PubMed: 24297256]
21. Lappalainen T, Sammeth M, Friedlander MR, et al.Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 2013; 501: 506–11. [PubMed: 24037378]

22. Zhai R, Yu X, Wei Y, et al.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]
23. Aulchenko YS, Ripke S, Isaacs A, et al.GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007; 23: 1294–6. [PubMed: 17384015]
24. Wacholder S, Chanock S, Garcia-Closas M, et al.Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst. 2004; 96: 434–42. [PubMed: 15026468]
25. Ward LD, Kellis M.HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic acids research. 2012; 40: D930–4. [PubMed: 22064851]
26. Xu Z, Taylor JA.SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic acids research. 2009; 37: W600–5. [PubMed: 19417063]
27. Boyle AP, Hong EL, Hariharan M, et al.Annotation of functional variation in personal genomes using RegulomeDB. Genome research. 2012; 22: 1790–7. [PubMed: 22955989]
28. Consortium GT.Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015; 348: 648–60. [PubMed: 25954001]
29. Cancer Genome Atlas Research N.Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014; 511: 543–50. [PubMed: 25079552]
30. Chambless LE, Diao G.Estimation of time-dependent area under the ROC curve for long-term risk prediction. Stat Med. 2006; 25: 3474–86. [PubMed: 16220486]
31. 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: 2592–601. [PubMed: 26757251]
32. Gyorffy B, Surowiak P, Budczies J, et al.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. [PubMed: 24367507]
33. 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: 607–16. [PubMed: 27158780]
34. Imielinski M, Berger AH, Hammerman PS, et al.Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell. 2012; 150: 1107–20. [PubMed: 22980975]
35. Puchalska P, Crawford PA.Multi-dimensional Roles of Ketone Bodies in Fuel Metabolism, Signaling, and Therapeutics. Cell Metab. 2017; 25: 262–84. [PubMed: 28178565]
36. Phan LM, Yeung SC, Lee MH.Cancer metabolic reprogramming: importance, main features, and potentials for precise targeted anti-cancer therapies. Cancer Biol Med. 2014; 11: 1–19. [PubMed: 24738035]
37. Khodadadi S, Sobhani N, Mirshekar S, et al.Tumor Cells Growth and Survival Time with the Ketogenic Diet in Animal Models: A Systematic Review. Int J Prev Med. 2017; 8: 35. [PubMed: 28584617]
38. Puisac B, Ramos M, Arnedo M, et al.Characterization of splice variants of the genes encoding human mitochondrial HMG-CoA lyase and HMG-CoA synthase, the main enzymes of the ketogenesis pathway. Mol Biol Rep. 2012; 39: 4777–85. [PubMed: 21952825]
39. Li P, Wang Y, Zhang L, et al.The Expression Pattern of PLIN2 in Differentiated Adipocytes from Qinchuan Cattle Analysis of Its Protein Structure and Interaction with CGI-58. Int J Mol Sci. 2018; 19.
40. Itabe H, Yamaguchi T, Nimura S, et al.Perilipins: a diversity of intracellular lipid droplet proteins. Lipids Health Dis. 2017; 16: 83. [PubMed: 28454542]
41. Cao Q, Ruan H, Wang K, et al.Overexpression of PLIN2 is a prognostic marker and attenuates tumor progression in clear cell renal cell carcinoma. Int J Oncol. 2018; 53: 137–47. [PubMed: 29749470]
42. Magne J, Aminoff A, Perman Sundelin J, et al.The minor allele of the missense polymorphism Ser251Pro in perilipin 2 (PLIN2) disrupts an alpha-helix, affects lipolysis, and is associated with

reduced plasma triglyceride concentration in humans. FASEB J. 2013; 27: 3090–9. [PubMed: 23603836]

43. Wilborn TW, Lang NP, Smith M, et al.Association of SULT2A1 allelic variants with plasma adrenal androgens and prostate cancer in African American men. The Journal of steroid biochemistry and molecular biology. 2006; 99: 209–14. [PubMed: 16617014]
44. Girard H, Butler LM, Villeneuve L, et al.UGT1A1 and UGT1A9 functional variants, meat intake, and colon cancer, among Caucasians and African-Americans. Mutation research. 2008; 644: 56–63. [PubMed: 18675828]
45. Dellinger RW, Fang JL, Chen G, et al.Importance of UDP-glucuronosyltransferase 1A10 (UGT1A10) in the detoxification of polycyclic aromatic hydrocarbons: decreased glucuronidative activity of the UGT1A10139Lys isoform. Drug metabolism and disposition: the biological fate of chemicals. 2006; 34: 943–9. [PubMed: 16510539]
46. Fang JL, Beland FA, Doerge DR, et al.Characterization of benzo(a)pyrene-trans-7,8-dihydrodiol glucuronidation by human tissue microsomes and overexpressed UDP-glucuronosyltransferase enzymes. Cancer research. 2002; 62: 1978–86. [PubMed: 11929814]
47. Girard H, Thibaudeau J, Court MH, et al.UGT1A1 polymorphisms are important determinants of dietary carcinogen detoxification in the liver. Hepatology. 2005; 42: 448–57. [PubMed: 15986396]
48. Yueh MF, Nguyen N, Famourzadeh M, et al.The contribution of UDP-glucuronosyltransferase 1A9 on CYP1A2-mediated genotoxicity by aromatic and heterocyclic amines. Carcinogenesis. 2001; 22: 943–50. [PubMed: 11375903]
49. Girard H, Court MH, Bernard O, et al.Identification of common polymorphisms in the promoter of the UGT1A9 gene: evidence that UGT1A9 protein and activity levels are strongly genetically controlled in the liver. Pharmacogenetics. 2004; 14: 501–15. [PubMed: 15284532]
50. Levesque E, Delage R, Benoit-Biancamano MO, et al.The impact of UGT1A8, UGT1A9, and UGT2B7 genetic polymorphisms on the pharmacokinetic profile of mycophenolic acid after a single oral dose in healthy volunteers. Clinical pharmacology and therapeutics. 2007; 81: 392–400. [PubMed: 17339869]

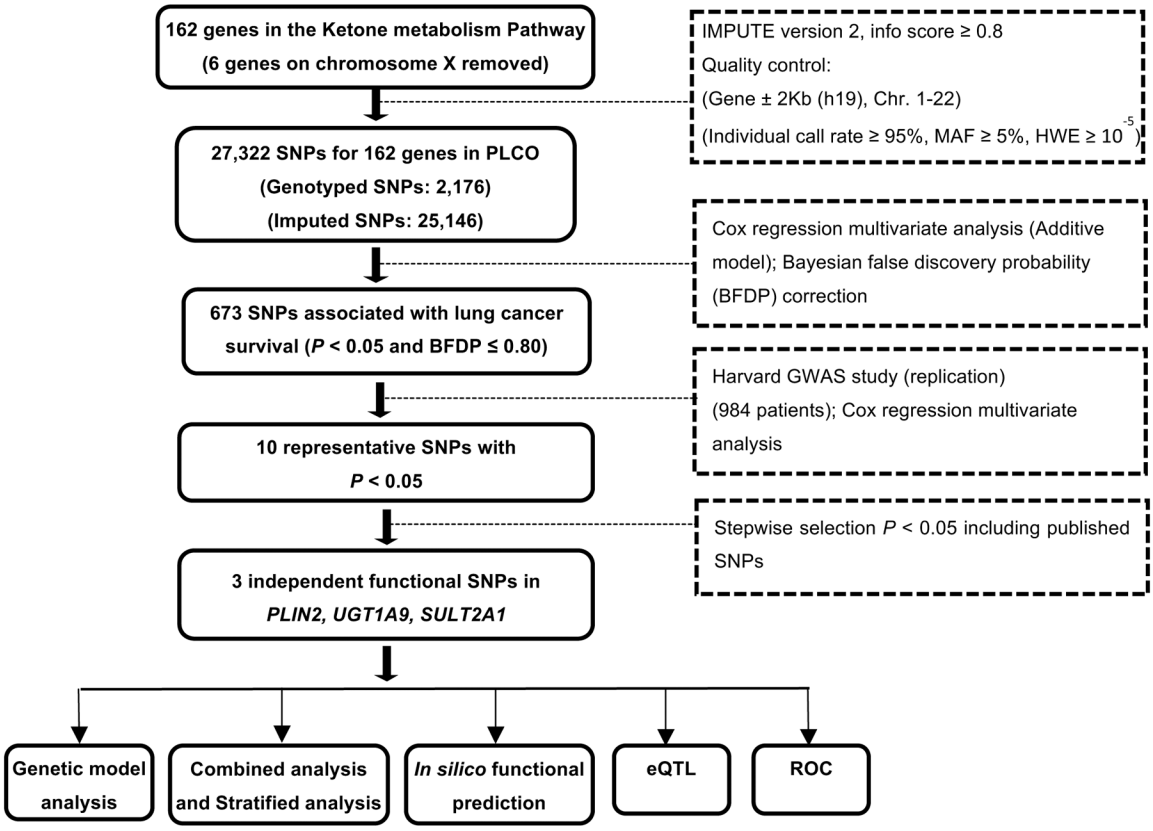

Figure 1.

Study workflow chart

Abbreviations: SNP, single nucleotide polymorphism; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; BFDP, Bayesian false discovery probability; eQTL, expression quantitative trait loci; ROC, receiver operating characteristic.

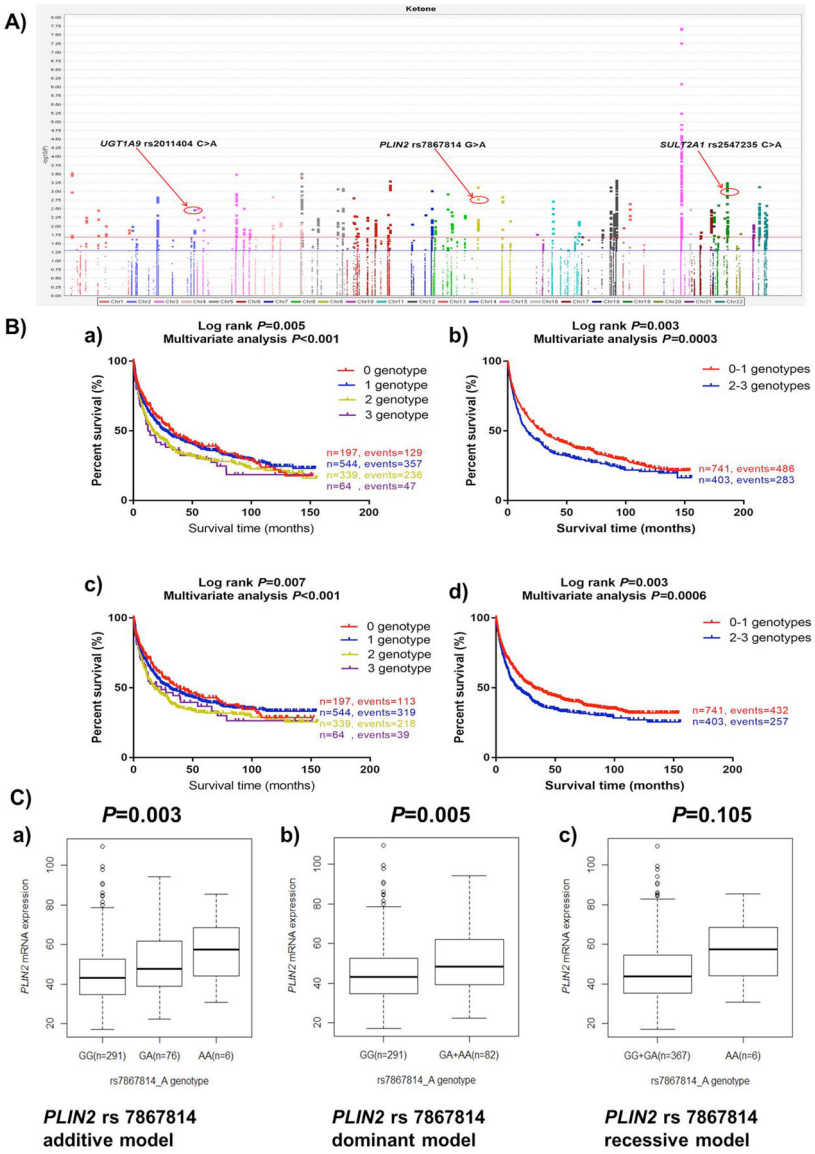

Figure 2.

Functional and survival-associated SNPs. **A.** Manhattan plot of 27,322 SNPs of the Ketone metabolism Pathway genes in the PLCO trial. The statistical values across the autosomes for associations between 27,322 SNPs and overall survival of patients with NSCLC are plotted as −log10 *P* values. The blue horizontal line indicates *P* = 0.05 and the red line indicates BFDP = 0.8. **B.** Kaplan-Meier (KM) survival curves for NSCLC patients of three validated SNPs. (a) OS analysis of combined risk genotypes in the PLCO trial by 0, 1, 2 and 3 unfavorable genotypes (log-rank test for trend: *P*), and (b) by 0–1 and 2–3 unfavorable genotypes (Log-rank test and multivariate analysis: *P*) in the PLCO trial; (c) DSS analysis of combined protective genotypes in the PLCO trial by 0, 1, 2 and 3 unfavorable genotypes (log-rank test for trend: *P*), and (d) by 0–1 and 2–3 unfavorable genotypes (Log-rank test and multivariate analysis: *P*) in the PLCO trial. **C.** eQTL analyses of *PLIN2* rs7867814 genotype and corresponding gene mRNA expression (n=373). All the data were from the

1,000 Genomes Project dataset. (**a**) rs7867814 additive model (*P*=0.003); (**b**) rs7867814 dominant model (*P*=0.005); and (**c**) rs7867814 recessive model (*P*=0.105.
Abbreviations: BFDP, Bayesian false-discovery probability; NSCLC, non-small cell lung cancer; PLCO, Prostate, Lung, Colorectal and Ovarian cancer screening trial.

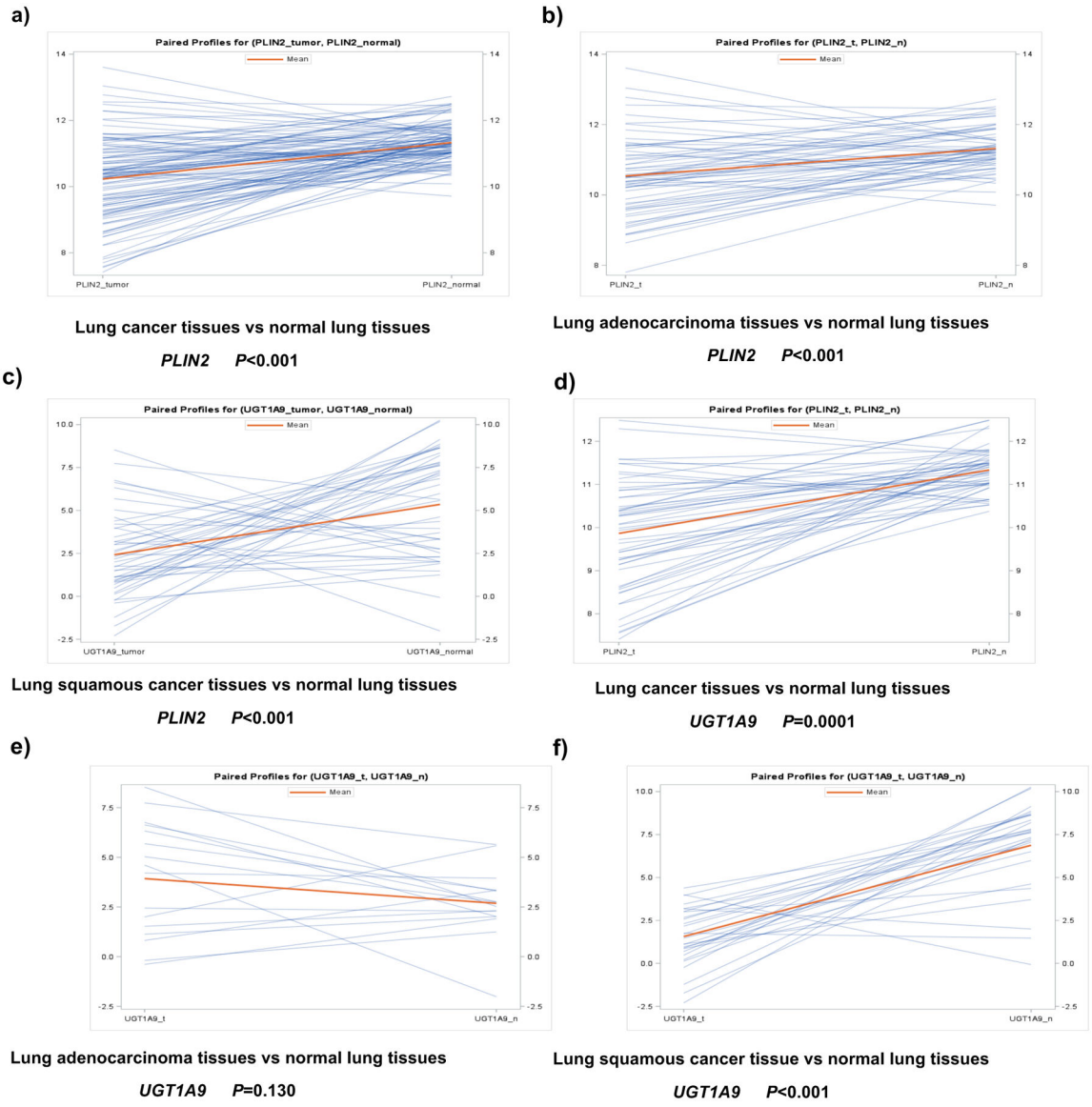

Figure 3.

Comparison of mRNA expression levels of *PLIN2* and *UGT1A9* between lung cancer tissue and adjacent normal lung tissues in the TCGA dataset [1]

.**a.** The *PLIN2* mRNA expression levels in lung cancer tissues were significantly higher than that in normal lung tissues ($P<0.001$); **b.** The *PLIN2* mRNA expression levels in lung adenocarcinoma tissues were significantly higher than that in normal lung tissues ($P<0.001$); **c.** The *PLIN2* mRNA expression levels in lung squamous tissues were significantly higher than that in normal lung tissues ($P<0.001$); **d.** The *UGT1A9* mRNA expression levels in lung cancer tissues were significantly lower than that in normal lung tissues ($P=0.0001$); **e.** The *UGT1A9* mRNA expression levels in lung adenocarcinoma tissues were not higher than that in normal lung tissues ($P=0.130$); **f.** The *UGT1A9* mRNA expression levels in lung squamous tissues were significantly lower than that in normal lung tissues ($P<0.001$); *PLIN2*_t= lung cancer tissues; *PLIN2*_n= adjacent normal lung tissues; *UGT1A9*_t= lung

cancer tissues; *UGT1A9*_n= adjacent normal lung tissues. 1 Including 51 pairs of lung adenocarcinoma tissues and 60 pairs of lung squamous tissues; in total, there were 111 pairs of tissues in the analysis.

Table 1.

Combined analysis of the ten significant SNPs in both PLCO and HLCS datasets

SNP	Allele[a]	Gene	PLCO (n=1,185)				HLCS (n=984)			Combined analysis			
			EAF	BFDP	FDR	HR (95% CI)[b]	P_P[b]	EAF	HR(95% CI)[c]	P[c]	HR(95% CI)[d]	P[d]	P_{het}[e]	I^2
rs786781	G>A	*PLIN2*	0.12	0.31	0.24	1.27 (1.09–1.47)	0.002	0.16	1.17 (1.00–1.36)	0.048	1.22 (1.09–1.36)	0.0003	0.439	0
rs731581	C>G	*PRKAG2*	0.12	0.58	0.48	1.23 (1.06–1.42)	0.006	0.12	1.27 (1.06–1.52)	0.008	1.25 (1.11–1.40)	0.0001	0.775	0
rs254723	C>T	SULT2A1	0.18	0.29	0.19	1.25 (1.10–1.43)	0.0008	0.20	1.17 (1.01–1.35)	0.036	1.21 (1.10–1.33)	0.0001	0.486	0
rs201140	C>T	*UGT1A9*	0.17	0.57	0.37	0.82 (0.71–0.94)	0.003	0.17	0.82 (0.70–0.96)	0.011	0.82 (0.74–0.91)	0.0002	0.979	0
rs780232	T>A	*RORA*	0.15	0.67	0.51	0.83 (0.72–0.95)	0.009	0.13	0.81 (0.68–0.96)	0.016	0.82 (0.74–0.92)	0.0004	0.821	0
rs763828	G>A	*RORA*	0.15	0.67	0.51	0.83 (0.72–0.95)	0.009	0.13	0.81 (0.68–0.96)	0.017	0.82 (0.74–0.92)	0.0004	0.831	0
rs697066	A>C	*PRKAG2*	0.20	0.77	0.54	1.17 (1.03–1.32)	0.016	0.21	1.16 (1.01–1.33)	0.037	1.16 (1.06–1.28)	0.001	0.908	0
rs697072	G>C	*PRKAG2*	0.20	0.77	0.54	1.17 (1.03–1.32)	0.016	0.21	1.16 (1.01–1.33)	0.040	1.16 (1.06–1.28)	0.001	0.895	0
rs697036	A>G	*PRKAG2*	0.20	0.77	0.54	1.17 (1.03–1.32)	0.016	0.21	1.15 (1.00–1.33)	0.043	1.16 (1.06–1.27)	0.001	0.876	0
rs102778	T>C	*PRKAG2*	0.19	0.77	0.54	1.17 (1.03–1.32)	0.017	0.20	1.15 (1.00–1.32)	0.046	1.16 (1.06–1.27)	0.001	0.848	0

[a] Effect/reference 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 an additive model;

[e] *P* value for heterogeneity by Cochrane's Q test.

Abbreviations: PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; HLCS = Harvard Lung Cancer Susceptibility study; SNPs = Single nucleotide polymorphisms; EAF = Effect allele frequency; HR = Hazards ratio; CI = Confidence interval.

Table 2.

Predictors of OS obtained from the final Cox hazards regression analysis in the PLCO dataset

Variables	Category	Frequency	HR (95% CI)[a]	P[a]	HR (95% CI)[b]	P[b]
Age	Continuous	1185	1.03 (1.02–1.05)	<0.0001	1.04 (1.02–1.05)	<0.0001
Sex	Male	698	1.00		1.00	
	Female	487	0.79 (0.68–0.93)	0.003	0.77 (0.66–0.91)	0.002
Smoking status	Never	115	1.00		1.00	
	Current	423	1.75 (1.30–2.35)	0.0002	1.93 (1.43–2.61)	<0.0001
	Former	647	1.77 (1.34–2.34)	<0.0001	1.94 (1.46–2.59)	<0.0001
Histology	AD	577	1.00		1.00	
	SC	285	1.25 (1.03–1.51)	0.023	1.28 (1.05–1.56)	0.013
	Others	323	1.31 (1.10–1.56)	0.002	1.38 (1.15–1.66)	0.0005
Stage	I-IIIA	655	1.00		1.00	
	IIIB-IV	528	3.21 (2.62–3.91)	<0.0001	3.29 (2.68–4.03)	<0.0001
Chemotherapy	No	639	1.00		1.00	
	Yes	538	0.54 (0.45–0.65)	<0.0001	0.56 (0.46–0.67)	<0.0001
Radiotherapy	No	762	1.00		1.00	
	Yes	415	0.99 (0.84–1.16)	0.871	0.99 (0.83–1.17)	0.889
Surgery	No	637	1.00		1.00	
	Yes	540	0.22 (0.17–0.28)	<0.0001	0.20 (0.16–0.27)	<0.0001
PLIN2 rs7867814 G>A	GG/GA/AA	916/248/21	1.24 (1.07–1.44)	0.005	1.27 (1.09–1.48)	0.003
SULT2A1 rs2547235 C>T	CC/CT/TT	805/338/42	1.23 (1.07–1.40)	0.003	1.22 (1.07–1.40)	0.004
UGT1A9 rs2011404 C>T	CC/CT/TT	793/332/29	0.83 (0.73–0.96)	0.010	0.85 (0.73–0.98)	0.025

[a]The final analysis included age, sex, smoking status, tumor stage, tumor histology, chemotherapy, radiotherapy, surgery, top four principal components and three new validated SNPs (*SULT2A1* rs2547235, *UGT1A9* rs2011404 and *PLIN2* rs7867814 in an additive model);

[b]Fifteen published SNPs were used for post-stepwise adjustment. Five SNPs were reported in previous publication (PMID: 27557513); One SNP was reported in the previous publication (PMID: 29978465); Two SNPs were reported in the previous publication (PMID: 30259978); Two SNPs were reported in the previous publication (PMID: 26757251); Three SNPs were reported in the previous publication (PMID: 30650190); Two SNPs were reported in the previous publication (PMID: 30989732);

Abbreviations: OS = Overall survival; PLCO = Prostate, Lung, Colorectal and Ovarian cancer trial; HR = Hazards ratio; CI = Confidence interval.

Table 3.

Associations of the three independent and validated SNPs in the Ketone metabolism pathway genes with OS and DSS of NSCLC in the PLCO dataset

	OS Multivariate analysis[a]				DSS Multivariable analysis[a]			
Genotype	**Frequency (n=1,144)**	**Death (%)**	**HR (95% CI)**	***P***	**Frequency (n=1,144)**	**Death (%)**	**HR (95% CI)**	***P***
***PLIN2* rs7867814 G>A**								
GG	908	604 (66.5)	1.00		908	544 (59.9)	1.00	
GA	246	168 (68.3)	1.27 (1.07–1.52)	0.006	246	148 (60.2)	1.25 (1.04–1.51)	0.017
AA	21	17 (81.0)	1.58 (0.97–2.59)	0.069	21	17 (81.0)	1.65 (1.01–2.72)	0.048
Trend test				0.002				0.003
GA+AA	267	185 (69.3)	1.30 (1.10–1.53)	0.003	267	165 (61.8)	1.28 (1.07–1.53)	0.006
***SULT2A1* rs2547235 C>T**								
CC	795	530 (66.7)	1.00		795	474 (59.6)	1.00	
CT	338	226 (66.9)	1.15 (0.98–1.35)	0.084	338	204 (64.0)	1.13 (0.96–1.34)	0.152
TT	42	33 (78.6)	2.03 (1.42–2.92)	0.0001	42	31 (73.8)	2.06 (1.42–2.99)	0.0001
Trend test				0.0008				0.002
CT+TT	380	259 (68.2)	1.22 (1.05–1.42)	0.011	380	235 (61.8)	1.20(1.02–1.41)	0.025
***UGT1A9* rs2011404 C>T**								
CC	783	533 (68.1)	1.00		783	479 (61.2)	1.00	
CT	332	218 (65.7)	0.85 (0.73–1.00)	0.049	332	194 (58.4)	0.84 (0.71–1.00)	0.048
TT	29	18 (62.1)	0.57 (0.35–0.92)	0.022	29	16 (55.2)	0.57 (0.34–0.94)	0.029
Trend test				0.005				0.006
CT+TT	361	236 (65.4)	0.82 (0.70–0.96)	0.015	361	210 (58.2)	0.82 (0.69–0.96)	0.016
NUG[b]								
0	197	129 (65.5)	1.00		197	113 (57.4)	1.00	
1	544	357 (65.6)	1.26 (1.02–1.55)	0.029	544	319 (58.6)	1.27 (1.02–1.58)	0.035
2	339	236 (69.6)	1.51 (1.21–1.88)	0.0003	339	218 (64.3)	1.54 (1.22–1.94)	0.0003
3	64	47 (73.4)	1.86 (1.32–2.63)	0.0004	64	39 (60.9)	1.76 (1.21–2.55)	0.003
Trend test				<0.001				<0.001
NUG[b]								
0–1	741	486 (65.6)			741	432 (58.3)		
2–3	403	283 (70.2)	1.32 (1.13–1.53)	0.0003	403	257 (63.8)	1.32 (1.13–1.54)	0.0006

[a]Multivariate Cox hazards regression analyses were adjusted for age, sex, smoking, stage, histology, chemotherapy, radiotherapy, surgery, and top four principal components in the PLCO dataset; there were 31 subjects with missing information for genotype and 10 with missing phenotype data.

[b]NUG=number of unfavorable genotypes; NUG were *PLIN2* rs7867814 GA+AA, *SULT2A1* rs2547235 CT+TT and *UGT1A9* rs2011404 CC+CT;

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

Abbreviations: SNPs = Single nucleotide polymorphisms; OS = Overall survival; NSCLC = Non-small cell lung cancer; PLCO = Prostate, Lung, Colorectal and Ovarian cancer trial; NPG = Number of protective genotypes; HR = Hazards ratio; CI = Confidence interval; N/A = Not applicable