Original Article

Novel genetic variants in genes of the Fc gamma receptor-mediated phagocytosis pathway predict non-small cell lung cancer survival

Danwen Qian[1,2,3], Hongliang Liu[2,3], Lingling Zhao[2,3], Xiaomeng Wang[2,3], Sheng Luo[4], Patricia G. Moorman[2,5], Edward F. Patz Jr[2,6], Li Su[7], Sipeng Shen[7], David C. Christiani[7,8], Qingyi Wei[2,3,9]

[1]Cancer Institute, Fudan University Shanghai Cancer Center; Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China; [2]Duke Cancer Institute, Duke University Medical Center, Durham, NC, USA; [3]Department of Population Health Sciences, [4]Department of Biostatistics and Bioinformatics, [5]Department of Family Medicine and Community Health, Duke University School of Medicine, Durham, NC, USA; [6]Department of Radiology, Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, NC, USA; [7]Department of Environmental Health and Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA; [8]Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; [9]Department of Medicine, Duke University School of Medicine, Durham, NC, USA

Contributions: (I) Conception and design: D Qian, Q Wei; (II) Administrative support: Q Wei, H Liu; (III) Provision of study materials or patients: Q Wei, H Liu, L Su, S Shen, DC Christiani; (IV) Collection and assembly of data: D Qian, H Liu, L Su, S Shen, DC Christiani, Q Wei; (V) Data analysis and interpretation: D Qian, H Liu, L Zhao, X Wang, S Luo, DC Christiani, Q Wei; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

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

Background: Both antibody-dependent cellular cytotoxicity and phagocytosis activate innate immunity, and the Fc gamma receptor (FCGR)-mediated phagocytosis is an integral part of the process. We assessed associations between single-nucleotide polymorphisms (SNPs) in FCGR-related genes and survival of patients with non-small cell lung cancer (NSCLC).

Methods: We evaluated associations between 24,734 (SNPs) in 97 FCGR-related genes and survival of 1,185 patients with NSCLC using a published genome-wide association study (GWAS) dataset from the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial and validated the results in another independent dataset of 894 NSCLC patients.

Results: In the single-locus analysis with Bayesian false discovery probability (BFDP) for multiple testing correction, we found 1,084 SNPs to be significantly associated overall survival (OS) ($P<0.050$ and BFDP ≤0.80), of which two independent SNPs (*PLCG2* rs9673682 T>G and *PLPP1* rs115613985 T>A) were further validated in another GWAS dataset of 894 patients from the Harvard Lung Cancer Susceptibility (HLCS) Study, with combined allelic hazards ratios for OS of 0.87 [95% confidence interval (CI): 0.81–0.94 and $P=5.90\times10^{-4}$] and 1.18 (95% CI: 1.08–1.29 and 1.32×10^{-4}, respectively). Expression quantitative trait loci analysis showed that the rs9673682 G allele was significantly correlated with increased mRNA expression levels of *PLCG2* in 373 transformed lymphoblastoid cell-lines ($P=7.20\times10^{-5}$). Additional evidence from differential expression analysis further supported a tumor-suppressive effect of *PLCG2* on OS of patients with lung cancer, with lower mRNA expression levels in both lung squamous carcinoma and adenocarcinoma than in adjacent normal tissues.

Conclusions: Genetic variants in *PLCG2* of the FCGR-mediated phagocytosis pathway may be promising predictors of NSCLC survival, possibly through modulating gene expression, but additional investigation of the molecular mechanisms of *PLPP1* rs115613985 is warranted.

Keywords: Non-small cell lung cancer (NSCLC); Fc gamma receptor (FCGR); genome-wide association study (GWAS); single-nucleotide polymorphism (SNP); survival

Submitted Aug 11, 2019. Accepted for publication Mar 17, 2020.
doi: 10.21037/tlcr-19-318
View this article at: http://dx.doi.org/10.21037/tlcr-19-318

Introduction

Global cancer statistics in 2018 indicated that lung cancer was the most common cancer and the leading cause of cancer-related deaths, with an estimated 2,093,876 new cases and 1,761,007 cancer deaths worldwide (1). Non-small cell lung cancer (NSCLC) accounts for 85% of lung cancer cases, which mainly includes squamous cell carcinomas and adenocarcinomas. Recently developed immunotherapy has become a clinical research focus for the treatment of lung cancer, especially clinical application of the check-point inhibitors, such as PD-1/PD-L1 inhibitors (2). However, predicting the response to these therapies, has been challenging, because individuals' response to the same treatment remains heterogeneous, suggesting that host factors including genetic variants also play an important role in variable treatment response. For example, single-nucleotide polymorphisms (SNPs) have been shown to be associated with both short-term response and long-term prognosis of cancer patients (3-5). Therefore, identifying the role of these genetic variants may help us generate a better model to predict outcomes and optimize therapies in the treatment of NSCLC patients.

Antibodies, produced in response to foreign antigens, are the key components that are responsible for innate and adaptive immunity. For example, immunoglobulin G (IgG) antibodies recognize tumor cells and form immune complexes to mediate effector activities by interactions between the Fc domains of IgG antibodies and the Type I and type II Fc gamma receptors (FCGRs) (6,7). In addition, the FCGR-mediated phagocytosis plays a central role in tumor immunity by tyrosine phosphorylation of multiple immunoreceptors, resulting in phagocytosis of target tumor cells (8). Furthermore, antibody-coated tumor cells interact with FCGRs and engage monocytes, macrophages or natural killer cells, leading to phagocytosis or cell-mediated cytotoxicity of tumor cells (9). Recent research has showed that antibody-dependent cellular phagocytosis (ADCP) is an important immune mechanism by which antibodies attack foreign antigens. For example, FCGR2A (CD32a) is thought to be a dominant player in the induction of ADCP by macrophages (10).

Several studies have reported that SNPs in the coding regions of some FCGR-related genes may have an effect on clinical response to mAbs. For instance, genetic variants in *FCGR2A* and *FCGR3A* predicted the response to trastuzumab in both metastatic HER2-positive breast cancer patients and gastric cancer patients (11,12), and the *FCGR2A* H/H genotype was associated with a better overall survival (OS) in cetuximab-treated colorectal cancer patients with a wild-type *KRAS* (13). But there are no reports about the effect of SNPs in the FCGR-related genes on survival of NSCLC. Thus, we hypothesize that genetic variants in the FCGR-mediated phagocytosis pathway genes are associated with a heterogeneous anti-tumor immune response, resulting in variable survival of NSCLC patients. In the present study, we tested this hypothesis by using publicly available genome-wide association study (GWAS) datasets to evaluate associations between genetic variants of genes in the FCGR-mediated phagocytosis pathway and NSCLC survival.

Methods

Study populations

We used one GWAS dataset for 1,185 NSCLC patients from the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial as the discovery and another for 984 NSCLC patients from the Harvard Lung Cancer Susceptibility (HLCS) study as the validation. The PLCO dataset had additional data on OS and disease-free survival (DSS) but not progression-free survival (PFS), while the HLCS dataset had data only on OS for analysis. The access to the PLCO GWAS dataset was approved by the dbGAP from the National Cancer Institute (the approval number: PLCO-95 and Project #6404), for which genomic DNA samples extracted from the whole blood were genotyped with Illumina HumanHap240Sv1.0, HumanHap300v1.1 and HumanHap550v3.0 (dbGaP accession: phs000093.v2.p2 and phs000336.v1.p1) (14,15), while the DNA samples from HLCS patients extracted from the whole blood were genotyped with Illumina Humanhap610-Quad arrays. Details of data collection and participants' characteristics in both GWAS datasets have been described elsewhere (16,17). The two original studies were approved by the institutional review boards of the National Cancer Institute and Massachusetts General Hospital, respectively,

with a written informed consent obtained from each participant. The distributions of population characteristics in the PLCO and HLCS studies are shown in *Table S1*. Because all the data used in the present study were de-identified and previously collected, this exempt research was approved by Duke Internal Review Board (Pro00054575).

Gene and SNP selection

In total, 97 genes in the FCGR-mediated phagocytosis pathway were selected as the candidate gene-set after excluding duplicate genes and pseudo genes, according to the online datasets the Molecular Signatures Database (http://software.broadinstitute.org/gsea/msigdb/index.jsp) and the Human Biological Pathway Unification Database (http://pathcards.genecards.org) by the keyword "Fc gamma R" (*Table S2*). We imputed additional SNPs for these genes and their ±500 kb flanking buffer regions with IMPUTE2 and the 1,000 Genomes Project data (phase 3). SNPs in these genes and within their ±2 kb flanking regions were selected for further analysis based on the following criteria: a minor allele frequency ≥0.05, a genotyping rate ≥95%, a Hardy-Weinberg equilibrium P value $\geq 1\times10^{-5}$, and imputation infor score ≥0.8.

In silico functional analysis

We used three online bioinformatics tools to identify potential functional SNPs, including RegulomeDB (http://regulomedb.org/), HaploReg (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) and SNPinfo (https://snpinfo.niehs.nih.gov/snpinfo/snpfunc.html). We further performed expression quantitative trait loci (eQTL) analysis to assess correlations between SNPs and mRNA expression levels. These mRNA expression data were obtained from lymphoblastoid cell lines derived from 373 European descendants included in the 1,000 Genomes Project (18) as well as from the whole blood and normal lung tissues in the genotype-tissue expression (GTEx) project (19). In addition, we compared the differences in mRNA expression levels for the paired tumor and adjacent normal tissues using the data from the Cancer Genome Atlas (TCGA) database (dbGaP Study Accession: phs000178.v9.p8) (20).

Statistical analysis

In the PLCO dataset, multivariate Cox proportional hazards regression analysis was used to evaluate associations between each of the SNPs and OS (in an additive model) with adjustment for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery and the first four principal components of the population structures. In the single-locus analysis, we estimated the hazards ratio (HR) and its 95% confidence interval (CI) by using the GenABEL package of R software (21). The false discovery rate (FDR) with a cut-off value of 0.20 was used to reduce the probability of false-positives findings (22), and the recommended Bayesian false-discovery probability (BFDP) with a cutoff value of 0.800 was also used for multiple test corrections under the consideration that the majority of SNPs were imputed with a high level of linkage disequilibrium (LD) among the SNPs under investigation (23). Then, SNPs with P≤0.005 were chosen for validation using the HLCS GWAS dataset. Meta-analysis was performed to combine the results of both discovery and validation datasets. If the Cochran's Q-test P>0.100 and the heterogeneity statistic (I^2) <50%, a fixed-effects model was applied; otherwise, a random-effects model was employed. In the PLCO dataset, different genetic models were performed to explore the associations between the validated SNPs and OS/DSS, and Kaplan-Meier curves and log-rank tests were used to compare the effects of genotypes on OS/DSS. To observe the combined effect of significant SNPs, we used the number of risk genotypes to estimate cumulative effects of the identified SNPs. Finally, we performed the receiver operating characteristic (ROC) curve and time-dependent ROC analysis with the time ROC package of R software to illustrate the prediction ability of the genetic model (24). In the eQTL analysis, we calculated correlations between SNPs and mRNA expression levels by using a general linear regression model. Additionally, we compared the mRNA expression levels between paired tumor and adjacent normal lung tissues available in the TCGA database by the paired *t* test. Statistical analyses were performed using PLINK (version 1.9), SAS software (version 9.4; SAS Institute, Cary, NC, USA) and R software (version 3.5.1). The Manhattan plots were generated by Haploview v4.2 and regional association plots were constructed by LocusZoom (http://locuszoom.sph.umich.edu/locuszoom/).

Results

Associations between SNPs in the FCGR-mediated phagocytosis pathway genes and NSCLC OS in both PLCO and HLCS datasets

Since only OS was available in both PLCO and HLCS

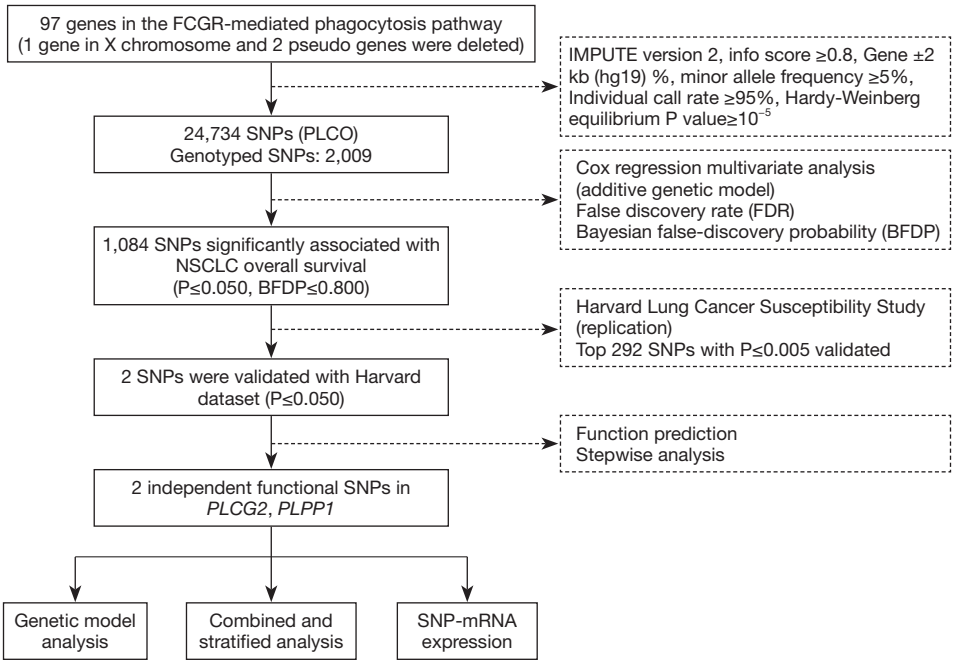

Figure 1 The flowchart of study. FCGR, Fc gamma receptors; SNP, single-nucleotide polymorphism; PLCO, Prostate, Lung, Colorectal and Ovarian cancer screening trial; NSCLC, non-small cell lung cancer; FDR, false discovery rate; BFDP, Bayesian false-discovery probability.

datasets, we used OS to identified independent SNPs as NSCLC survival predictors. Basic demographic and clinical characteristics of the 1,185 NSCLC patients in the PLCO discovery dataset have been described elsewhere (25), with the results of multivariate Cox regression analysis adjusted for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery and first four principal components (detailed information on principal components was provided in *Table S3*). As shown in the flowchart (*Figure 1*), after imputation and quality control, we used 24,734 SNPs (including 2,009 genotyped and 22,725 imputed SNPs) in the multivariate Cox regression analysis. As a result, we identified 1,084 SNPs to be significantly associated with NSCLC OS after multiple test correction by BFDP ≤0.800. A Manhattan plot showing the associations between the SNPs and NSCLC OS is presented in *Figure S1*. Only the top 292 SNPs with P≤0.005 in multivariate Cox regression analysis in the PLCO datasets were included for further validation by the HLCS dataset. Two SNPs in intron regions of two different genes (i.e., rs9673682 in *PLCG2* and rs115613985 in *PLPP1*, both of which were imputed) were validated after multivariate Cox regression analysis with adjustment for age, sex, histology, stage, smoking status, chemotherapy, radiotherapy, surgery and first three principal components. In the subsequent combined-analysis of these two datasets, the *PLCG2* rs9673682 G allele was found to be associated with a better NSCLC OS (HR 0.87, 95% CI: 0.81–0.94, $P=5.90\times10^{-4}$), while the *PLPP1* rs115613985 A allele was associated with a poorer NSCLC OS (HR 1.18, 95% CI: 1.08–1.29, $P=1.32\times10^{-4}$), and no heterogeneity was observed between the two datasets (*Table 1*).

Identification of potentially functional and independent SNPs as NSCLC survival predictors

First, we used three bioinformatics tools (i.e., SNPinfo, RegulomeDB and HaploReg) to search for potential functions of the two SNPs and found that both SNPs were located in the intron regions of their genes with some considerable evidence of functions. As indicated by RegulomeDB, *PLCG2* rs9673682 and *PLPP1* rs115613985 had a score of 4 and 6, respectively (*Table S4*), but neither had functions based on the SNPinfo. In HaploReg, however, *PLCG2* rs9673682 and *PLPP1* rs115613985 were predicted to be located in histone marks, DNAse or motifs, which may have an effect on transcriptional activity. Therefore, these two SNPs were included in the further stepwise Cox

Table 1 Combined-analysis of two validated SNPs using two previously published NSCLC GWAS datasets

SNP	Allele	Gene	PLCO (n=1,185)					HLCS (n=984)			Combined-analysis			
			EAF	HR (95% CI)[1]	P[1]	FDR[2]	BFDP[2]	EAF	HR (95% CI)[3]	P[3]	P_{het}[4]	I^2	HR (95% CI)[5]	P[5]
rs9673682	T>G	*PLCG2*	0.34	0.86 (0.77–0.95)	0.004	0.493	0.457	0.34	0.89 (0.79–1.00)	0.050	0.705	0	0.87 (0.81–0.94)	5.90×10^{-4}
rs115613985	T>A	*PLPP1*	0.38	1.16 (1.05–1.29)	0.005	0.493	0.604	0.35	1.22 (1.05–1.42)	0.008	0.560	0	1.18 (1.08–1.29)	1.32×10^{-4}

[1], obtained from an additive genetic model with adjustment for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3, and PC4 in the PLCO datasets; [2], FDR and BFDP were available in the PLCO dataset because the HLCS study provided only the summary data; [3], obtained from an additive genetic model with adjustment for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, and PC3 in the HLCS dataset; [4], P_{het}: P value for heterogeneity by Cochrane's Q test; [5], meta-analysis using a fixed-effects model. SNP, single-nucleotide polymorphism; NSCLC, non-small cell lung cancer; GWAS, genome-wide association study; PLCO, Prostate, Lung, Colorectal and Ovarian cancer screening trial; HLCS, Harvard Lung Cancer Susceptibility; EAF, effect allele frequency; HR, hazards ratio; CI, confidence interval; FDR, false discovery rate; BFDP, Bayesian false-discovery probability.

Table 2 Predictors of OS obtained from stepwise multivariate Cox regression analysis of selected variables in the PLCO Trial

Variables[1]	Category	Frequency[3]	HR (95% CI)	P
Age	Continuous	1,162	1.03 (1.02–1.05)	<0.001
Sex	Male	688	1.00	
	Female	474	0.80 (0.69–0.93)	0.005
Smoking status	Never	114	1.00	
	Former	638	1.64 (1.25–2.16)	<0.001
	Current	410	1.65 (1.23–2.20)	0.001
Histology	AD	570	1.00	
	SC	279	1.14 (0.94–1.37)	0.185
	others	313	1.32 (1.11–1.57)	0.001
Stage	I–IIIA	646	1.00	
	IIIB–IV	516	2.80 (2.31–3.40)	<0.001
Chemotherapy	No	630	1.00	
	Yes	532	0.58 (0.49–0.69)	<0.001
Radiotherapy	No	752	1.00	
	Yes	410	0.95 (0.81–1.12)	0.553
Surgery	No	628	1.00	
	Yes	534	0.21 (0.16–0.27)	<0.001
PLCG2 rs9673682[2]	TT/TG/GG	513/518/131	0.85 (0.77–0.95)	0.003
PLPP1 rs115613985[2]	TT/TA/AA	451/545/166	1.17 (1.06–1.30)	0.003

[1], stepwise analysis in the final model including age, sex, smoking status, tumor stage, histology, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3, PC4 and two SNPs; [2], rs9673682 and rs115613985 in an additive genetic model. The leftmost genotype was used as the reference; [3], there were 23 samples with missing date and thus were excluded. OS, overall survival; PLCO, Prostate, Lung, Colorectal and Ovarian cancer screening trial; HR, hazards ratio; CI, confidence interval; AD, adenocarcinoma; SC, squamous cell carcinoma.

regression analysis. As the HLCS dataset only provided the summary data, we had to use the PLCO dataset to perform the multivariate stepwise Cox regression analysis with adjustment for clinical variables and the first four principal components. The results suggest that both of the SNPs were significant and independent predictors of NSCLC OS (*Table 2*). In addition, we also conducted an independent test for all previously published SNPs identified in the present study populations in the PLCO dataset; as a result, both of these SNPs remained associated with NSCLC OS (HR 0.89,

Table 3 Associations between two independent and functional SNPs and survival of NSCLC in the PLCO Trial

Genotype	Number[1]	OS[2]			DSS[2]		
		Death (%)	HR (95% CI)	P	Death (%)	HR (95% CI)	P
PLCG2 rs9673682 T>G							
TT	513	350 (68.23)	1.00	–	324 (63.16)	1.00	–
TG	518	346 (66.80)	0.90 (0.77–1.04)	0.163	308 (59.46)	0.88 (0.75–1.03)	0.112
GG	131	86 (65.65)	0.70 (0.55–0.90)	0.004	77 (58.78)	0.70 (0.54–0.90)	0.006
Trend	–	–	–	0.005	–	–	0.004
TG + GG	649	432 (66.56)	0.85 (0.74–0.99)	0.030	385 (59.32)	0.84 (0.72–0.97)	0.021
PLPP1 rs115613985 T>A							
TT	451	289 (64.08)	1.00	–	260 (57.65)	1.00	–
TA	545	375 (68.81)	1.20 (1.03–1.40)	0.023	337 (61.83)	1.19 (1.01–1.41)	0.036
AA	166	118 (71.08)	1.33 (1.06–1.65)	0.012	105 (63.25)	1.32 (1.05–1.67)	0.020
Trend	–	–	–	0.005	–	–	0.009
TA + AA	711	493 (69.34)	1.23 (1.06–1.42)	0.008	442 (62.17)	1.22 (1.04–1.43)	0.013
Number of risk genotypes[3]							
0	236	152 (64.41)	1.00	–	136 (57.63)	1.00	–
1	628	417 (66.40)	1.27 (1.05–1.54)	0.013	371 (59.08)	1.26 (1.03–1.54)	0.023
2	298	213 (71.48)	1.45 (1.18–1.80)	0.001	195 (65.44)	1.48 (1.18–1.84)	0.001
Trend	–	–	–	0.001	–	–	0.001
0–1	864	569 (65.86)	1.00	–	507 (58.68)	1.00	–
2	298	213 (71.48)	1.22 (1.04–1.44)	0.013	195 (65.44)	1.25 (1.06–1.48)	0.009

[1], there were 23 samples with missing date and thus were excluded; [2], adjusted for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery and principal components; [3], risk genotypes were *PLCG2* rs9673682 TT and *PLPP1* rs115613985 TA + AA. SNP, single-nucleotide polymorphism; NSCLC, non-small cell lung cancer; PLCO, Prostate, Lung, Colorectal and Ovarian cancer screening trial; OS, overall survival; DSS, disease-specific survival; HR, hazards ratio; CI, confidence interval.

95% CI: 0.80–1.00, and P=0.040 for *PLCG2* rs9673682 and HR 1.23, 95% CI: 1.10–1.37 and P<0.001 for *PLPP1* rs115613985), independent of other previously published SNPs (*Table S5*).

In the PLCO dataset, multivariate Cox regression analyses with different genetic models were also used to evaluate the effect of each SNP on risk of death. The *PLCG2* rs9673682 G allele was significantly associated with a decrease in risk of death [for OS: HR 0.90, 95% CI: 0.77–1.04 for TG, 0.70 (0.55–0.90) for GG compared with TT, P_{trend}=0.005; for DSS: 0.88 (0.75–1.03) for TG, 0.70 (0.54–0.90) for GG compared with TT, P_{trend}=0.004], while the *PLPP1* rs115613985 A allele was significantly associated with an increase in risk of death [for OS: 1.20 (1.03–1.40) for TA, 1.33 (1.06–1.65) for AA compared with TT, P_{trend}=0.005; for DSS: 1.19 (1.01–1.41) for TA; 1.32 (1.05–1.67) for AA compared with TT, P_{trend}=0.009; *Table 3*]. Kaplan-Meier curves of the associations between these SNPs and NSCLC OS and DSS are presented in *Figure S2*, and the regional association plot of each SNP is shown in *Figure S3*. In a dominant genetic model, *PLCG2* rs9673682 TG + GG genotypes were consistently associated with better OS (HR 0.85, 95% CI: 0.74–0.99; P=0.030) and DSS (HR 0.84, 95% CI: 0.72–0.97; P=0.021), while *PLPP1* rs115613985 TA + AA were associated with poorer OS (HR 1.23, 95% CI: 1.06–1.42; P=0.008) and DSS (HR 1.22, 95% CI: 1.04–1.43; P=0.013) (*Table 3*).

Combined and stratified analyses of the two independent and functional SNPs in the PLCO dataset

To better estimate the joint effect of these two SNPs on risk of death, we combined their risk genotypes (i.e., rs9673682 TT and rs115613985 TA + AA) into a genetic score as the

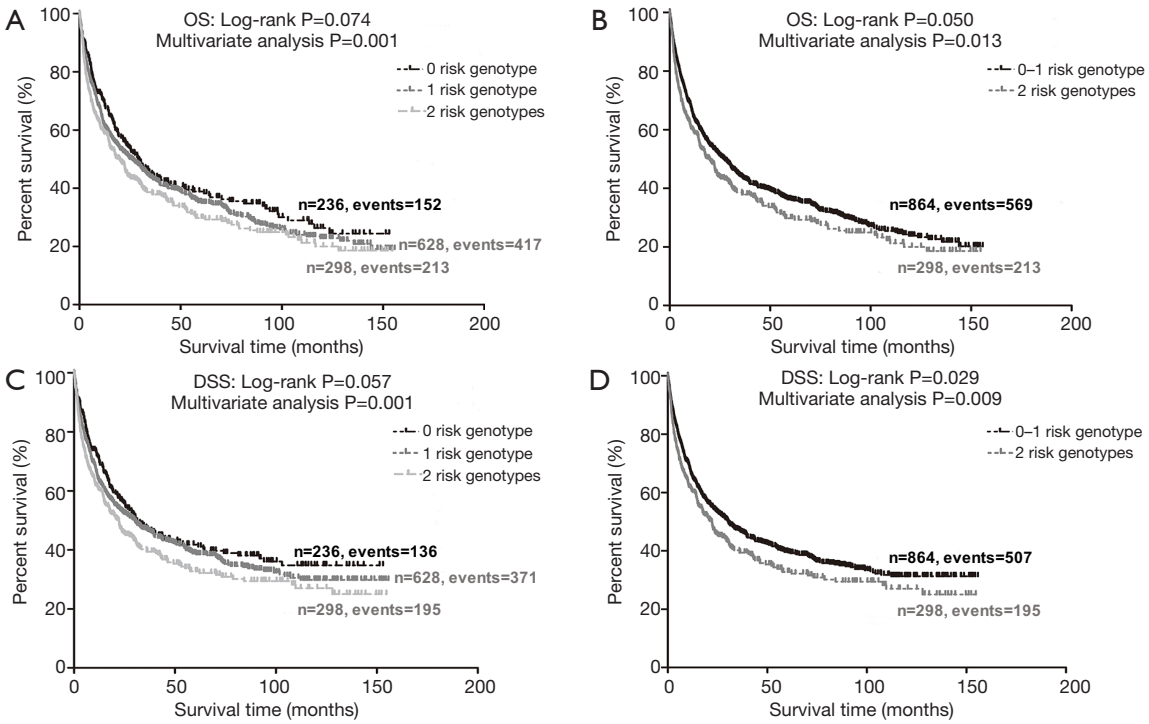

Figure 2 Kaplan-Meier analysis for patients with NSCLC by the combined risk genotypes. (A,C) By 0, 1 and 2 risk genotypes (log-rank test for trend and multivariate analysis for OS and DSS, respectively); (B,D) by 0-1 and 2 risk genotypes (log-rank test and multivariate analysis) for OS and DSS, respectively in the PLCO trial. OS, overall survival; DSS, disease-specific survival.

number of risk genotypes. The trend test indicated that an increased genetic score was associated with a poorer survival (P_{trend}=0.001 for OS and P_{trend}=0.001 for DSS; *Table 3*). Based on similar survival rates of the subgroups, we divided all patients into a low-risk group (0–1 risk score) and a high-risk group (2 risk scores). Compared with the low-risk group, patients in the high-risk group had a higher death risk in the PLCO dataset (OS: HR 1.22, 95% CI: 1.04–1.44 and P=0.013; DSS: HR 1.25, 95% CI: 1.06–1.48 and P=0.009, *Table 3*). Moreover, when we created a genetic score by combining these two SNPs with all other previously reported SNPs identified in the present study populations, we found that the effect on NSCLC OS/DSS increased as the genetic score increased (P_{trend}<0.001) after adjustment for other covariates and that the high-risk group (i.e., 8–13 *vs.* 3–7 risk genotypes) was associated with a poorer survival (OS: HR 1.99, 95% CI: 1.71–2.31 and P<0.001; DSS: HR 1.94, 95% CI: 1.66–2.28 and P<0.001, *Table S6*). To visualize these effects, Kaplan-Meier curves for the associations between risk genotypes and NSCLC OS and DSS are shown in *Figure 2* and *Figure S4*.

Stratified analysis was performed to investigate whether the joint effect of risk genotypes on NSCLC survival was modified by the covariates, including age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy and surgery. Except for radiotherapy (OS: P=0.011; DSS: P=0.020), no significant interactions with these covariates were found (P>0.05 for both OS and DSS, *Table S7*). Therefore, we further reassessed the associations between the combined effect of the SNPs and NSCLC OS/DSS by subgroups of patients with or without radiotherapy (*Table S8*), and only in the subgroup of patients with radiotherapy, an increased genetic score was associated with a poorer survival (OS: P_{trend}=0.001; HR 1.52, 95% CI: 1.19–1.96 and P=0.001 for 2 scores *vs.* 0–1 score; DSS: P_{trend}<0.001; HR 1.55, 95% CI: 1.20–2.01 and P=0.001 for 2 scores *vs.* 0–1 score). To assess whether the SNP-associated outcomes in patients with NSCLC in early and advanced stages were also different, we calculated the survival for stage I–IIIA and IIIB–IV, separately. As a result, a significant association between an increased genetic score and a poorer survival was more obvious in patients with stage I–IIIA than in those with stage IIIB–IV [I–IIIA: P_{trend}=0.030 and 0.017 for OS and DSS, respectively; scores

2 *vs.* 0–1: HR 1.27 (0.99–1.64) and 1.37 (1.03–1.81) for OS and DSS, respectively; IIIB–IV: P_{trend}=0.012 and 0.017 for OS and DSS, respectively; HR 1.13 (0.92–1.39) and 1.14 (0.91–1.40) for OS and DSS, respectively, *Table S9*). These discrepancies in the survival between patients with stage I–IIIA and IIIB–IV were likely because the study population had more patients with stage I–IIIA than stage IIIN-IV, as no interaction was observed between the genetic scores and NSCLC stages.

ROC and time-dependent area under the receiver operating characteristic curve (AUC) estimation for NSCLC survival prediction

To explore the ability of the combined model to predict survival of NSCLC, we estimated the predictive value of the combined risk genotypes with ROC and time-dependent AUC curves. We compared the AUC for clinical variables with or without the risk genotypes of the two SNPs. After adding the risk genotypes, the 5-year OS AUC increased non-significantly from 89.12% to 89.32% (P=0.347, *Figure S5A*) and the 5-year DSS AUC increased non-significantly from 88.92% to 89.11% (P=0.416, *Figure S5B*). However, the 5-year OS AUC significantly increased from 88.87% to 89.88% (P=0.040, *Figure S5C*) and the 5-year DSS AUC significantly increased from 88.57% to 89.76% (P=0.025, *Figure S5D*), when combining risk genotypes of these two SNPs and all other previously reported SNPs identified from the present study population, suggesting the importance of identifying additional risk genotypes in this study populations. Furthermore, time-dependent AUC curves of these two SNPs and all other previously reported SNPs identified in the present study populations were provided to assess the ability of model to predict NSCLC OS and DSS in this study populations through the entire follow-up period (*Figure S5E,F,G,H*).

In silico functional validation

We further evaluated the correlations between SNPs and mRNA expression using RNA-seq of 373 transformed lymphoblastoid cell-lines in the 1,000 genomes project. Notably, the rs9673682 G allele was significantly correlated with increased mRNA expression levels of *PLCG2* in both additive and dominant models ($P=7.20\times10^{-5}$ and 0.002, respectively), but not for *PLPP1* rs115613985 (P=0.814 and 0.850, respectively) (*Figure 3*). Then, we extracted the eQTL analysis results from the database of the GTEx project and found that the rs9673682 G allele had a non-significant trend in correlation with mRNA expression levels of *PLCG2* in the whole blood cells (n=369, P=0.645) and normal lung tissues (n=383, P=0.885); however, there were no *PLPP1* expression data for further analysis (*Table S10*).

Additional differential expression analysis further revealed that both *PLCG2* and *PLPP1* had lower mRNA expression levels in 107 paired tumor tissues than in the adjacent normal tissues in the TCGA database (P≤0.001, *Figure S6*), and their lower expression levels were associated with a poorer NSCLC OS (26) (*Figure S7*).

Discussion

In the present study, we re-analyzed the published GWAS datasets by a hypothesis-based gene-set approach in the post-GWAS era to explore the roles of genetic variants in clinical outcomes of NSCLC patients. We chose the FCGR-mediated phagocytosis pathway genes as the candidate gene-set. As a result, we identified two novel SNPs (i.e., *PLCG2* rs9673682 T>G and *PLPP1* rs115613985 T>A) as independent predictors of NSCLC survival after adjustment for available covariates as well as all other SNPs previously reported in the PLCO study population. We also assessed the effects of the combined risk genotypes of these two SNPs in the prediction model and found that they collectively predicted a poorer survival of NSCLC, which is consistent with the fact that both *PLCG2* and *PLPP1* harboring the SNPs are functioning in the same FCGR-mediated phagocytosis pathway. Further functional prediction analysis by assessing publicly available databases found that the *PLCG2* rs9673682 G allele, which was associated with an improved survival, was also significantly associated with increased mRNA expression levels of *PLCG2*. Consistent with this, lower mRNA expression levels of *PLCG2* were observed in tumor tissues than in adjacent normal tissues and were also associated with a poorer survival of lung cancer patients. Although we did not find evidence for a correlation between the *PLPP1* rs115613985 A allele and its gene expression, the observed lower mRNA expression levels of *PLPP1* in tumor tissues than in adjacent normal tissues were also associated with a poorer survival. In addition, a prediction model that included all identified risk genotypes of previously reported SNPs in the same study population had significantly improved 5-year survival prediction as indicated by the ROC curve. Taken together, our findings of associations between genetic variants in the

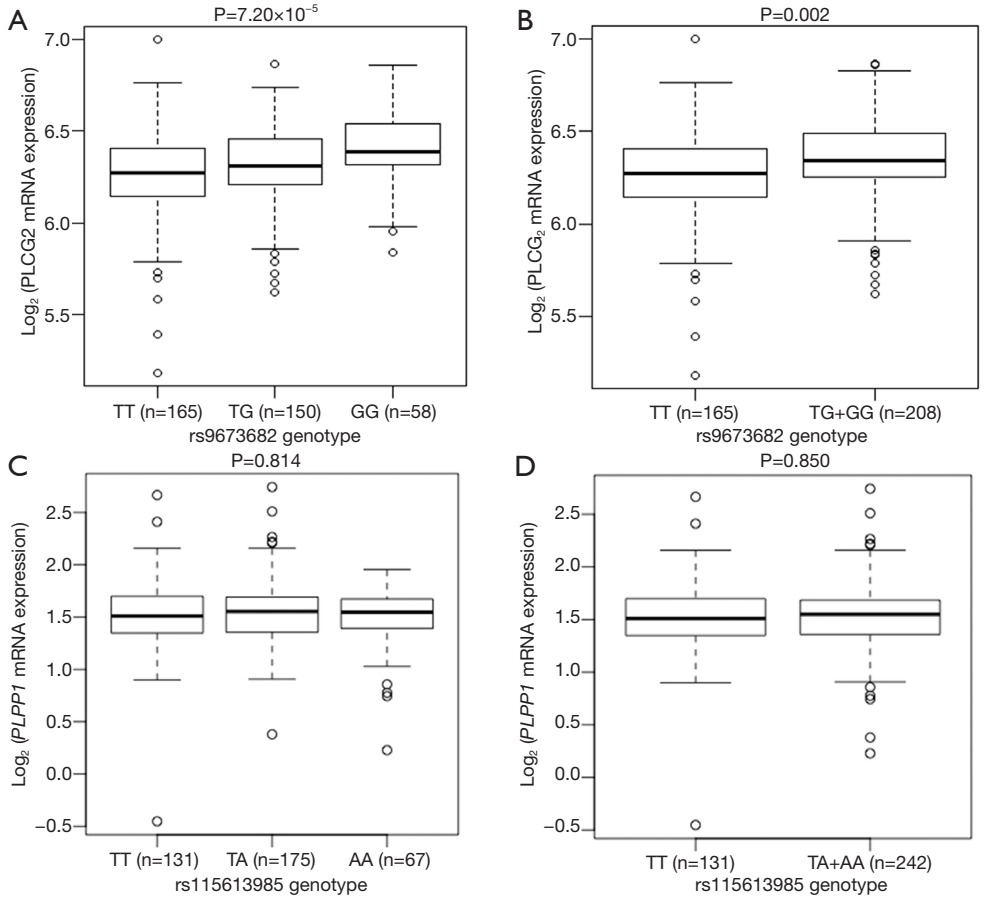

Figure 3 Correlation of the SNPs with mRNA expression in 373 transformed lymphoblastoid cells in the 1,000 Genomes Project. (A,B) Correlation between rs9673682 and *PLCG2* mRNA expression levels in additive and dominant model; (C,D) correlation between rs115613985 and *PLPP1* mRNA expression levels in additive and dominant model.

FCGR-mediated phagocytosis pathway genes and NSCLC survival are biologically plausible.

PLCG2 is located on chromosome region 16q23.3 and encodes phosphoinositide specific phospholipase C family protein PLCG2, which is critical for the modulation of the calcium signals in response to the stimulation of immune receptors (27). The expression of PLCG2 is widely distributed on hematopoietic cells, including macrophages, natural killer cells, mast cells, and neutrophils (28). In macrophages, FCGR-mediated Ca^{2+} increase is dependent on PLCG2, and interestingly, this abrogation of Ca^{2+} increase changes the FCGR-mediated inflammatory reaction, but not the FCGR-mediated phagocytosis, suggesting other mechanisms may be involved in this process (29). Myeloid-derived suppressor cells (MDSCs) support tumor growth in many cancers by suppressing anti-tumor T cell responses. For example, one study demonstrated that down-regulation of PLCG2 signaling in MDSCs could result in their aberrant expansion during tumor progression, suggesting the possibility that PLCG2 may be the target of anti-cancer therapy (30). Moreover, mutations in *PLCG2* are required for ibrutinib resistance in chronic lymphocytic leukemia, and patients with *PLCG2* mutations are more likely to progress (31). In NSCLC, *PLCG2* had a low somatic mutation rate according to the public available data from the database of the cBioPortal for Cancer Genomics (http://www.cbioportal.org). As shown in *Figure S8*, the mutation rate is low in both lung adenocarcinoma [2.19% (4/183) in the Broad study, 1.31% (12/915) in the MSK study, 1.23% (2/163) in the TSP study and 1.19% (7/586) in the TCGA study] and squamous cell carcinoma [0.78% (4/511) in the TCGA study], suggesting that germline genetic variants in *PLCG2* may play a more important role in the regulation of mRNA expression

of this gene. In the present study, we found that the rs9673682 minor G allele showed a significant association with a decreased risk of death in NSCLC patients and an increased *PLCG2* expression level in normal blood cells. Additional experimental data from the ENCODE project suggest that *PLCG2* rs9673682 T>G is located in a DNase I hypersensitive site with considerable levels of histone modification H3K4Me1, which may lead to an enhanced transcriptional activity (*Figure S9A*).

PLPP1, also known as *PPAP2A* or *LPP1*, is located on chromosome 5 and encodes a protein belonging to the phosphatidic acid phosphatase (PAP) family (32). One major function of PLPP1 is to dephosphorylate extracellular lysophosphatidic acid (LPA) and sphingosine-1-phosphate (S1P), which are important regulators of cell division, migration, and survival (33). It was shown that S1P might enhance the phagocytic activity of macrophages by up-regulating the expression of FCGR2A, suggesting PLPP1 might also contribute to the FCGR-mediated phagocytosis (34). Many cancer cells express low levels of PLPP1, which is an important contributor to the degradation of LPA, and high expression of LPA would produce resistance to chemotherapy and radiotherapy (35). Some *in vivo* studies showed that an increasing expression of PLPP1 in breast, ovarian and thyroid cancer cells would decrease tumor growth and metastasis in mouse models (36,37). In the present study, we observed that NSCLC patients with the *PLPP1* rs115613985 A allele had a worse survival. However, the eQTL results showed that the rs115613985 A allele had a non-significant trend in a correlation with mRNA expression levels of *PLPP1*, indicating other mechanisms, such as somatic mutation or other epigenetic effects, might be involved in the transcription of the *PLPP1* gene. Furthermore, *PLPP1* rs115613985 T>A is located in an H3K4Me1 mark and a Tbp motif based on the ENCODE Project data and transcription factor CHIP-seq data (*Figure S9B*), suggesting the minor A allele might affect the transcription factors.

There are limitations in the present study. Although we identified associations of genetic variants in *PLCG2* and *PLPP1* with NSCLC survival, the results may be limited to Caucasian populations and thus not generalizable to other ethnic populations. The sample size of the PLCO was not large enough to pass the FDR cut-off value in the multiple test correction, for which we used the relatively less strident BFDP method as a result of the highly correlated SNPs from imputation. In the PLCO dataset, only a few clinical variables were available, and other needed information, especially the details of treatment and PFS, were not accessible. The HLCS dataset only provided the summary genotyping results and OS data but not DSS and PFS, and thus we used OS to identify independent SNPs as NSCLC survival predictors. In addition, although we found some functional evidence to support the observed associations, the potential molecular mechanisms (e.g., several molecular signature associated with cell survival, migration, and DNA damage response) between the SNPs and prognosis in NSCLC remain unclear, and further functional experiments are essential to explore these molecular mechanisms underlying the observed associations, which will eventually provide their clinical utility in future precision medicine.

Conclusions

In conclusion, we identified two independent, novel SNPs (i.e., *PLCG2* rs9673682 T>G and *PLPP1* rs115613985 T>A) that were significantly associated with NSCLC survival in both PLCO and HLCS GWAS datasets, of which *PLCG2* rs9673682 T>G may have an effect on its mRNA expression. Our findings provided a rationale to further investigate the functions of survival-associated SNPs in the FCGR-mediated phagocytosis pathway genes in the development and progression of NSCLC.

Acknowledgments

We thank all the participants of the PLCO Cancer Screening Trial. We 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 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 these studies can be found in the supplemental data.

Funding: QW 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). DQ was supported by the State Scholarship Fund, China Scholarship Council, the Ministry of Education of China through Fudan University Shanghai Cancer Center. The Harvard Lung Cancer Susceptibility Study was supported

by NIH grants (CA092824, CA074386, CA090578) to DCC. SL was supported by National Institutes of Health (Grant ID: R01NS091307 and R56AG062302).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tlcr-19-318). EFP reports personal fees from Grid Therapeutics, personal fees from OncoCyte, outside the submitted work. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The original studies were approved by the institutional review boards of the National Cancer Institute and Massachusetts General Hospital, respectively, with written informed consent obtained from each participant.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
2. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature 2018;553:446-54.
3. Jia M, Zhu M, Zhou F, et al. Genetic variants of JNK and p38alpha pathways and risk of non-small cell lung cancer in an Eastern Chinese population. Int J Cancer 2017;140:807-17.
4. Zienolddiny S, Skaug V. Single nucleotide polymorphisms as susceptibility, prognostic, and therapeutic markers of nonsmall cell lung cancer. Lung Cancer (Auckl) 2011;3:1-14.
5. Stenzel-Bembenek A, Sagan D, Guz M, et al. Single nucleotide polymorphisms in lung cancer patients and cisplatin treatment. Postepy Hig Med Dosw (Online) 2014;68:1361-73.
6. Thulin NK, Wang TT. The Role of Fc Gamma Receptors in Broad Protection against Influenza Viruses. Vaccines (Basel) 2018;6:36.
7. Cassard L, Cohen-Solal J, Camilleri-Broet S, et al. Fc gamma receptors and cancer. Springer Semin Immunopathol 2006;28:321-8.
8. Nimmerjahn F, Ravetch JV. Fcgamma receptors as regulators of immune responses. Nat Rev Immunol 2008;8:34-47.
9. Pincetic A, Bournazos S, DiLillo DJ, et al. Type I and type II Fc receptors regulate innate and adaptive immunity. Nat Immunol 2014;15:707-16.
10. Weiskopf K, Weissman IL. Macrophages are critical effectors of antibody therapies for cancer. MAbs 2015;7:303-10.
11. Tamura K, Shimizu C, Hojo T, et al. FcgammaR2A and 3A polymorphisms predict clinical outcome of trastuzumab in both neoadjuvant and metastatic settings in patients with HER2-positive breast cancer. Ann Oncol 2011;22:1302-7.
12. Wang DS, Wei XL, Wang ZQ, et al. FcgammaRIIA and IIIA polymorphisms predict clinical outcome of trastuzumab-treated metastatic gastric cancer. Onco Targets Ther 2017;10:5065-76.
13. Shepshelovich D, Townsend AR, Espin-Garcia O, et al. Fc-gamma receptor polymorphisms, cetuximab therapy, and overall survival in the CCTG CO.20 trial of metastatic colorectal cancer. Cancer Med 2018;7:5478-87.
14. Tryka KA, Hao L, Sturcke A, et al. NCBI's Database of Genotypes and Phenotypes: dbGaP. Nucleic Acids Res 2014;42:D975-9.
15. Mailman MD, Feolo M, Jin Y, et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet 2007;39:1181-6.
16. Asomaning K, Miller DP, Liu G, et al. Second hand smoke, age of exposure and lung cancer risk. Lung Cancer 2008;61:13-20.
17. Qian D, Liu H, Wang X, et al. Potentially functional genetic variants in the complement-related immunity gene-set are associated with non-small cell lung cancer survival. Int J Cancer 2019;144:1867-76.
18. Lappalainen T, Sammeth M, Friedlander MR, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013;501:506-11.
19. Consortium GT. Human genomics. The Genotype-

Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015;348:648-60.
20. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014;511:543-50.
21. Aulchenko YS, Ripke S, Isaacs A, et al. GenABEL: an R library for genome-wide association analysis. Bioinformatics 2007;23:1294-6.
22. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Statistical Methodology 1995;57:289-300.
23. Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. Am J Hum Genet 2007;81:208-27.
24. Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. Stat Med 2006;25:3474-86.
25. 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.
26. Győrffy 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.
27. Wilde JI, Watson SP. Regulation of phospholipase C gamma isoforms in haematopoietic cells: why one, not the other? Cell Signal 2001;13:691-701.
28. Bae YS, Lee HY, Jung YS, et al. Phospholipase Cgamma in Toll-like receptor-mediated inflammation and innate immunity. Adv Biol Regul 2017;63:92-7.
29. Wen R, Jou ST, Chen Y, et al. Phospholipase C gamma 2 is essential for specific functions of Fc epsilon R and Fc gamma R. J Immunol 2002;169:6743-52.
30. Capietto AH, Kim S, Sanford DE, et al. Down-regulation of PLCgamma2-beta-catenin pathway promotes activation and expansion of myeloid-derived suppressor cells in cancer. J Exp Med 2013;210:2257-71.
31. Ahn IE, Underbayev C, Albitar A, et al. Clonal evolution leading to ibrutinib resistance in chronic lymphocytic leukemia. Blood 2017;129:1469-79.
32. Carman GM, Han GS. Phosphatidic acid phosphatase, a key enzyme in the regulation of lipid synthesis. J Biol Chem 2009;284:2593-7.
33. Tang X, Benesch MG, Brindley DN. Lipid phosphate phosphatases and their roles in mammalian physiology and pathology. J Lipid Res 2015;56:2048-60.
34. Duong CQ, Bared SM, Abu-Khader A, et al. Expression of the lysophospholipid receptor family and investigation of lysophospholipid-mediated responses in human macrophages. Biochim Biophys Acta 2004;1682:112-9.
35. Brindley DN, Lin FT, Tigyi GJ. Role of the autotaxin-lysophosphatidate axis in cancer resistance to chemotherapy and radiotherapy. Biochim Biophys Acta 2013;1831:74-85.
36. Tang X, Benesch MG, Dewald J, et al. Lipid phosphate phosphatase-1 expression in cancer cells attenuates tumor growth and metastasis in mice. J Lipid Res 2014;55:2389-400.
37. Nakayama J, Raines TA, Lynch KR, et al. Decreased peritoneal ovarian cancer growth in mice lacking expression of lipid phosphate phosphohydrolase 1. PLoS One 2015;10:e0120071.

Cite this article as: Qian D, Liu H, Zhao L, Wang X, Luo S, Moorman PG, Patz EF Jr, Su L, Shen S, Christiani DC, Wei Q. Novel genetic variants in genes of the Fc gamma receptor-mediated phagocytosis pathway predict non-small cell lung cancer survival. Transl Lung Cancer Res 2020;9(3):575-586. doi: 10.21037/tlcr-19-318

Acknowledgments

We wish to thank all of the investigators and funding agencies that enabled the deposition of data in dbGaP and PLCO that we used in this study.

The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000336.v1.p1 and phs000093.v2.p2. Principal Investigators: Maria Teresa Landi. Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. Neil E. Caporaso. Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

Funding support for the GWAS of Lung Cancer and Smoking was provided through the NIH Genes, Environment and Health Initiative (GEI) (Z01 CP 010200). The human subjects participating in the GWAS derive from The Environment and Genetics in Lung Cancer Etiology (EAGLE) case-control study and the Prostate, Lung Colon and Ovary Screening Trial and these studies are supported by intramural resources of the National Cancer Institute. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the Gene Environment Association Studies, GENEVA Coordinating Center (U01HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438).

PLCO was also supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. The authors thank PLCO screening center investigators and staff, and the staff of Information Management Services Inc. and Westat Inc. Most importantly, we acknowledge trial participants for their contributions that made this study possible.

Table S1 Comparison of characteristics of the subjects between the PLCO trial and HLCS study

Characteristics	PLCO		HLCS		Chi-square
	Frequency	Deaths, n (%)	Frequency	Deaths, n (%)	
Total	1,185	798 (67.3)	984	665 (67.5)	–
Median overall survival (months)	23.8		39.9		–
Age, years					<0.0001
≤71	636	400 (62.9)	654	428 (65.4)	
>71	549	398 (72.5)	330	237 (71.8)	
Sex					0.001
Male	698	507 (72.6)	507	379 (74.7)	
Female	487	291 (59.8)	477	286 (59.9)	
Smoking status					0.166
Never	115	63 (54.8)	92	52 (56.5)	
Current	423	272 (64.3)	390	266 (68.2)	
Former	647	463 (71.6)	502	347 (69.1)	
Histology					<0.0001
Adenocarcinoma	577	348 (60.3)	597	378 (63.3)	
Squamous cell carcinoma	285	192 (67.4)	216	156 (72.2)	
Others	323	258 (79.9)	171	131 (76.6)	
Stage					0.038
I–IIIA	655	315 (48.1)	606	352 (58.0)	
IIIB–IV	528	482 (91.3)	377	313 (83.0)	
Missing	2	–	–	–	–

PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; HLCS, Harvard Lung Cancer Susceptibility study.

Table S2 List of 97 selected genes in the FCGR-mediated phagocytosis pathway

Dataset	Name of pathway[1]	Selected genes[2]	Number of genes
Genecards	Fc gamma R-mediated phagocytosis	*AKT1, AKT2, AKT3, AMPH, ARF6, ARPC1A, ARPC1B, ARPC2, ARPC3, ARPC4, ARPC5, ARPC5L, ASAP1, ASAP2, ASAP3, BIN1, CDC42, CFL1, CFL2, CRK, CRKL, DNM2, DOCK2, FCGR1A, FCGR2A, FCGR2B, FCGR3A, GAB2, GSN, HCK, INPP5D, INPPL1, LAT, LIMK1, LIMK2, LYN, MAP2K1, MAPK1, MAPK3, MARCKS, MARCKSL1, MYO10, NCF1, PAK1, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3, PIP5K1A, PIP5K1B, PIP5K1C, PLA2G4A, PLA2G4B, PLA2G4D, PLA2G4E, PLA2G4F, PLA2G6, PLCG1, PLCG2, PLD1, PLD2, PLPP1, PLPP2, PLPP3, PRKCA, PRKCB, PRKCD, PRKCE, PRKCG, PTPRC, RAC1, RAC2, RAF1, RPS6KB1, RPS6KB2, SCIN, SPHK1, SPHK2, SYK, VASP, VAV1, VAV2, VAV3, WASF1, WASF2, WASF3, WASL*	89
MSigDB	KEGG FCGR-mediated phagocytosis	*DNM1, DNM1L, DNM3, FCGR2C, PIK3CG, PIK3R5, PIKFYVE, PIP4K2B*	8
Total			97

[1], genes were selected based on online datasets and literatures; [2], duplicated genes, pseudo genes and genes in X chromosome had been removed, 2 genes removed in the GeneCards dataset and 89 genes removed in the MSigDB dataset.

Table S3 Associations of the first 10 principal components and OS of NSCLC in the PLCO trial (n=1,185)

PC*	Parameter estimate	Standard error	Chi-square	P
PC1	4.821	1.353	12.697	<0.001
PC2	–0.681	1.228	0.308	0.579
PC3	–3.054	0.949	10.351	0.001
PC4	–2.837	1.246	5.184	0.023
PC5	–0.910	1.232	0.546	0.460
PC6	1.355	1.252	1.172	0.279
PC7	–0.236	1.218	0.038	0.846
PC8	–1.684	1.322	1.622	0.203
PC9	–1.886	1.267	2.216	0.137
PC10	0.347	1.240	0.078	0.180

*, the first 4 were used for the adjustment for population stratification in the multivariate analysis. OS, overall survival; NSCLC, non-small cell lung cancer; PLCO, the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; PC, principal component.

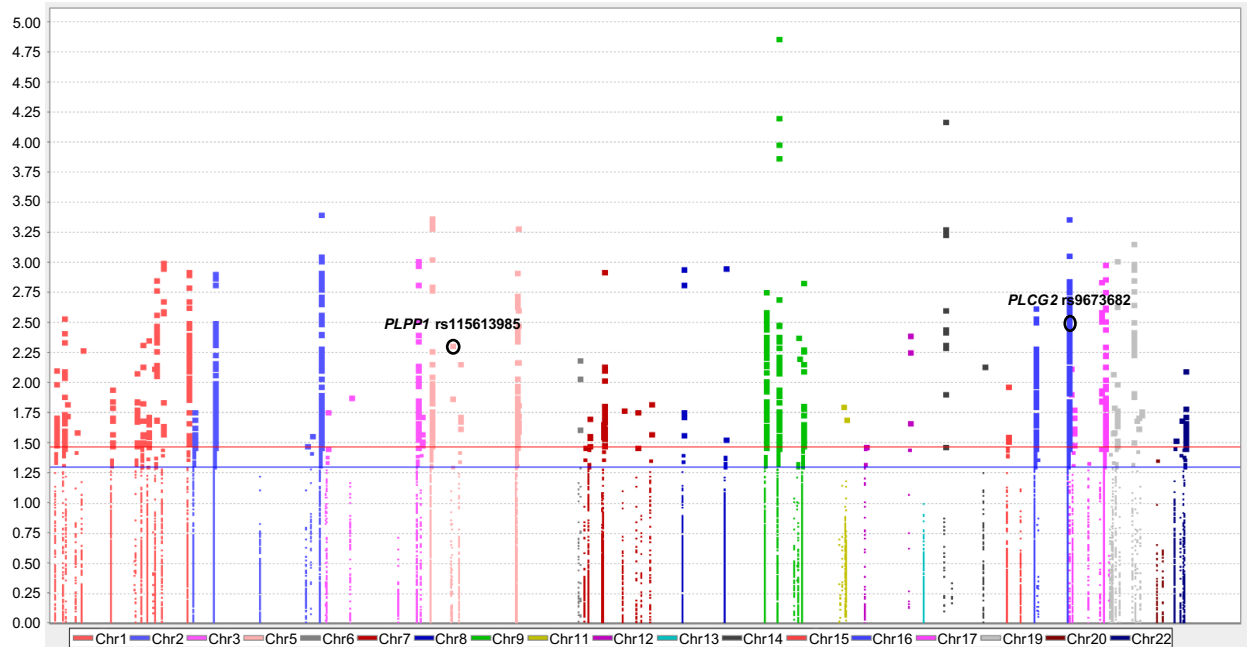

Figure S1 Manhattan plot of the discovery set in genotype data of the PLCO trial. The statistical values across the autosomes of associations between 24,734 SNPs and overall survival are plotted as –log10 P values. The blue horizontal line indicates P=0.050 and the red line indicates BFDP =0.80.

Table S4 Function prediction of the two validated SNPs associated with survival of NSCLC*

SNP	Gene	Chromosome	Source of genotype	RegDB[1]	Haploreg v4.1[2]				
					Enhancer histone marks	DNAse	Motifs changed	Selected eQTL hits	dbSNP function annotation
rs9673682 T>G	*PLCG2*	16	Imputed	4	4 tissues	BRST, CRVX, SKIN	STAT	7 hits	Intronic
rs115613985 T>A	*PLPP1*	5	Imputed	6	5 tissues	–	9 altered motifs	–	Intronic

*, both of the two SNPs had no function in the SNPinfo: https://snpinfo.niehs.nih.gov/snpinfo/snpfunc.html; [1], RegulomeDB: http://regulomedb.org/; [2], Haploreg: http://archive.broadinstitute.org/mammals/haploreg/haploreg.php. SNP, single-nucleotide polymorphism; NSCLC, non-small cell lung cancer.

Table S5 Independent test for SNPs identified in this study and all previously published in the present study populations in the PLCO dataset

Variables[1]	Category[2]	Frequency[3]	HR (95% CI)	P
Age	Continuous	1,124	1.04 (1.02–1.05)	<0.001
Sex	Male/female	665/459	0.79 (0.67–0.92)	0.003
Smoking status	Never/former	111/610	1.85 (1.39–2.46)	<0.001
	Never/current	111/403	1.85 (1.37–2.50)	<0.001
Histology	AD/SC	555/268	1.16 (0.96–1.41)	0.127
	AD/others	555/301	1.38 (1.16–1.65)	<0.001
Stage	I I–IIIA/IIB-IV	626/498	2.99 (2.46–3.64)	<0.001
Chemotherapy	No/yes	609/515	0.59 (0.49–0.71)	<0.001
Radiotherapy	No/yes	727/397	0.98 (0.83–1.16)	0.818
Surgery	No/yes	605/519	0.20 (0.15–0.26)	<0.001
ABCG1 rs225390[4]	AA/AG/GG	529/475/120	0.88 (0.78–0.98)	0.021
ADAM12 rs10794069[5]	AA/AG/GG	638/411/75	1.19 (1.06–1.34)	0.003
DTX1 rs1732793[5]	GG/GA/AA	750/340/34	1.22 (1.06–1.40)	0.006
E2F3 rs3806116[5]	GG/GT/TT	428/530/166	1.17 (1.05–1.30)	0.004
IRAK2 rs779901[6]	CC/CT/TT	831/271/22	0.81 (0.69–0.94)	0.007
VWF rs73049469[7]	CC/CA/AA	892/222/10	1.20 (1.01–1.41)	0.035
ITGB2 rs3788142[7]	GG/GA/AA	650/403/71	1.14 (1.01–1.28)	0.034
RUNX3 rs7553295[8]	GG/GT/TT	585/457/82	0.83 (0.73–0.94)	0.003
AMD1 rs1279590[8]	GG/GA/AA	819/280/25	0.80 (0.69–0.93)	0.004
MSRA rs73534533[8]	CC/CA/AA	937/177/10	0.78 (0.64–0.95)	0.011
TNFRSF1B rs677844[9]	TT/TC/CC	620/429/75	0.84 (0.74–0.94)	0.003
IKBKAP rs4978754[9]	CC/CT/TT	945/169/10	0.82 (0.67–0.99)	0.041
PLCG2 rs9673682[10]	TT/TG/GG	501/496/127	0.89 (0.80–1.00)	0.040
PLPP1 rs115613985[10]	TT/TA/AA	438/526/160	1.23 (1.10–1.37)	<0.001

[1], stepwise analysis included age, sex, smoking status, tumor stage, histology, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3, PC4 and SNPs identified in this study and previous studies; [2], the leftmost genotype was used as the reference; [3], 61 missing date were excluded; [4], two SNPs were reported in the previous publication (PMID: 26757251); SNP rs225388 could not enter the final model. [5], five SNPs were reported in previous publication (PMID: 27557513); SNP rs199731120 and rs35970494 could not enter the final model; [6], one SNP was reported in the previous publication (PMID: 29978465); [7], two SNPs were reported in the previous publication (PMID: 30259978); [8], three SNPs were reported in the previous publication (PMID: 30650190); [9], two SNPs were reported in the previous publication (PMID: 30989732); [10], two SNPs identified in this study. SNP, single-nucleotide polymorphisms; PLCO, Prostate, Lung, Colorectal and Ovarian cancer screening trial; HR, hazards ratio; CI, confidence interval; AD, adenocarcinoma; SC, squamous cell carcinoma.

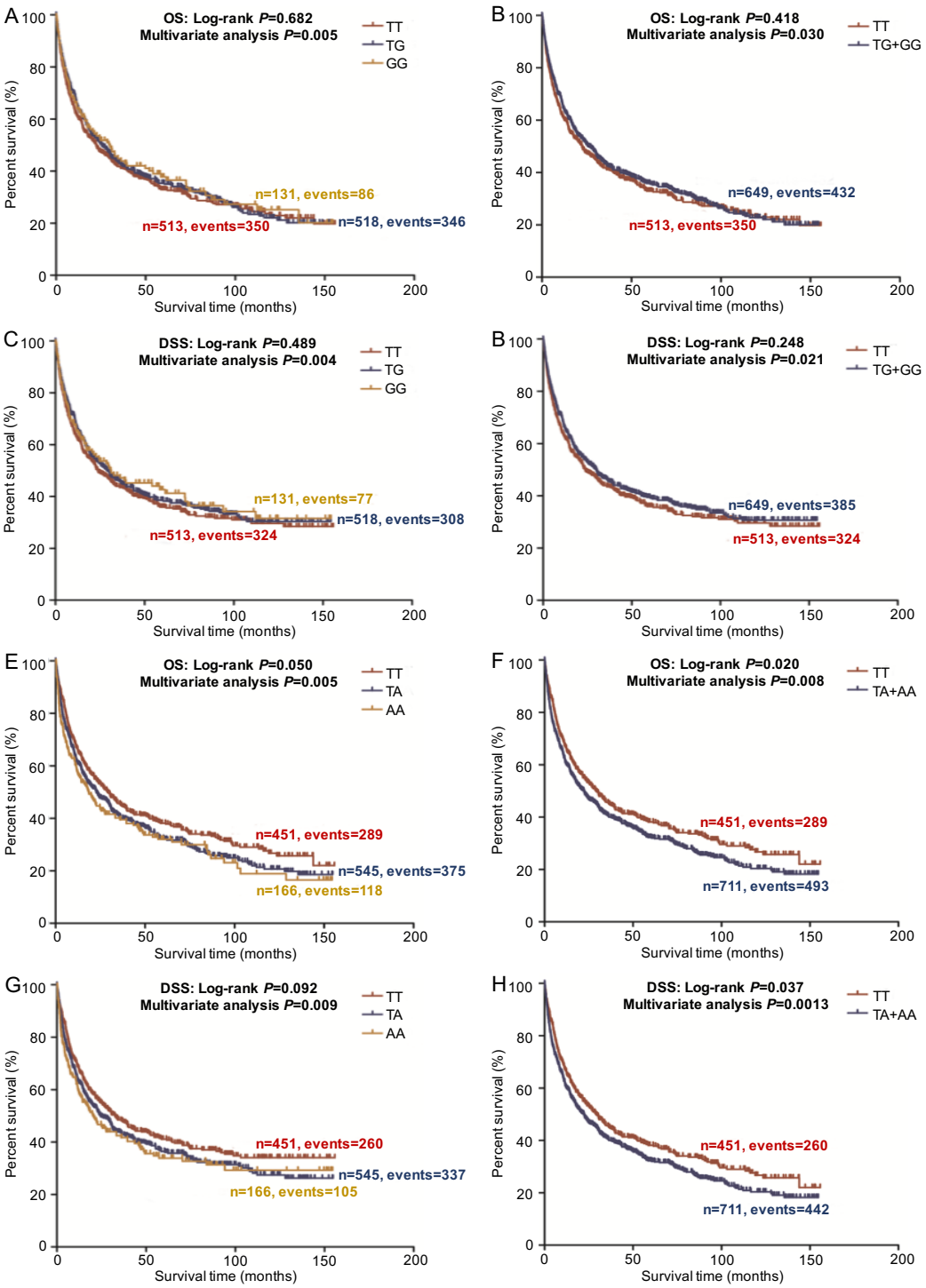

Figure S2 Kaplan-Meier analysis for patients with NSCLC by SNPs. (A,B,C,D) Kaplan-Meier survival curves of rs9673682 in additive and dominant models; (E,F,G,H) Kaplan-Meier survival curves of rs115613985 in additive and dominant models.

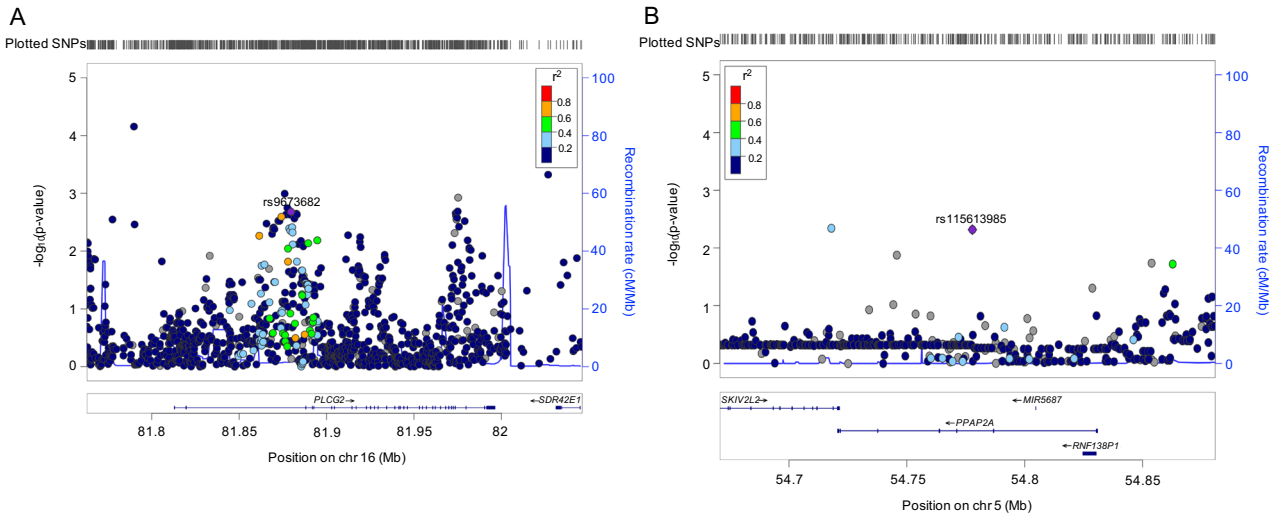

Figure S3 Regional association plots for the two independent SNPs in the FCGR-mediated phagocytosis pathway genes. SNPs in the region of 50 kilobases up or downstream of (A) *PLCG2* rs9673682 and (B) *PLPP1* rs115613985. Data points are colored according to the level of linkage disequilibrium of each pair of SNPs based on the hg19/1000 Genomes European population. The left-hand y-axis shows the association P value of individual SNPs in the discovery dataset, which is plotted as –log10 (P) against chromosomal base-pair position. The right-hand y-axis shows the recombination rate estimated from HapMap Data Rel 22/phase II European population.

Table S6 Associations between combined SNPs identified in this study and all previously published in the present study populations and survival of NSCLC in the PLCO Trial

Number of risk genotypes[1]	Number[2]	OS[3]			DSS[3]		
		Death, n (%)	HR (95% CI)	P	Death, n (%)	HR (95% CI)	P
3	10	7 (70.00)	1.00		5 (50.00)	1.00	
4	31	20 (64.52)	0.63 (0.26–1.54)	0.312	15 (48.39)	0.69 (0.24–1.94)	0.479
5	75	36 (48.00)	1.13 (0.49–2.56)	0.778	34 (45.33)	1.45 (0.56–3.74)	0.444
6	157	100 (63.69)	1.75 (0.80–3.82)	0.162	87 (55.41)	2.01 (0.81–5.01)	0.135
7	215	119 (55.35)	1.78 (0.82–3.88)	0.146	111 (51.63)	2.26 (0.91–5.62)	0.078
8	251	176 (70.12)	2.69 (1.24–5.81)	0.012	159 (63.35)	3.27 (1.33–8.07)	0.010
9	197	153 (77.66)	2.97 (1.37–6.44)	0.006	136 (69.04)	3.43 (1.39–8.49)	0.008
10	123	96 (78.05)	3.15 (1.43–6.93)	0.004	85 (69.11)	3.66 (1.46–9.17)	0.006
11	57	44 (77.19)	3.99 (1.76–9.04)	0.001	40 (70.18)	4.90 (1.90–12.63)	0.001
12	8	8 (100.00)	5.07 (1.77–14.52)	0.003	7 (87.50)	6.01 (1.84–19.58)	0.003
13	1	0 (0.00)	–	–	0 (0.00)	–	–
Trend				<0.001			<0.001
3–7	488	282 (57.79)	1.00	–	252 (51.64)	1.00	–
8–13	637	477 (74.88)	1.99 (1.71–2.31)	<0.001	427 (67.03)	1.94 (1.66–2.28)	<0.001

[1], risk genotypes were *ABCG1* rs225390 AA, *ADAM12* rs10794069 AG+GG, *DTX1* rs1732793 GA + AA, *E2F3* rs3806116 GT+TT, *IRAK2* rs779901 CC, *VWF* rs73049469 CA + AA, *ITGB2* rs3788142 GA + AA, *RUNX3* rs7553295 GG, *AMD1* rs1279590 GG, *MSRA* rs73534533 CC, *TNFRSF1B* rs677844 TT, *IKBKAP* rs4978754 CC, *PLCG2* rs9673682 TT and *PLPP1* rs115613985 TA + AA; [2], 60 missing date were excluded, no patients with 0, 1, 2 or 14 risk genotypes; [3], adjusted for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery and principal components. SNP, single-nucleotide polymorphism; NSCLC, non-small cell lung cancer; PLCO, Prostate, Lung, Colorectal and Ovarian cancer screening trial; OS, overall survival; DSS, disease-specific survival; HR, hazards ratio; CI, confidence interval.

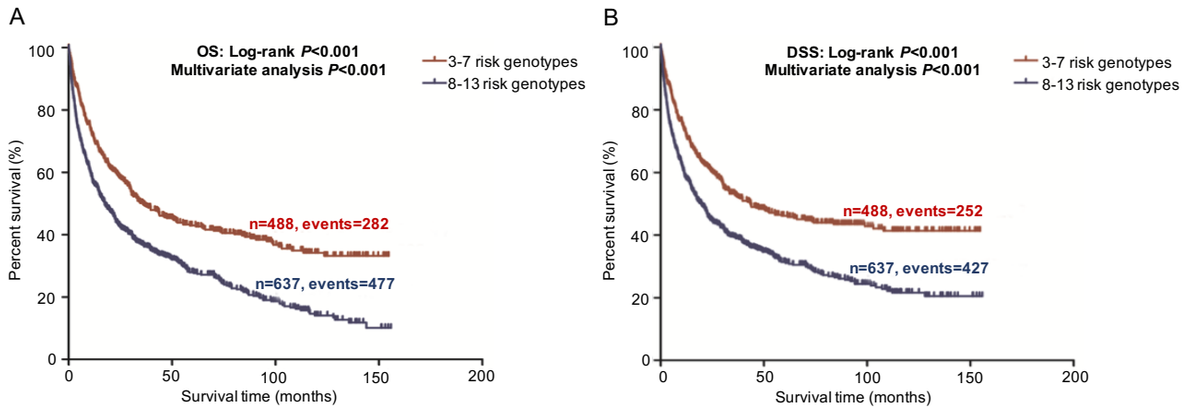

Figure S4 Kaplan-Meier survival curves of combined all the SNPs identified in this study and previously published in the present study populations.

Table S7 Stratified analysis for association between the risk genotypes and survival of NSCLC in the PLCO trial[1]

Characteristics	0–1 risk genotype			2 risk genotypes			OS[2]			DSS[2]		
	All	OS death (%)	DSS death (%)	All	OS death (%)	DSS death (%)	HR (95% CI)	P	P_{inter}[3]	HR (95% CI)	P	P_{inter}[3]
Age (years)									0.383			0.425
≤71	463	291 (62.85)	259 (55.94)	159	101 (63.52)	93 (58.49)	1.12 (0.89–1.40)	0.353		1.15 (0.90–1.46)	0.262	
>71	401	278 (69.33)	248 (61.85)	139	112 (80.58)	102 (73.38)	1.38 (1.10–1.73)	0.005		1.39 (1.10–1.77)	0.006	
Sex									0.388			0.240
Male	503	355 (70.58)	307 (61.03)	185	146 (78.92)	133 (71.89)	1.31 (1.08–1.59)	0.007		1.36 (1.11–1.68)	0.003	
Female	361	214 (59.28)	200 (55.40)	113	67 (59.29)	62 (54.87)	1.06 (0.80–1.42)	0.676		1.06 (0.78–1.43)	0.711	
Smoking status									0.975			0.965
Never	88	45 (51.14)	44 (50.00)	26	17 (65.38)	16 (61.54)	1.81 (0.91–3.61)	0.091		1.83 (0.90–3.72)	0.094	
Former	473	338 (71.46)	303 (64.06)	165	119 (72.12)	107 (64.85)	1.28 (1.03–1.58)	0.024		1.27 (1.02–1.59)	0.034	
Current	303	186 (61.39)	160 (52.81)	107	77 (71.96)	72 (67.29)	1.07 (0.81–1.40)	0.652		1.12 (0.84–1.49)	0.434	
Histology									0.505			0.459
Adenocarcinoma	416	248 (59.62)	229 (55.05)	154	97 (62.99)	91 (59.09)	1.12 (0.88–1.42)	0.361		1.12 (0.87–1.43)	0.379	
Squamous cell carcinoma	201	126 (62.69)	102 (50.75)	78	61 (78.21)	55 (70.51)	1.56 (1.13–2.15)	0.007		1.78 (1.26–2.53)	0.001	
Others	247	195 (78.95)	176 (71.26)	66	55 (83.33)	49 (74.24)	1.20 (0.89–1.64)	0.237		1.20 (0.86–1.66)	0.279	
Tumor stage									0.952			0.638
I–IIIA	486	228 (46.91)	181 (37.24)	160	84 (52.50)	70 (43.75)	1.27 (0.98–1.64)	0.067		1.37 (1.03–1.81)	0.029	
IIIB–IV	378	341 (90.21)	326 (86.24)	138	129 (93.48)	125 (90.58)	1.13 (0.92–1.39)	0.240		1.14 (0.92–1.41)	0.226	
Chemotherapy									0.410			0.387
No	458	255 (55.68)	211 (46.07)	172	109 (63.37)	94 (54.65)	1.22 (0.97–1.54)	0.085		1.23 (0.96–1.58)	0.102	
Yes	406	314 (77.34)	296 (72.91)	126	104 (82.54)	101 (80.16)	1.17 (0.93–1.46)	0.185		1.20 (0.95–1.50)	0.130	
Radiotherapy									0.011			0.020
No	561	328 (58.47)	283 (50.45)	191	117 (61.26)	104 (54.45)	1.01 (0.81–1.25)	0.946		1.03 (0.81–1.29)	0.837	
Yes	303	241 (79.54)	224 (73.93)	107	96 (89.72)	91 (85.05)	1.54 (1.20–1.97)	0.001		1.56 (1.21–2.02)	0.001	
Surgery									0.654			0.708
No	466	409 (87.77)	383 (82.19)	162	150 (92.59)	145 (89.51)	1.21 (1.00–1.46)	0.054		1.23 (1.01–1.50)	0.038	
Yes	398	160 (40.20)	124 (31.16)	136	63 (46.32)	50 (36.76)	1.26 (0.94–1.69)	0.125		1.32 (0.95–1.85)	0.101	

[1], 23 missing date were excluded; [2], adjusted for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3, and PC4; [3], P_{inter}: P value for interaction analysis between characteristic and risk genotypes. OS, overall survival; DSS, disease-specific survival; NSCLC, non-small cell lung cancer; PLCO, the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; HR, hazards ratio; CI, confidence interval.

Table S8 Associations between two combined SNPs and survival by subgroups of NSCLC patients with or without radiotherapy in the PLCO Trial[1]

Number of risk genotypes	Multivariate analysis (without radiotherapy)[2]				Multivariate analysis (with radiotherapy)[2]			
	All	Death, n (%)	HR (95% CI)	P	All	Death, n (%)	HR (95% CI)	P
OS								
0	156	88 (56.41)	1.00		80	64 (80.00)	1.00	
1	405	240 (59.26)	1.26 (0.98–1.61)	0.075	223	177 (79.37)	1.24 (0.92–1.68)	0.162
2	191	117 (61.26)	1.18 (0.89–1.57)	0.260	107	96 (89.72)	1.78 (1.28–2.49)	0.001
Trend of OS				0.312				0.001
0–1	561	328 (58.47)	1.00		303	241 (79.54)	1.00	
2	191	117 (61.26)	1.00 (0.81–1.24)	0.998	107	96 (89.72)	1.52 (1.19–1.96)	0.001
DSS								
0	156	78 (50.00)	1.00		80	58 (72.50)	1.00	
1	405	205 (50.62)	1.23 (0.94–1.60)	0.137	223	166 (74.44)	1.25 (0.91–1.71)	0.162
2	191	104 (54.45)	1.18 (0.87–1.60)	0.285	107	91 (85.05)	1.83 (1.29–2.58)	0.001
Trend of DSS				0.328				<0.001
0–1	561	283 (50.45)	1.00		303	224 (73.93)	1.00	
2	191	104 (54.45)	1.02 (0.81–1.28)	0.867	107	91 (85.05)	1.55 (1.20–2.01)	0.001

[1], 23 missing date were excluded; [2], adjusted for age, sex, stage, histology, smoking status, chemotherapy, surgery, PC1, PC2, PC3, and PC4. SNP, single-nucleotide polymorphism; OS, overall survival; DSS, disease-specific survival; NSCLC, non-small cell lung cancer; PLCO, the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; HR, hazards ratio; CI, confidence interval.

Table S9 Associations between two combined SNPs and survival by subgroups of NSCLC patients with I–IIIA or IIIB–IV stage in the PLCO Trial[1]

Number of risk genotypes	Multivariate analysis (I–IIIA stage)[2]				Multivariate analysis (IIIB–IV stage)[2]			
	All	Death, n (%)	HR (95% CI)	P	All	Death, n (%)	HR (95% CI)	P
OS								
0	132	58 (43.94)	1.00		104	94 (90.38)	1.00	
1	354	170 (48.02)	1.21 (0.89–1.65)	0.231	274	247 (90.15)	1.41 (1.11–1.81)	0.006
2	160	84 (52.50)	1.46 (1.03–2.06)	0.032	138	129 (93.48)	1.44 (1.01–1.89)	0.009
Trend of OS				0.030				0.012
0–1	486	228 (46.91)	1.00		378	341 (90.21)	1.00	
2	160	84 (52.50)	1.27 (0.99–1.64)	0.065	138	129 (93.48)	1.13 (0.92–1.39)	0.256
DSS								
0	132	46 (34.85)	1.00		104	90 (86.54)	1.00	
1	354	135 (38.14)	1.22 (0.86–1.72)	0.264	274	236 (86.13)	1.40 (1.09–1.80)	0.008
2	160	70 (43.75)	1.58 (1.08–2.31)	0.019	138	125 (90.58)	1.44 (1.09–1.90)	0.010
Trend of DSS				0.017				0.013
0–1	486	181 (37.24)	1.00		378	326 (86.24)	1.00	
2	160	70 (43.75)	1.37 (1.03–1.81)	0.030	138	125 (90.58)	1.14 (0.91–1.40)	0.236

[1], 23 missing date were excluded; [2], adjusted for age, sex, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3, and PC4. SNP, single-nucleotide polymorphism; OS, overall survival; DSS, disease-specific survival; NSCLC, non-small cell lung cancer; PLCO, the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; HR, hazards ratio; CI, confidence interval.

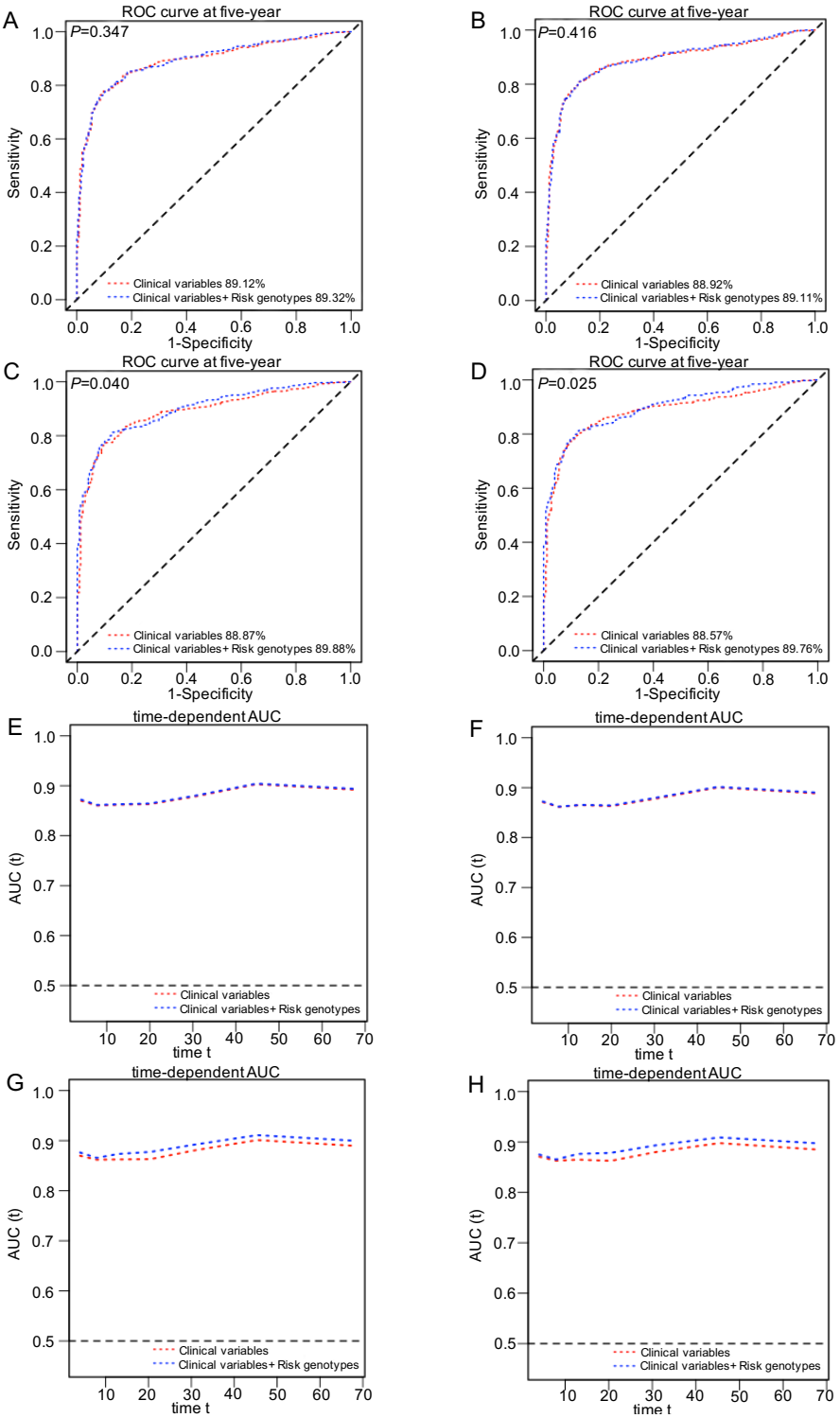

Figure S5 ROC curve and time-dependent AUC estimation for prediction of NSCLC survival in the PLCO dataset. Five-year survival prediction by ROC curve: (A,B) for OS and DSS of the two SNPs identified in the present study, (C,D) for OS and DSS of all the SNPs identified in this study and previously published in the present study populations. Time-dependent AUC estimation: (E,F) for OS and DSS of the two SNPs identified in the present study, (G,H) for OS and DSS of all the SNPs identified in this study and previously published in the present study populations.

Table S10 Correlation between the two SNPs and their mRNA expression in whole blood and normal lung tissue in the GTEx project

SNP	Gene	Whole blood (n=369)			Normal lung tissue (n=383)		
		β	SE	P	β	SE	P
rs9673682 T>G	*PLCG2*	0.012	0.258	0.645	−0.005	0.038	0.885
rs115613985 T>A	*PLPP1*	–	–	–	–	–	–

SNP, single-nucleotide polymorphism; GTEx, genotype-tissue expression project; SE, standard error.

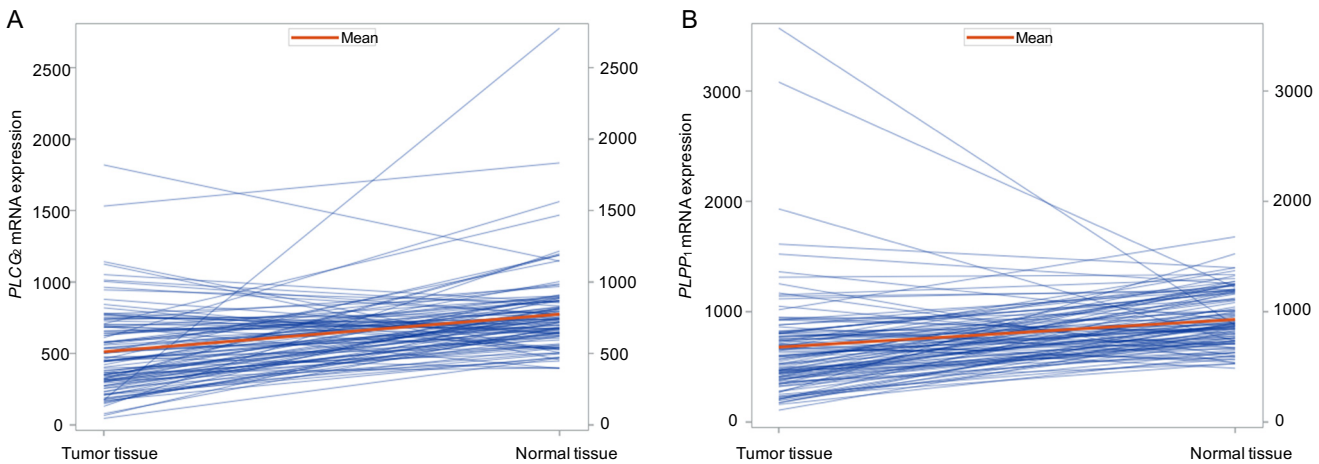

Figure S6 Differential mRNA expression analysis by using the data of the Cancer Genome Atlas (TCGA). Compared to the adjacent normal tissues, lower expression of *PLCG2* (A) and *PLPP1* (B) were found in the 107 paired tumor tissues (P≤0.001 for both).

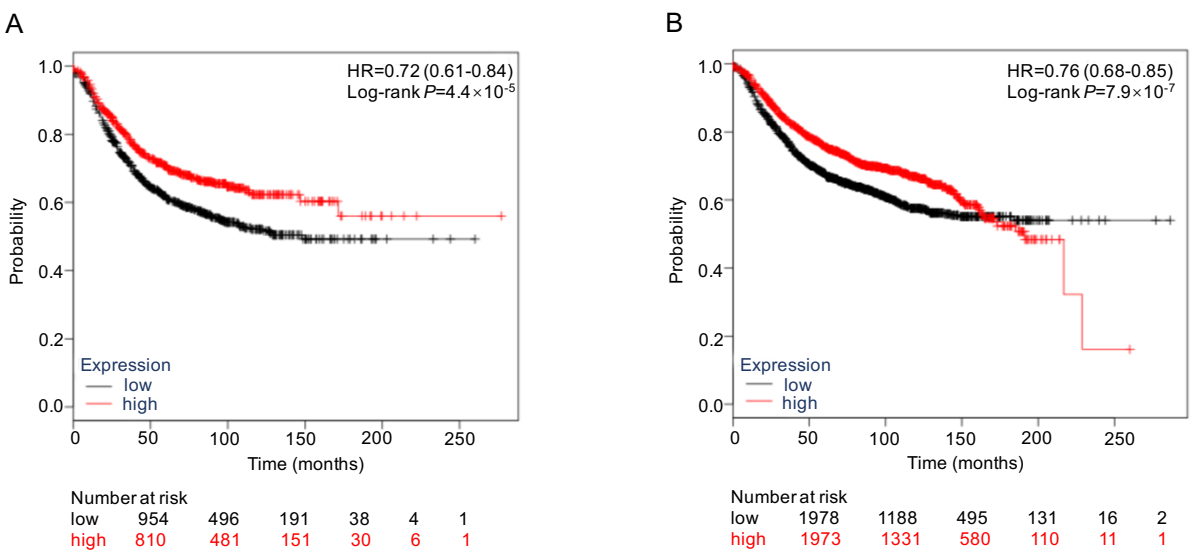

Figure S7 Kaplan-Meier analysis for patients with NSCLC by the two genes. Based on online survival analysis software, lower expression levels of *PLCG2* (A) and *PLPP1* (B) were associated with a poorer survival of NSCLC.

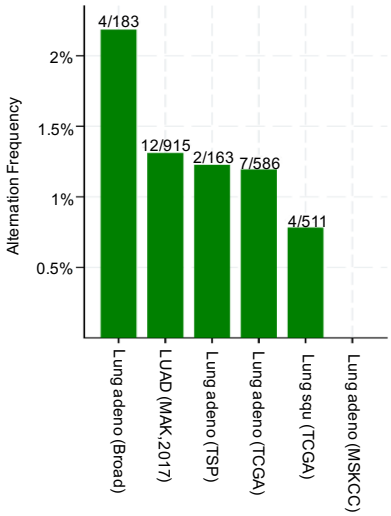

Figure S8 Mutation analysis of *PLCG2* gene in lung tumor tissues by using public available data in the database of the cBioPortal for Cancer Genomics (http://www.cbioportal.org). *PLCG2* gene had low mutation rates in both lung adenocarcinoma [2.19% (4/183) in the Broad study, 1.31% (12/915) in the MSK study, 1.23% (2/163) in the TSP study and 1.19% (7/586) in the TCGA study] and squamous cell carcinoma [0.78% (4/511) in the TCGA study].

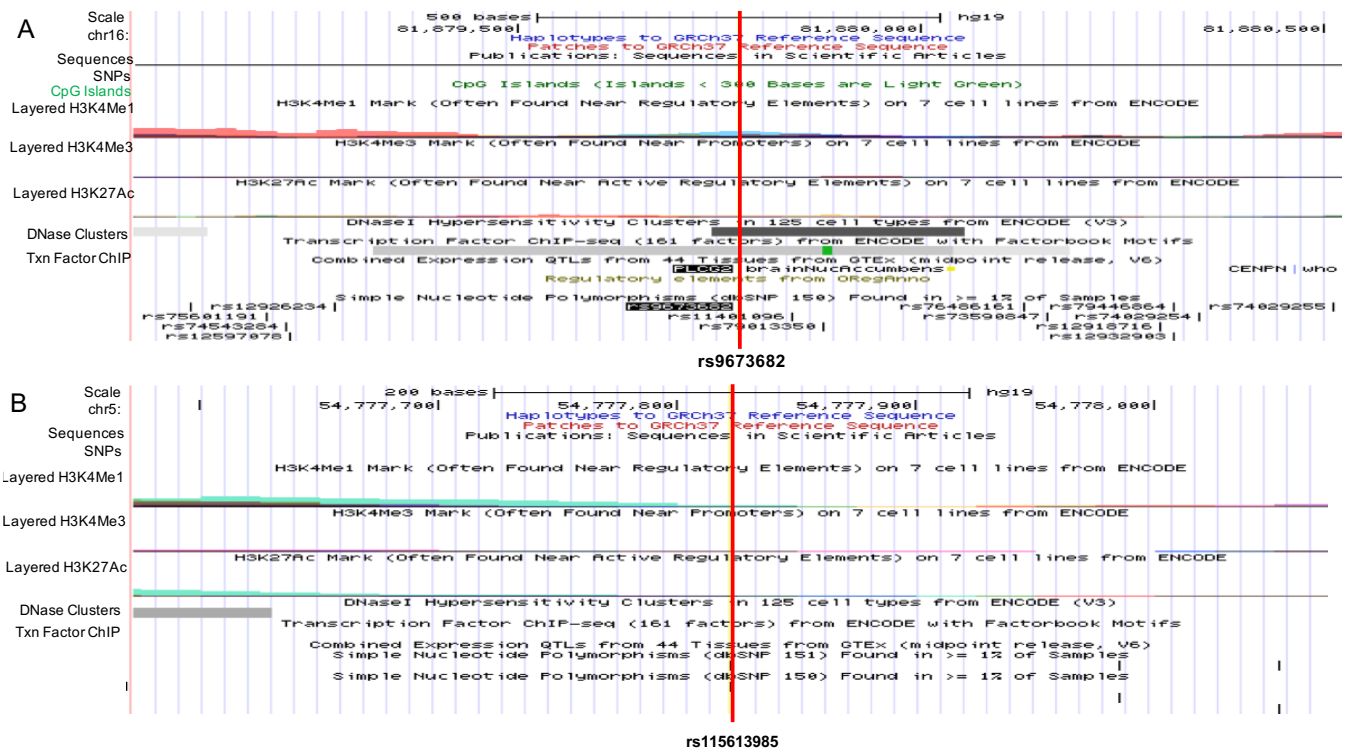

Figure S9 Functional prediction of SNPs in the ENCODE project. (A) Location and functional prediction of SNP rs9673682; (B) location and functional prediction of SNP rs115613985.