Am J Cancer Res 2020;10(6):1770-1784
www.ajcr.us /ISSN:2156-6976/ajcr0114150

Original Article

Novel genetic variants of *KIR3DL2* and *PVR* involved in immunoregulatory interactions are associated with non-small cell lung cancer survival

Yufeng Wu[1,2,4], Sen Yang[1,2,4], Hongliang Liu[2,4], Sheng Luo[5], Thomas E Stinchcombe[2,3], Carolyn Glass[2,6], Li Su[7], Sipeng Shen[7], David C Christiani[7,8], Qiming Wang[1], Qingyi Wei[2,3,4]

[1]Department of Internal Medicine, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China; [2]Duke Cancer Institute, [3]Department of Medicine, Duke University Medical Center, Durham, NC 27710, USA; Departments of [4]Population Health Sciences, [5]Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC 27710, USA; [6]Department of Pathology, Duke University School of Medicine, Durham, NC 27710, USA; [7]Department of Environmental Health and Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA; [8]Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA

Received May 11, 2020; Accepted May 19, 2020; Epub June 1, 2020; Published June 15, 2020

Abstract: Immunoregulatory interactions play a pivotal role in immune surveillance, recognition, and killing, particularly its internal pathway, likely playing an important role in immune escape. By using two genotyping datasets, one from the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer screening trial (n = 1,185) as the discovery, and the other from Harvard Lung Cancer Susceptibility (HLCS) study (n = 984) as the validation, we evaluated associations between 4,713 genetic variants (338 genotyped and 4,375 imputed) in 60 genes involved in immunoregulatory interactions and survival of non-small cell lung cancer (NSCLC). We found that 115 SNPs were significantly associated with NSCLC overall survival in the discovery, of which four remained significant after validation by the HLCS dataset after multiple test correction by Bayesian false discovery probability. Final combined analysis identified two independent SNPs (*KIR3DL2* rs4487030 A>G and *PVR* rs35385129 C>A) that predicted NSCLC survival with a combined hazards ratio of 0.84 (95% confidence interval = 0.76-0.93, *P* = 0.001) and 0.84 (95% confidence interval = 0.73-0.97, *P* = 0.021), respectively. Besides, expression quantitative trait loci analyses showed that these two survival-associated SNPs of *KRI3DL2* and *PVR* were significantly associated with their mRNA expression levels in both normal lung tissues and whole blood cells. Additional analyses suggested an oncogenic role for *KRI3DL2* and a suppressor role for *PVR* on the survival. Once further validated, genetic variants of *KIR3DL2* and *PVR* may be potential prognostic markers for NSCLC survival.

Keywords: Non-small cell lung cancer, immunoregulatory interactions, single nucleotide polymorphism, survival

Introduction

Lung cancer is the most commonly diagnosed cancer (11.6% of the total cases) and the leading cause of cancer deaths (18.4% of the total cancer deaths) in the world [1]. Non-small cell lung cancer (NSCLC) is the most common histologic type of lung cancer, accounting for about 80-85% of the total cases [2]. In the past decades, chemotherapy, radiotherapy, and surgery are the main treatments for NSCLC until the emergence of targeted therapies and immunotherapeutic drugs. After many years of disappointing results, treatment choice has finally changed, and immunotherapy has become a clinically validated treatment for many cancers, including NSCLC [3]. Although the use of immunotherapy has a much-improved treatment response, the overall 5-year survival rate of NSCLC is still less than 15% [4, 5] and current research has focused on identifying additional factors that improve the effectiveness of immunotherapy. Studies have shown that genetic variants, single-nucleotide polymorphisms (SNPs) in particular as one of the most common genetic variations, have a significant impact on outcomes of cancer treatment [6, 7].

The idea of using the host immune system to combat cancer dates back to decades ago.

Researchers have realized that various components of the immune system play pivotal roles in protecting humans from cancer. Following numerous disappointing efforts and clinical failures, cancer immunotherapy is now widely considered the "fourth-pillar" of cancer therapy in addition to the other three conventional treatments [8]. Currently immune checkpoint inhibitors alone or in combination with chemotherapy are the standard first-line therapy for patients with metastatic non-small cell lung cancer. One of the representative tumor immunotherapy agents is the autologous cellular immunotherapy sipuleucel-T and anti-cytotoxic T lymphocyte-associated protein 4 antibody, also so-called ipilimumab, which was approved for prostate cancer treatment in 2010 [9], and the other is the anti-programmed cell death protein 1 antibody (PD1) for the treatment of melanoma approved in 2014 [10]. However, the proportion of cancer patients with indications for current immunotherapies remains small, and there are also many problems encountered in the immunotherapies, such as the screening of sensitive patients and, more importantly, dealing with immune drug resistance [11-14]. Although the immune system can differentiate protein structures at the molecular level, cancer cells manage to escape the host immune recognition and subsequent immune killing [15]. In the process of immunotherapy, the immune recognition of malignant cells exerts some selective pressure on the developing tumors, resulting in the growth of tumor cells evolving with low immunogenicity but a strong anti-apoptotic ability. In this case, immune escape occurs, leading to the failure of immunotherapy. Therefore, it is crucial to identify additional survival-related factors involved in immunoregulatory interactions.

Genetic variation, including SNPs, in some key genes in the signaling pathway for immunoregulatory interactions may be involved in the disorder or over-activation of the entire signaling pathway, modulating the effect of the immune system on tumor growth and progression. However, the roles of genetic variants in candidate genes involved in the immunoregulatory interaction signaling and their biological functions in tumor growth or progress remain unknown. As a promising hypothesis-driven method in the post-GWAS (genome-wide association study) era, the biological pathway-based approach has been applied to reanalyze published GWAS datasets and to test the cumulative effect of SNPs across multiple genes in the same biological pathway [16]. Therefore, we hypothesized that genetic variants in genes involved in the immunoregulatory interaction pathway are associated with survival of lung cancer patients. We tested this hypothesis by using genotyping data from two independently published GWAS datasets, focusing on those SNPs that may alter their gene functions and thus most likely have biological and functional consequences.

Materials and methods

Study populations

In the present study, we used the GWAS genotyping dataset from the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial as the discovery dataset. PLCO is a randomized control study conducted by the National Cancer Institute (NCI), involving 77500 men and 77500 women, aged between 55 and 74, registered in 10 medical centers in the United States from 1993 to 2011. All the participants were randomized into either the intervention arm that received a trial screening or the control arm that received standard care, were followed up for at least 13 years after the enrollment [17]. Blood samples and personal information including smoking status, histologic diagnosis, tumor stage, treatment method and family history were provided at enrollment [18]. A total of 1,185 NSCLC patients were eligible for survival analysis after excluding two individuals who had no follow-up information. Genomic DNA extracted from the whole blood samples of the participants were genotyped with Illumina HumanHap240Sv1.0 and HumanHap550v3.0 (dbGaP accession: phs000093.v2.p2 and phs000336.v1.p1) [19, 20]. The 1,185 NSCLC patients with both complete follow-up information and genotype data were used for survival analysis. Each institutional review board of the participating institutions had approved the collection and all the participants had provided a written informed consent permitting the PLCO trial and use of the collected data.

Another GWAS dataset of 984 histology-confirmed Caucasian NSCLC patients from the Harvard Lung Cancer Susceptibility (HLCS) Study which began in 1992 was used as the validation dataset [21]. In the HLCS study, the

whole blood samples and personal information were collected after diagnosis, and DNA from the blood samples was extracted with Auto Pure Large Sample nucleic acid purification system (QIAGEN Company, Venlo, Limburg, Netherlands) and genotyped by using the Illumina Humanhap610-Quad array. The genotyped data were used for imputation with the MACH software based on the sequencing data from the 1,000 Genomes Project [21].

The use of these two GWAS datasets was approved by both the Internal Review Board of Duke University School of Medicine (#Pro00054575) and the dbGaP database administration (#6404). The comparison of the characteristics between the PLCO trial (n = 1,185) and the HLCS study (n = 984) is presented in Table S1.

Gene and SNP selection

The genes involved in the immunoregulatory interaction signaling pathway were selected using the Molecular Signatures Database (http://software.broadinstitute.org/gsea/msigdb/index.jsp) with the keyword "immunoregulatory AND interactions". After the removal of five pseudogenes, four genomic positions unavailable, and one gene on the X chromosome, 60 genes remained as candidate genes for further analysis (Table S2). Imputation with IMPUTE2 and the 1,000 Genomes Project data (phase 3) was performed for additional SNPs untyped for these candidate genes with their ± 500-kb flanking regions. After that, SNPs within the genes and their ± 2 kb flanking regions were extracted with the following criteria: an imputation information score ≥0.8 (Figure S1), a genotyping rate ≥95%, a minor allelic frequency (MAF) ≥5%, and Hardy-Weinberg equilibrium (HWE) $\geq 1\times10^{-5}$. As a result, a total of 4,713 (338 genotyped and 4,375 imputed) SNPs were included in the PLCO dataset.

Statistical analyses

The endpoints for the analyses included overall survival (OS) and disease-specific survival (DSS). In the single-locus analysis, we first used multivariate Cox proportional hazards regression analysis to assess the association between each of the SNPs and NSCLC survival in an additive model using the PLCO dataset, with adjustment for clinical variables such as age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery and the first four principal components (Table S3) by using the GenABEL package of R software [22]. We then used Bayesian false discovery probability (BFDP) with a cut-off value of 0.80 for multiple testing correction to lower the probability of potentially false-positive results as recommended [23, 24], because most of the SNPs under investigation are in high LD as a result of imputation. We assigned a prior probability of 0.10 and detected an upper boundary hazard ratio (HR) of 3.0 for an association with variant genotypes or minor alleles of the SNPs with $P<0.05$. After that, we validated those chosen SNPs by using the HLCS dataset. Next, we performed an inverse variance weighted meta-analysis to combine the results of both discovery and validation datasets. In the meta-analysis, Cochran's Q-test and the heterogeneity statistic (I^2) were performed to assess the inter-study heterogeneity. If no heterogeneity was observed between the two datasets ($P_{het}>0.10$ and $I^2<50\%$), a fixed-effects model was implemented; otherwise, a random-effects model was applied. Furthermore, a multivariate stepwise Cox model, including the first four principal components of the PLCO dataset, available demographic and clinical variables was performed to identify novel and independent SNPs. Finally, the model was further adjusted for 15 previously published SNPs for the survival of NSCLC from the same PLCO GWAS dataset.

We then used the combined genotypes or alleles to evaluate the cumulative effects of the identified SNPs and the Kaplan-Meier (KM) survival curves to show the survival probability associated with the combined genotypes or alleles. We also assessed possible interactions with an χ^2-based Q-test between subgroups in the stratified analysis. We then performed the receiver operating characteristic (ROC) curve and time-dependent area under the curve (AUC) with the timeROC package of R software (version 3.5.0) to illustrate the prediction accuracy of the model integrating the effects of both clinical and genetic variables on NSCLC survival [25]. To evaluate the correlations between SNPs and the corresponding mRNA expression levels, we performed the expression quantitative trait loci (eQTL) analyses with a linear regression model performed with the R software. The mRNA expression data of genes were obtained from two sources: 373 European individuals included in the 1,000

Genomes Project as well as the data from whole blood cell samples from 589 subjects and normal lung tissue samples from 454 subjects included in the genotype-tissue expression (GTEx) project [26, 27]. The bioinformatics functional prediction for the tagging SNPs was then performed with SNPinfo [28] (https://snpinfo.niehs.nih.gov), RegulomeDB [29] (http://www.regulomedb.org) and HaploReg [30] (http://archive.broadinstitute.org/mammals/haploreg/haploreg.php).

Lastly, the differences in mRNA expression levels were examined in 111 pairs of lung cancer tissues and adjacent normal tissues from the Cancer Genome Atlas (TCGA) database by using a paired *t*-test. We also assessed the differences in mRNA expression levels in a larger, but not paired, dataset from TCGA (http://ualcan.path.uab.edu), and the KM survival analysis was performed to assess the association between the mRNA expression levels and survival probability (http://kmplot.com/analysis/index.php?p=service&cancer=lung). All statistical analyses were performed with a statistical significance level of P<0.05 and using the SAS software (version 9.4; SAS Institute, Cary, NC, USA) unless otherwise indicated.

Data availability

The datasets are available from the National Center for Biotechnology Information Database of Genotypes and Phenotypes (dbGaP Study Accession: phs000093.v2.p2 and phs000336.v1.p1). Genome-wide imputation was performed based on the 1000 Genomes Project, phase III CEU, utilizing the IMPUTE2 software (October 2014 release).

Results

Associations between SNPs in the immunoregulatory interaction pathway genes and NSCLC survival

The overall flowchart in **Figure 1** describes the design in the present study with genotyping data from two previously published GWAS datasets. The discovery genotyping dataset from the PLCO trial included 1,185 NSCLC patients, and the validation genotyping dataset from the HLCS GWAS study enrolled 984 NSCLC patients [31]. In the discovery PLCO dataset, 338 genotyped and 4,375 imputed SNPs in the 60 immunoregulatory interactions pathway genes with NSCLC after single-locus analysis (all imputation information score ≥0.8, Figure S1). As a result, we identified 115 SNPs to be significantly associated with NSCLC OS (P<0.05), with multiple testing correction by BFDP ≤0.80. After further validation by the HLCS genotyping dataset, four SNPs in three genes remained significant. In combined analysis of the PLCO and HLCS genotyping datasets, three of these four newly identified SNPs were associated with a better survival, including rs4487030 A>G of *KIR3DL2* (HR = 0.84, P<0.001) and rs35385129 C>A (HR = 0.83, P<0.001) and rs28411142 C>T (HR = 0.84, P<0.001) of *PVR*, but rs3788142 G>A of *ITGB2* was associated with a poor survival (HR = 1.16, P<0.001) (**Table 1**).

Independent SNPs associated with NSCLC survival in the PLCO dataset

Subsequently, these four SNPs were tested for their independence in the multivariate stepwise Cox regression model using the PLCO dataset (because the HLCS dataset did not have individual genotyping data). Two SNPs (*KIR3DL2* rs4487030 A>G) and (*PVR* rs35385129 C>A) remained significantly associated with a better survival (**Table 2**), after adjustment for other 15 previously reported survival-associated SNPs in the same PLCO GWAS dataset. As showed in Figure S2, the two SNPs from both PLCO and HLCS datasets are summarized in a Manhattan plot, respectively, and the regional association plot for each of these two SNPs is also shown in Figure S3.

In the PLCO dataset with complete adjustment for available covariates, patients with the protective *KIR3DL2* rs4487030 G allele (i.e., AG+GG) or the *PVR* rs35385129 A allele (i.e., CA+AA) had a better OS and DSS (P_{trend}<0.001 and P_{trend} = 0.003 for *KIR3DL2* rs4487030 G, respectively and P_{trend} = 0.010 and P_{trend} = 0.039 for *PVR* rs35385129 A, respectively) (**Table 3**). In comparison with the AA genotype, the *KIR3DL2* rs4487030 GG genotype was associated with a decreased risk of death (HR = 0.73, 95% CI = 0.54-0.82 and P<0.001 for OS and HR = 0.70, 95% CI = 0.56-0.87 and P = 0.001 for DSS), but the *KIR3DL2* rs4487030 AG genotype was associated with a non-significant

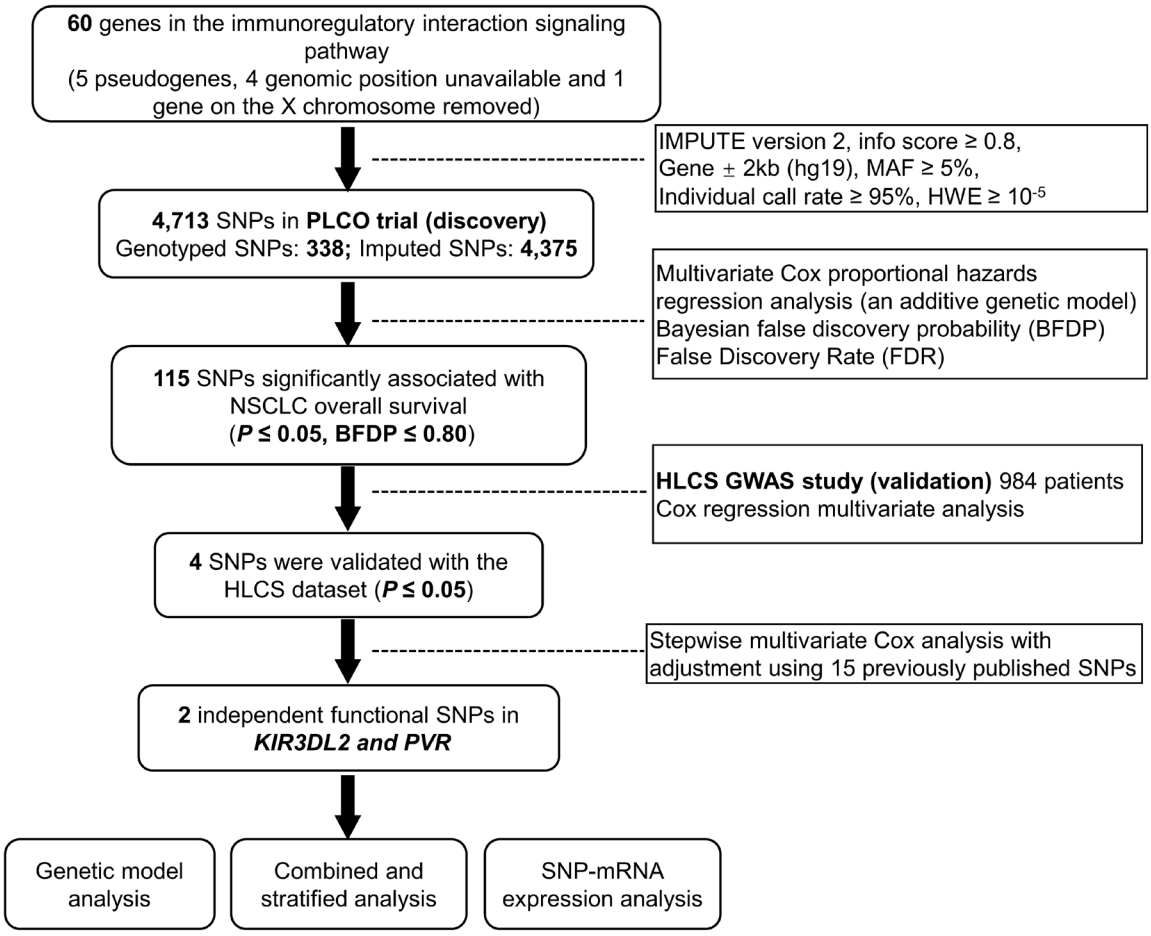


Figure 1. The flowchart of the present study. Abbreviations: SNP, single nucleotide polymorphism; PLCO, Prostate, Lung, Colorectal and Ovarian cancer screening trial; NSCLC, non-small cell lung cancer; HLCS, Harvard lung cancer susceptibility study; KIR3DL2: killer cell immunoglobulin like receptor, three Ig domains and long cytoplasmic tail 2; PVR: PVR cell adhesion molecule.

better survival (HR = 0.94, 95% CI = 0.80-1.11 and P = 0.033 for OS and HR = 0.84, 95% CI = 0.72-0.99 and P = 0.033 for DSS). Meanwhile, in comparison with the CC genotype, the *PVR* rs35385129 CA genotype was associated with a better significant better survival (HR = 0.84, 95% CI = 0.72-0.99 and P = 0.034 for OS and HR = 0.86, 95% CI = 0.73-1.02 and P = 0.079 for DSS), but the *PVR* rs35385129 AA genotype was associated with a non-significant survival (HR = 0.63, 95% CI = 0.35-1.12 and P = 0.112 for OS and HR = 0.70, 95% CI = 0.39-1.24 and P = 0.224 for DSS), likely due to a small observations for the AA genotype (**Table 3**).

Combined effects of the two independent SNPs in the PLCO dataset

Consequently, we also utilized the PLCO dataset to assess the combined effect of the two independent SNPs on NSCLC OS and DSS. First, we combined the significant protective genotypes (i.e., *KIR3DL2* rs4487030 AG+GG and *PVR* rs35385129 CA+AA) into a genetic score as the number of protective genotypes (NPGs). As shown in **Table 3**, the increased genetic score was associated with a better survival in the multivariate analysis of the PLCO dataset (P_{trend} = 0.001 for OS and P_{trend} = 0.003 for DSS). When we dichotomized all the patients into genetic scores of 0-1 and 2 NPGs, the score 2 group had a significantly better survival (HR = 0.72, 95% CI = 0.60-0.86 and P<0.001 for OS and HR = 0.74, 95% CI = 0.61-0.90 and P = 0.003 for DSS), in comparison with the score 0-1 group. We further plotted Kaplan-Meier survival curves to depict these associations of protective genotypes with NSCLC OS (**Figure 2A** and **2B**) and DSS (**Figure 2C** and **2D**).

Table 1. Associations between four validated significant SNPs and overall survival in both discovery and validation genotyping datasets from two previously published NSCLC GWASs

SNP	Allele[a]	Gene	PLCO (n = 1185)			Harvard (n = 984)			Combined-analysis			
			EAF	HR (95% CI)[b]	*P*[b]	EAF	HR (95% CI)[c]	*P*[c]	P_{het}[d]	I^2	HR (95% CI)[e]	*P*[e]
rs4487030	A>G	*KIR3DL2*	0.44	0.83 (0.75-0.92)	<0.001	0.41	0.85 (0.74-0.97)	0.018	0.828	0	0.84 (0.77-0.91)	<0.001
rs3788142	G>A	*ITGB2*	0.24	1.17 (1.04-1.32)	0.008	0.24	1.15 (1.01-1.32)	0.030	0.886	0	1.16 (1.07-1.27)	<0.001
rs35385129	C>A	*PVR*	0.16	0.83 (0.72-0.96)	0.010	0.15	0.82 (0.70-0.97)	0.018	0.956	0	0.83 (0.74-0.92)	<0.001
rs28411142	C>T	*PVR*	0.16	0.83 (0.72-0.96)	0.012	0.15	0.85 (0.72-0.99)	0.037	0.862	0	0.84 (0.75-0.93)	0.001

[a]Major allele > minor allele; [b]Adjusted for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3 and PC4; [c]Adjusted for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2 and PC3; [d]P_{het}: *P* value for heterogeneity by Cochrane's Q test; [e]Meta-analysis in the fix-effects model; Abbreviations: EAF: effect allele frequency; HR: hazards ratio; CI: confidence interval; PLCO: Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; HLCS: Harvard Lung Cancer Susceptibility Study.

Table 2. Three indenpenden SNPs associated with OS in multivariate Cox proportional hazards regression analysis with adjustment for other covariates and previously published SNPs in the PLCO GWAS dataset

Variables[a]	Category	Frequency	HR (95% CI)[a]	*P*[a]	HR (95% CI)[b]	*P*[b]
Age	Continuous	1185	1.03 (1.02-1.05)	<0.001	1.04 (1.02-1.05)	<0.001
Sex	Male	698	1.00		1.00	
	Female	487	0.77 (0.67-0.90)	0.001	0.78 (0.66-0.91)	0.002
Smoking status	Never	115	1.00		1.00	
	Current	423	1.64 (1.22-2.19)	0.001	1.86 (1.37-2.50)	<0.001
	Former	647	1.62 (1.23-2.14)	0.001	1.83 (1.38-2.43)	<0.001
Histology	AD	577	1.00		1.00	
	SC	285	1.16 (0.97-1.40)	0.112	1.21 (1.00-1.47)	0.048
	others	323	1.31 (1.10-1.55)	0.002	1.35 (1.13-1.61)	0.001
Tumor stage	I-IIIA	655	1.00		1.00	
	IIIB-IV	528	2.89 (2.37-3.51)	<0.001	3.05 (2.50-3.72)	<0.001
Chemotherapy	No	639	1.00		1.00	
	Yes	538	0.56 (0.47-0.67)	<0.001	0.57 (0.47-0.68)	<0.001
Radiotherapy	No	762	1.00		1.00	
	Yes	415	0.98 (0.83-1.15)	0.786	0.96 (0.81-1.14)	0.628
Surgery	No	637	1.00		1.00	
	Yes	540	0.21 (0.16-0.27)	<0.001	0.20 (0.15-0.26)	<0.001
KIR3DL2 rs4487030 A>G	AA/GA/GG	381/560/244	0.84 (0.76-0.92)	**<0.001**	0.84 (0.76-0.94)	**0.001**
PVR rs35385129 C>A	CC/CA/AA	826/334/25	0.82 (0.72-0.95)	**0.008**	0.84 (0.73-0.97)	**0.021**

[a]Stepwise analysis included age, sex, smoking status, tumor stage, tumor histology, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3, PC4, and two newly validated SNPs in an additive model; *P* values for significant SNPs were in bold. [b]Fifteen previously published SNPs were used for the 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 Screening Trial; GWAS: genome-wide association study; HR: hazards ratio; CI: confidence interval; AD: Adenocarcinoma; SC: Squamous cell carcinoma.

Stratified analysis for associations between NPGs and NSCLC survival

To evaluate whether the combined effects of protective alleles were modified by other clinical variables, we performed stratified analysis by age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy and surgery in the PLCO dataset. As showed in Table S4, we found that the smoke status and chemotherapy had a significant effect on DSS. Patients with two protective genotypes had a better DSS (HR = 0.62, 95% CI = 0.48-0.80 and *P*<0.001) than those with 0-1 protective geno-

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

Genotype	Frequency	OS[a]			DSS[a]		
		Death (%)	HR (95% CI)	*P**	Death (%)	HR (95% CI)	*P**
KIR3DL2 rs4487030 A>G[b]							
AA	376	264 (70.21)	1.00		234 (62.23)	1.00	
AG	557	377 (67.68)	0.94 (0.80-1.11)	0.486	342 (61.40)	0.98 (0.82-1.16)	0.789
GG	242	148 (61.16)	0.73 (0.54-0.82)	**<0.001***	133 (54.96)	0.70 (0.56-0.87)	**0.001***
Trend				**<0.001***			**0.003***
AG+GG	799	525 (65.71)	0.85 (0.73-0.99)	0.034*	475 (59.45)	0.88 (0.75-1.04)	0.125
PVR rs35385129 C>A[b]							
CC	820	559 (68.17)	1.00		500 (60.98)	1.00	
CA	330	218 (66.06)	0.84 (0.72-0.99)	**0.034***	197 (59.70)	0.86 (0.73-1.02)	0.079
AA	25	12 (48.00)	0.63 (0.35-1.12)	0.112	12 (48.00)	0.70 (0.39-1.24)	0.224
Trend				**0.010***			**0.039***
CA+AA	355	230 (64.79)	0.83 (0.71-0.97)	**0.016***	209 (58.87)	0.85 (0.72-1.00)	**0.049***
NPG[1b,c]							
0	254	178 (70.08)	1.00		155 (61.02)	1.00	
1	688	467 (67.88)	0.95 (0.79-1.13)	0.542	424 (61.63)	0.99 (0.82-1.19)	0.884
2	233	144 (61.80)	0.69 (0.55-0.86)	**0.001***	130 (55.79)	0.73 (0.58-0.93)	**0.012***
Trend				**0.001***			**0.013***
0-1	942	645 (68.47)	1.00		579 (61.46)		
2	233	144 (61.80)	0.72 (0.60-0.86)	**<0.001***	130 (55.79)	0.74 (0.61-0.90)	**0.003***

[a]Adjusted for age, sex, smoking status, histology, tumor stage, chemotherapy, surgery and principal component. [b]10 missing date were excluded; [c]protective genotypes were *KIR3DL2* rs4487030 AG+GG and *PVR* rs35385129 CA+AA; *P values for significant genotypes were in bold. Abbreviations: 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; NPG: number of protective genotypes.

types in former-smoke patients (P_{inter} = 0.045). Meanwhile, patients who received chemotherapy had a better OS (HR = 0.61, 95% CI = 0.46-0.81 and *P*<0.001) and DSS (HR = 0.63, 95% CI = 0.46-0.86 and *P*<0.003) than those who did not receive chemotherapy (P_{inter} = 0.047 and 0.039, respectively).

The ROC curves and time-dependent AUC

We further assessed predictive value of the two SNPs with the time-dependent AUC and ROC curves for the five-year survival in the PLCO dataset. Compared with the model of covariates including age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery and the first four principal components, the time-dependent AUC plot with an additional two independent SNPs did not improve prediction performance of the model for five-year OS and DSS (Figure S4A-D). On the other hand, when we calculated the time-dependent AUC and ROC curves for the 5-year (at the 60th month) survival and the combine two independent SNPs together with the previously published 15 SNPs, the prediction performance of the model for OS and DSS has been significantly improved. The AUCs changed from 88.35% to 89.71% (*P* = 0.010) for OS and from 88.18% to 89.64% (*P* = 0.014) for DSS (Figure S4F and S4H).

The eQTL analysis

To assess the correlations between two independent SNPs and their corresponding mRNA expression levels, we firstly performed the eQTL analysis by using the data of 589 whole blood samples and 454 normal lung tissues from the GTEx project. We found that the rs4487030 G allele was significantly associated with a decreased mRNA expression level of *KIR3DL2* in normal lung tissues ($P = 4.3\times10^{-7}$) (**Figure 3A**) and whole blood samples ($P = 4.6\times10^{-12}$) (**Figure 3B**), while the rs35385129 A allele also associated with an increased

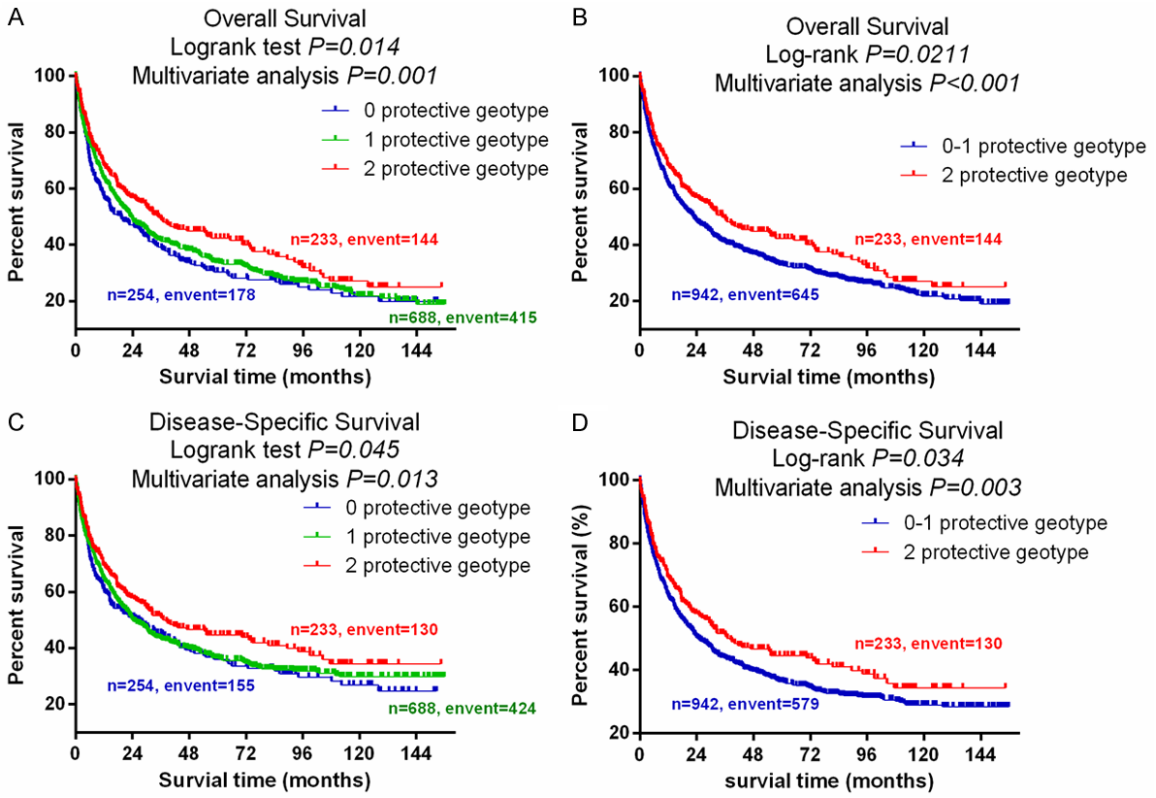


Figure 2. Prediction of survival with combined protective genotypes. (A) Kaplan-Meier survival curves for the overall survival of the combined protective genotypes (*KIR3DL2* rs4487030 AG+GG and *PVR* rs35385129 CA+AA) and (B) dichotomized groups of the NUG in the PLCO dataset; (C) Kaplan-Meier survival curves for the disease-specific survival of the combined protective genotypes and (D) dichotomized groups of the NUG in the PLCO dataset. Abbreviations: NPG, number of protective genotypes; PLCO, The Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.

mRNA expression level of *PVR* both in normal lung tissues ($P = 7.4\times10^{-6}$) (**Figure 3D**) and whole blood samples ($P = 7.6\times10^{-5}$) (**Figure 3E**).

Additional eQTL analysis was performed for the *PVR* rs35385129 A allele using the RNA-Seq data of lymphoblastoid cell lines from 373 European descendants available from the 1000 Genomes Project (but no data for the rs4487030 G allele). However, there was no significant correlation between the *PVR* rs35385129 A allele and mRNA expression levels of *PVR* in all three genetic models (Figure S7A).

Finally, we performed functional prediction for rs4487030 and rs35385129 as well as other SNPs in high LD, utilizing three different online bioinformatics tools, SNPinfo, RegulomeDB, and HaploReg, to predict their biological functions. In summary, rs35385129 may cause missense mutations, with some substantial functions based on RegulomeDB and HaploReg, particularly in the transcription factor binding or DNase peak. Moreover, rs35385129 is in high LD with other SNPs that may have an effect on other genes and proteins (Table S5). On the other hand, rs4487030 may have an impact on proteins CDP and p300 based on HaploReg without predicted functions based on RegulomeDB. The detailed results are summarized in Figure S6 and Table S5.

Differential mRNA expression analysis and survival of NSCLC

Subsequently, we assessed mRNA expression levels of the two genes identified by the SNPs in 111 pairs of tumor and adjacent normal tissue samples obtained from NSCLC patients in the TCGA database and non-paired tumor and normal tissue samples from the UALCAN database (http://ualcan.path.uab.edu/). For non-paired tumor tissues, as shown in Figure S5,

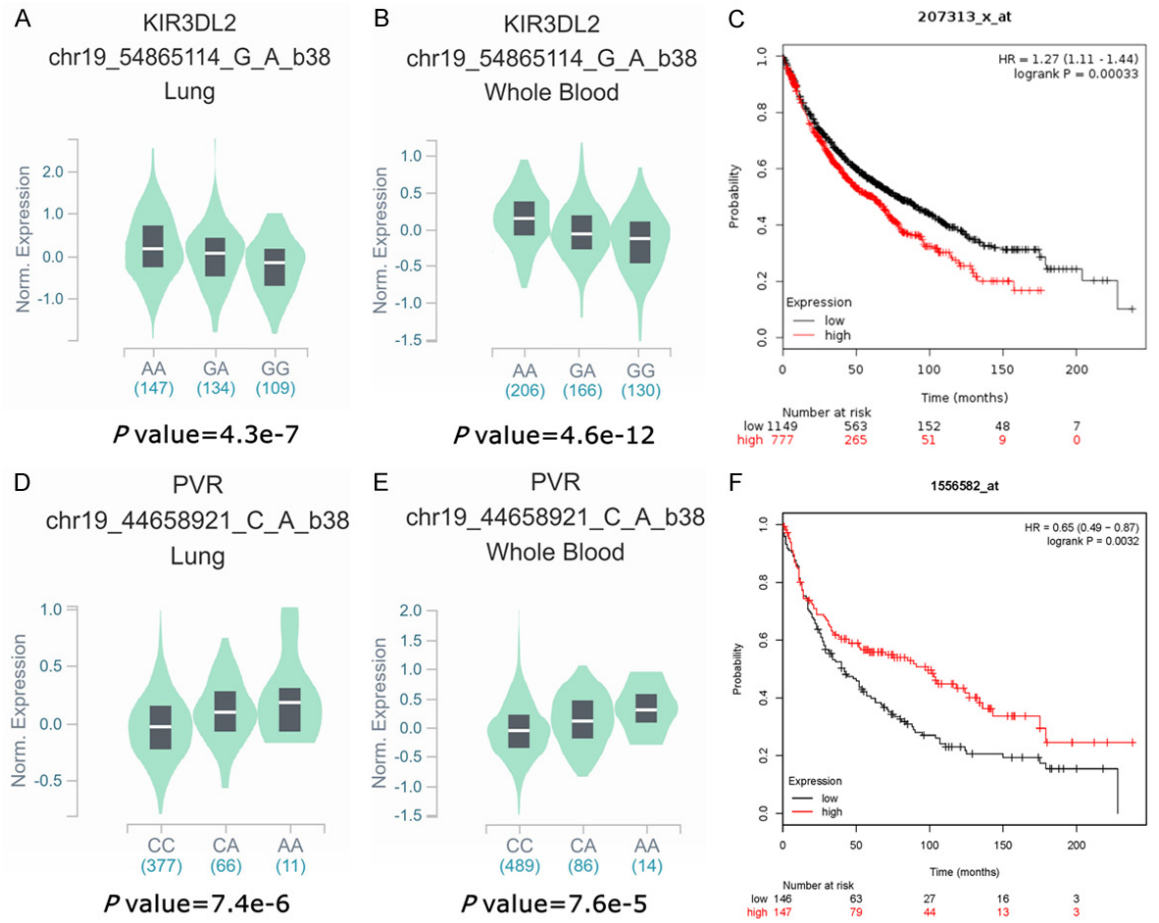


Figure 3. Correlations between genotypes of the significant SNPs and their corresponding mRNA expression levels. The *KIR3DL2* rs4487030 A allele was associated with higher mRNA expression levels of *KIR3DL2* (A) in normal lung tissue and (B) whole blood from the GTEx project; higher expression levels of *KIR3DL2* (C) were associated with a worse survival in patients with lung cancer; The *PVR* rs35385129 A allele was associated with higher mRNA expression levels of *PVR* (D) in normal lung tissue and (E) whole blood from the GTEx project; higher expression levels of *PVR* (F) were associated with a better survival in patients with lung cancer (GSE30219 dataset). Abbreviations: KIR3DL2: killer cell immunoglobulin like receptor, three Ig domains and long cytoplasmic tail 2; PVR: PVR cell adhesion molecule.

mRNA expression levels of *KIR3DL2* in lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) were significantly lower than that in adjacent normal tissues (P< 0.0001 for both), while the mRNA expression levels of *PVR* were also significantly lower in tumor tissues in LUSC, but not in LUAD, than adjacent normal tissues (P = 0.005 and P = 0.478, respectively) (Figure S7C). For the paired tumor samples, as shown in Figure S7B, compared with adjacent normal tissues, the mRNA expression levels of *PVR* were significantly lower in LUSC (P = 0.0001) and LUAD+ LUSC (P = 0.0005) than adjacent normal tissues, but not in LUAD alone. Subsequently, we analyzed the correlation between the expression levels of *KIR3DL2* and *PVR* and the survival curve of lung cancer by online KM plotter (http://km-plot.com/analysis). As showed in **Figure 3C**, a higher expression level of *KIR3DL2* was associated with a worse NSCLC survival, but the results of *PVR* were inconsistent in different datasets of the online KM survival curve plotter (Figure S8). In GSE30219 and GSE19-188 datasets, *PVR* was associated with a better survival but with a poor survival in GSE31-210 and All-combined databases (**Figure 3F**).

Discussion

In the present study, we used the existing genotyping data of two previously published

GWASs to analyze the association between genetic variation of immunoregulatory interaction genes and survival of NSCLC. We found and verified two potentially functional and independent SNPs (i.e., *KIR3DL2* rs4487030 A>G and *PVR* rs35385129 C>A) that were significantly associated with survival of NSCLC in populations of European descendants. Our eQTL analysis found that the rs4487030 G and rs35385129 A alleles were significantly correlated with mRNA expression levels of *KIR3DL2* and *PVR* in normal lung tissues and whole blood cells from 369 subjects in the GTEx program, respectively (**Figure 3**). These results were also consistent with the gene expression analysis of paired tumor and adjacent normal tissue samples and survival analysis in the TCGA database.

KIR3DL2 is a member of a killer Ig-like receptors (KIRs), also known as CD158k, and is expressed as a disulfide bond-linked homodimer. Each chain consists of three immunoglobulin-like domains and a long cytoplasmic tail containing two tyrosine inhibitory motifs of immune receptors. At the genomic level, KIRs are located in the leukocyte receptor cluster of chromosomes 19q13.4, while the *Ly49* gene of rodents is located in the NK complex of chromosome 6. Human haplotypes encode the *KIR* content, and their polymorphic alleles are very different, as many as 14 genes and three frame genes, namely *KIR2DL4, KIR3DL2* and *KIR3DL3* [32]. KIR3DL2 is expressed not only in NK cells but also in rare circulating T lymphocytes, mainly CD8+ [33]. An enriched KIR3DL2 expression in the memory CD45RO+ CD28-CCR7-CD62L-T cells has also been reported. For example, KIR3DL2 is capable, upon the ligation at the surface of NK cells, of inhibiting IFNγ production and cytotoxic function [34], which disrupts the tumor-killing capacity of the immune system.

In the present study, we found that the rs4487030 G allele was significantly associated with lower mRNA expression levels of *KIR3DL2* in normal lung tissues and whole blood cells (**Figure 3**); this finding is consistent with the results from other studies [34], suggesting that the rs4487030 G allele may prolong survival of NSCLC patients through decreasing *KIR3DL2* mRNA expression levels. Moreover, mRNA of *KIR3DL2* expression levels were lower in the non-paired tumor tissues in both LUSC and LUAD than in normal tissues. The observation of low expression levels of *KIR3DL2* associated with a poor survival in NSCLC patients suggested an oncogenic role of the gene. There are two explanations for this observation; one is that the inflammatory reaction and immune killing activity in tumor tissues are more significant than in adjacent normal tissues, so a lower *KIR3DL2* expression level is the consequence rather than the cause. On the other hand, normal tissues have a complete immune escape. It is evident that immune evasion in tumor tissues has occurred, likely as a result of low *KIR3DL2* expression in tumor microenvironment. According to the ENCODE database, rs4487030 is located in a DNase I hypersensitive site with observable levels of histone modifications in H3K4Me1 acetylation, suggesting that the rs4487030 SNP may lead to altered transcriptional activities of *KIR3DL2*. Therefore, once our findings are validated by other investigators, the exact molecular mechanisms underlying the observed *KIR3DL2* rs4487030 G allele-associated survival warrant additional experimental and mechanistic studies.

PVR was initially identified as a poliovirus receptor, with conserved amino acids and domain structure characteristically similar to the immunoglobulin superfamily [35]. PVR is the fifth member in the nectin-like molecule family, also referred to as CD155 or necl-5. Nectin-like molecule family, which has a domain structure similar to the nectins, plays a crucial role in cell adhesion and polarization, and *PVR* may have a similar function. However, many studies have investigated the biological role of *PVR*, but the results were inconsistent. For example, it has been reported that overexpression of *PVR* can promote tumor cell metastasis [36, 37], stimulate cell proliferation, or enhance cell proliferation induced by growth factor [38]. In fact, overexpression of *PVR* can also reduce contact inhibition by decreasing the expression of nectin-3 [39]. In animal experiments, $PVR^{-/-}$ mice displayed reduced tumor growth and metastasis via DNAM-1 upregulation and enhanced effector function of CD8+ T and NK cells, respectively, while *PVR*-deleted tumor cells also displayed slower tumor growth and reduced metastases, demonstrating the importance of a tumor-intrinsic

role of *PVR* [40]. On the other hand, *PVR* seems to play a dual function in onco-immunity and is considered to have the characteristics of a tumor suppressor gene. For example, PVR is the ligand for CD226, and interaction of CD226 with PVR triggers NK or T cell-mediated cytotoxicity, accompanied by an increase in cytokine production [41, 42]. Further *in vitro* studies showed that tumor cells with a higher *PVR* expression were more susceptible to the CD226-induced killing [43, 44]. Accordingly, the balance between PVR/CD226 and PVR/TIGIT or PVR/CD96 maintains normal NK and T cell function. However, this balance may be destroyed in the tumor microenvironment, affecting the progression of the tumor. According to the ENCODE database, rs35385129 is located in both H3K4Me1 and H3K27Ac, suggesting that rs35385129 may lead to an altered transcription, too.

In the present study, we found that the rs35385129 A allele was associated with higher mRNA expression levels of *PVR* in normal lung tissues and whole blood cells (**Figure 3**). This is consistent with the results of rs35385129 A allele-associated survival in both PLCO and HLCS datasets. Published studies reported that *PVR* expression is not detectable in most normal tissues but can be up-regulated in a series of human malignant tumors, including colon cancer, lung adenocarcinoma, melanoma, pancreatic cancer and glioblastoma [45]. This is inconsistent with the mRNA expression results in the TCGA database, including paired and non-paired samples. It is likely that these reported studies may have included different study populations or tumor tissues from those of the TCGA database. Apparently, the function of *PVR* is still controversial, and the reported effects of *PVR* on the survival of NSCLC are different as well. However, the findings of the present study provide additional support for a tumor suppressor role of *PVR*.

Although the associations between SNPs in the immunoregulatory interaction pathway genes had been comprehensively analyzed in the present study, it should also be mentioned that the present study has some limitations. Firstly, both the discovery and validation datasets were from Caucasian populations; thus, our findings may not be generalizable to other ethnic communities. Secondly, we did not have any information on the detailed treatment, particularly for immunotherapies. Additionally, we only analyzed the associations of genetic variants in the identified genes with the survival, and the complicated molecular mechanisms that underly these observed associations should be further explored. Finally, the role of viral infection in tumor tissues and their implications on the survival was not evaluated because the related data were not made available.

In conclusion, two independent SNPs, *KIR3DL2* rs4487030 A>G and *PVR* rs35385129 C>A, were found to be significantly associated with NSCLC survival in both the PLCO trial and the HLCS study. The combined analysis revealed that the protective genotypes of the SNPs were associated with a better OS and DSS in a genotype-dose response manner. Such protective effect on survival is likely through the SNP-associated expression regulation of *KIR3DL2* and *PVR*. Additional studies are needed to substantiate our findings.

Acknowledgements

The authors thank all the participants of the PLCO Cancer Screening Trial. The authors also thank the National Cancer Institute for providing 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 the dbGaP repository for providing cancer genotyping datasets. The accession numbers for the datasets for 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). Sheng Luo was supported by NIH (grants R01NS091307, R56AG062302). The Harvard Lung Cancer Susceptibility Study was supported by NIH grants U01CA209414, CA092824, CA074386 and CA090578 to David C. Christiani. 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 the present study: The datasets used for the analyses described in the present study 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] (Z01CP010200). 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.

Disclosure of conflict of interest

None.

Abbreviations

NSCLC, Non-small cell lung cancer; SNPs, single nucleotide polymorphisms; GWAS, Genome-Wide Association Study; PLCO, the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; HLCS, Harvard Lung Cancer Susceptibility; OS, overall survival; LD, linkage disequilibrium; FDR, false discovery rate; BFDP, Bayesian false discovery probability; eQTL, expression quantitative trait loci; TCGA, the Cancer Genome Atlas; ROC, receiver operating characteristic; KIR3DL2, killer cell immunoglobulin like receptor, three Ig domains and long cytoplasmic tail 2; PVR, PVR cell adhesion molecule; EAF, effect allele frequency; HR, hazards ratio; CI, confidence interval; AUC, area under the receiver operating characteristic curve; ROC, receiver operating characteristic curve.

Address correspondence to: Qiming Wang, Department of Internal Medicine, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou 450008, China. Tel: +86-0375-65588421; E-mail: qimingwang1006@126.com; Qingyi Wei, Duke Cancer Institute, Duke University Medical Center, Durham, NC 27710, USA; Department of Population Health Sciences, Duke University School of Medicine, 905 S LaSalle Street, Durham, NC 27710, USA. Tel: 919-660-0562; E-mail: qingyi.wei@duke.edu

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. 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] Blandin Knight S, Crosbie PA, Balata H, Chudziak J, Hussell T and Dive C. Progress and prospects of early detection in lung cancer. Open Biol 2017; 7: 170070.

[3] Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ Jr, Wu YL and Paz-Ares L. Lung cancer: current therapies and new targeted treatments. Lancet 2017; 389: 299-311.

[4] Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, Powderly JD, Sosman JA, Atkins MB, Leming PD, Spigel DR, Antonia SJ, Drilon A, Wolchok JD, Carvajal RD, McHenry MB, Hosein F, Harbison CT, Grosso JF and Sznol M. Five-year survival and correlates among patients with advanced melanoma, renal cell carcinoma, or non-small cell lung cancer treated with nivolumab. JAMA Oncol 2019; 5: 1411-20.

[5] Garon EB, Hellmann MD, Rizvi NA, Carcereny E, Leighl NB, Ahn MJ, Eder JP, Balmanoukian AS, Aggarwal C, Horn L, Patnaik A, Gubens M, Ramalingam SS, Felip E, Goldman JW, Scalzo C, Jensen E, Kush DA and Hui R. Five-year over-

all survival for patients with advanced nonsmall-cell lung cancer treated with pembrolizumab: results from the phase I KEYNOTE-001 study. J Clin Oncol 2019; 37: 2518-2527.

[6] Feng Y, Wang Y, Liu H, Liu Z, Mills C, Owzar K, Xie J, Han Y, Qian DC, Hung Rj RJ, Brhane Y, McLaughlin J, Brennan P, Bickeböller H, Rosenberger A, Houlston RS, Caporaso N, Landi MT, Brüske I, Risch A, Ye Y, Wu X, Christiani DC, Amos CI and Wei Q. Novel genetic variants in the P38MAPK pathway gene ZAK and susceptibility to lung cancer. Mol Carcinog 2018; 57: 216-224.

[7] Wen J, Liu H, Wang L, Wang X, Gu N, Liu Z, Xu T, Gomez DR, Komaki R, Liao Z and Wei Q. Potentially functional variants of ATG16L2 predict radiation pneumonitis and outcomes in patients with non-small cell lung cancer after definitive radiotherapy. J Thorac Oncol 2018; 13: 660-675.

[8] Mizukoshi E and Kaneko S. Telomerase-targeted cancer immunotherapy. Int J Mol Sci 2019; 20: 1823.

[9] Topalian SL, Weiner GJ and Pardoll DM. Cancer immunotherapy comes of age. J Clin Oncol 2011; 29: 4828-4836.

[10] Sharma P and Allison JP. Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential. Cell 2015; 161: 205-214.

[11] Reck M, Rodriguez-Abreu D, Robinson AG, Hui R, Csoszi T, Fulop A, Gottfried M, Peled N, Tafreshi A, Cuffe S, O'Brien M, Rao S, Hotta K, Leiby MA, Lubiniecki GM, Shentu Y, Rangwala R and Brahmer JR; KEYNOTE-024 Investigators. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. N Engl J Med 2016; 375: 1823-1833.

[12] Gandhi L, Rodriguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, Domine M, Clingan P, Hochmair MJ, Powell SF, Cheng SY, Bischoff HG, Peled N, Grossi F, Jennens RR, Reck M, Hui R, Garon EB, Boyer M, Rubio-Viqueira B, Novello S, Kurata T, Gray JE, Vida J, Wei Z, Yang J, Raftopoulos H, Pietanza MC and Garassino MC; KEYNOTE-189 Investigators. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. N Engl J Med 2018; 378: 2078-2092.

[13] Hellmann MD, Paz-Ares L, Bernabe Caro R, Zurawski B, Kim SW, Carcereny Costa E, Park K, Alexandru A, Lupinacci L, de la Mora Jimenez E, Sakai H, Albert I, Vergnenegre A, Peters S, Syrigos K, Barlesi F, Reck M, Borghaei H, Brahmer JR, O'Byrne KJ, Geese WJ, Bhagavatheeswaran P, Rabindran SK, Kasinathan RS, Nathan FE and Ramalingam SS. Nivolumab plus ipilimumab in advanced non-small-cell lung cancer. N Engl J Med 2019; 381: 2020-2031.

[14] Socinski MA, Jotte RM, Cappuzzo F, Orlandi F, Stroyakovskiy D, Nogami N, Rodriguez-Abreu D, Moro-Sibilot D, Thomas CA, Barlesi F, Finley G, Kelsch C, Lee A, Coleman S, Deng Y, Shen Y, Kowanetz M, Lopez-Chavez A, Sandler A and Reck M; IMpower150 Study Group. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. N Engl J Med 2018; 378: 2288-2301.

[15] Wang RF and Wang HY. Immune targets and neoantigens for cancer immunotherapy and precision medicine. Cell Res 2017; 27: 11-37.

[16] Gallagher MD and Chen-Plotkin AS. The post-GWAS era: from association to function. Am J Hum Genet 2018; 102: 717-730.

[17] Hocking WG, Hu P, Oken MM, Winslow SD, Kvale PA, Prorok PC, Ragard LR, Commins J, Lynch DA, Andriole GL, Buys SS, Fouad MN, Fuhrman CR, Isaacs C, Yokochi LA, Riley TL, Pinsky PF, Gohagan JK and Berg CD; PLCO Project Team. Lung cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. J Natl Cancer Inst 2010; 102: 722-731.

[18] Oken MM, Marcus PM, Hu P, Beck TM, Hocking W, Kvale PA, Cordes J, Riley TL, Winslow SD, Peace S, Levin DL, Prorok PC and Gohagan JK; PLCO Project Team. Baseline chest radiograph for lung cancer detection in the randomized Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. J Natl Cancer Inst 2005; 97: 1832-1839.

[19] Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M, Kholodov M, Zbicz K, Beck J, Kimelman M, Shevelev S, Preuss D, Yaschenko E, Graeff A, Ostell J and Sherry ST. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet 2007; 39: 1181-1186.

[20] Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, Lee M, Popova N, Sharopova N, Kimura M and Feolo M. NCBI's database of genotypes and phenotypes: dbGaP. Nucleic Acids Res 2014; 42: D975-979.

[21] Zhai R, Yu X, Wei Y, Su L and Christiani DC. Smoking and smoking cessation in relation to the development of co-existing non-small cell lung cancer with chronic obstructive pulmonary disease. Int J Cancer 2014; 134: 961-970.

[22] Aulchenko YS, Ripke S, Isaacs A and van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics 2007; 23: 1294-1296.

[23] Park JH, Geum DI, Eisenhut M, van der Vliet HJ and Shin JI. Bayesian statistical methods in genetic association studies: empirical examination of statistically non-significant Genome

Wide Association Study (GWAS) meta-analyses in cancers: a systematic review. Gene 2019; 685: 170-178.
[24] Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. Am J Hum Genet 2007; 81: 208-227.
[25] Chambless LE and Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. Stat Med 2006; 25: 3474-3486.
[26] Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015; 348: 648-660.
[27] Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T, Ferreira PG, Barann M, Wieland T, Greger L, van Iterson M, Almlof J, Ribeca P, Pulyakhina I, Esser D, Giger T, Tikhonov A, Sultan M, Bertier G, MacArthur DG, Lek M, Lizano E, Buermans HP, Padioleau I, Schwarzmayr T, Karlberg O, Ongen H, Kilpinen H, Beltran S, Gut M, Kahlem K, Amstislavskiy V, Stegle O, Pirinen M, Montgomery SB, Donnelly P, McCarthy MI, Flicek P, Strom TM, Geuvadis C, Lehrach H, Schreiber S, Sudbrak R, Carracedo A, Antonarakis SE, Hasler R, Syvanen AC, van Ommen GJ, Brazma A, Meitinger T, Rosenstiel P, Guigo R, Gut IG, Estivill X and Dermitzakis ET. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013; 501: 506-511.
[28] Xu Z and Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res 2009; 37: W600-605.
[29] Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM and Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012; 22: 1790-1797.
[30] Ward LD and Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res 2016; 44: D877-881.
[31] Wang Y, Liu H, Ready NE, Su L, Wei Y, Christiani DC and Wei Q. Genetic variants in ABCG1 are associated with survival of nonsmall-cell lung cancer patients. Int J Cancer 2016; 138: 2592-2601.
[32] Wende H, Colonna M, Ziegler A and Volz A. Organization of the leukocyte receptor cluster (LRC) on human chromosome 19q13.4. Mamm Genome 1999; 10: 154-160.
[33] Chan AT, Kollnberger SD, Wedderburn LR and Bowness P. Expansion and enhanced survival of natural killer cells expressing the killer immunoglobulin-like receptor KIR3DL2 in spondylarthritis. Arthritis Rheum 2005; 52: 3586-3595.
[34] Schmitt C, Marie-Cardine A and Bensussan A. Therapeutic antibodies to KIR3DL2 and other target antigens on cutaneous T-cell lymphomas. Front Immunol 2017; 8: 1010.
[35] Gao J, Zheng Q, Xin N, Wang W and Zhao C. CD155, an onco-immunologic molecule in human tumors. Cancer Sci 2017; 108: 1934-1938.
[36] Ikeda W, Kakunaga S, Itoh S, Shingai T, Takekuni K, Satoh K, Inoue Y, Hamaguchi A, Morimoto K, Takeuchi M, Imai T and Takai Y. Tage4/Nectin-like molecule-5 heterophilically trans-interacts with cell adhesion molecule Nectin-3 and enhances cell migration. J Biol Chem 2003; 278: 28167-28172.
[37] Ikeda W, Kakunaga S, Takekuni K, Shingai T, Satoh K, Morimoto K, Takeuchi M, Imai T and Takai Y. Nectin-like molecule-5/Tage4 enhances cell migration in an integrin-dependent, Nectin-3-independent manner. J Biol Chem 2004; 279: 18015-18025.
[38] Kakunaga S, Ikeda W, Shingai T, Fujito T, Yamada A, Minami Y, Imai T and Takai Y. Enhancement of serum- and platelet-derived growth factor-induced cell proliferation by Necl-5/Tage4/poliovirus receptor/CD155 through the Ras-Raf-MEK-ERK signaling. J Biol Chem 2004; 279: 36419-36425.
[39] Minami Y, Ikeda W, Kajita M, Fujito T, Monden M and Takai Y. Involvement of up-regulated Necl-5/Tage4/PVR/CD155 in the loss of contact inhibition in transformed NIH3T3 cells. Biochem Biophys Res Commun 2007; 352: 856-860.
[40] Li XY, Das I, Lepletier A, Addala V, Bald T, Stannard K, Barkauskas D, Liu J, Aguilera AR, Takeda K, Braun M, Nakamura K, Jacquelin S, Lane SW, Teng MW, Dougall WC and Smyth MJ. CD155 loss enhances tumor suppression via combined host and tumor-intrinsic mechanisms. J Clin Invest 2018; 128: 2613-2625.
[41] Bottino C, Castriconi R, Pende D, Rivera P, Nanni M, Carnemolla B, Cantoni C, Grassi J, Marcenaro S, Reymond N, Vitale M, Moretta L, Lopez M and Moretta A. Identification of PVR (CD155) and Nectin-2 (CD112) as cell surface ligands for the human DNAM-1 (CD226) activating molecule. J Exp Med 2003; 198: 557-567.
[42] Tahara-Hanaoka S, Shibuya K, Kai H, Miyamoto A, Morikawa Y, Ohkochi N, Honda S and Shibuya A. Tumor rejection by the poliovirus receptor family ligands of the DNAM-1 (CD226) receptor. Blood 2006; 107: 1491-1496.
[43] Castriconi R, Dondero A, Corrias MV, Lanino E, Pende D, Moretta L, Bottino C and Moretta A.

Natural killer cell-mediated killing of freshly isolated neuroblastoma cells: critical role of DNAX accessory molecule-1-poliovirus receptor interaction. Cancer Res 2004; 64: 9180-9184.

[44] Pende D, Spaggiari GM, Marcenaro S, Martini S, Rivera P, Capobianco A, Falco M, Lanino E, Pierri I, Zambello R, Bacigalupo A, Mingari MC, Moretta A and Moretta L. Analysis of the receptor-ligand interactions in the natural killer-mediated lysis of freshly isolated myeloid or lymphoblastic leukemias: evidence for the involvement of the Poliovirus receptor (CD155) and Nectin-2 (CD112). Blood 2005; 105: 2066-2073.

[45] Bevelacqua V, Bevelacqua Y, Candido S, Skarmoutsou E, Amoroso A, Guarneri C, Strazzanti A, Gangemi P, Mazzarino MC, D'Amico F, McCubrey JA, Libra M and Malaponte G. Nectin like-5 overexpression correlates with the malignant phenotype in cutaneous melanoma. Oncotarget 2012; 3: 882-892.

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

Characteristics	PLCO		HLCS		*P**
	Frequency	Deaths (%)	Frequency	Deaths (%)	
Total	1185	798 (67.3)	984	665 (67.5)	
Median overall survival (months)	23.8		39.9		
Age					
≤71	636	400 (62.9)	654	428 (65.4)	<0.0001
>71	549	398 (72.5)	330	237 (71.8)	
Sex					
Male	698	507 (72.6)	507	379 (74.7)	0.0006
Female	487	291 (59.8)	477	286 (59.9)	
Smoking status					
Never	115	63 (54.8)	92	52 (56.5)	0.166
Current	423	272 (64.3)	390	266 (68.2)	
Former	647	463 (71.6)	502	347 (69.1)	
Histology					
Adenocarcinoma	577	348 (60.3)	597	378 (63.3)	<0.0001
Squamous cell carcinoma	285	192 (67.4)	216	156 (72.2)	
Others	323	258 (79.9)	171	131 (76.6)	
Stage					
I-IIIA	655	315 (48.1)	606	352 (58.0)	0.003
IIIB-IV	528	482 (91.3)	377	313 (83.0)	
Missing	2		--		

*Chi-square test for the comparison of the characteristics between the PLCO trial and Harvard study for each clinical variable. Abbreviations: PLCO: Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; HLCS: Harvard Lung Cancer Susceptibility.

Table S2. List of 60 selected genes in the immunoregulatory interaction related gene-set used in the discovery analysis

Dataset	Name of pathway	Selected genes[a]	Number of genes
REACTOME	REACTOME_IMMUNOREGULATORY_INTERACTIONS_BETWEEN_A_LYMPHOID_AND_A_NON_LYMPHOID_CELL	*AMICA1, B2M, C3, CD160, CD19, CD200R1, CD226, CD247, CD34, CD3D, CD3E, CD3G, CD40, CD40LG, CD81, CD8A, CD8B, CD96, CDH1, CRTAM, CXADR, FCGR2B, FCGR3A, HCST, HLA-A, HLA-B, HLA-C, HLA-F, HLA-G, HLA-K, ICAM1, ICAM2, ICAM3, ICAM4, IFITM1, ITGAL, ITGB1, ITGB2, ITGB7, KIR2DL1, KIR2DL2, KIR2DL3, KIR2DL4, KIR2DS1, KIR2DS2, KIR3DL1, KIR3DL2, KLRC1, KLRD1, KLRK1, LILRA1, LILRB1, LILRB2, LILRB3, LILRB4, LILRB5, LOC651536, LOC652578, LOC652626, LOC653879, LOC731066, PVR, PVRL2, RAET1E, SELL, TYROBP, ULBP1, ULBP2, ULBP3, VCAM1*	70
Total	After removing 5 pseudogenes, 4 genomic position unavailable, and 1 gene on the x chromosome		60[b]

[a]Genes were selected based on online datasets and literatures, including GO, KEGG, REACTOME; [b]Duplicated genes, pseudogene and genes in X chromosome had been removed; Keyword: immunoregulatory AND interactions Organism: Homo sapiens.

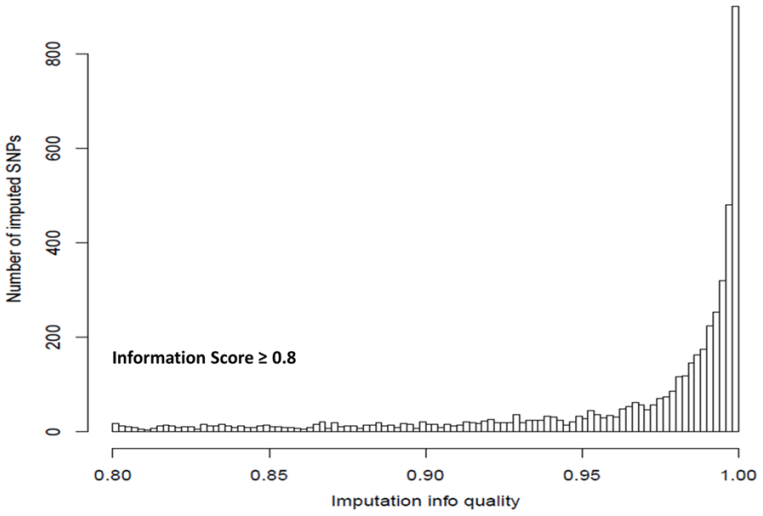


Figure S1. The distribution of the imputation information score of the present study.

Table S3. Associations of the first 10 principal components and overall survival of NSCLC in the PLCO trial

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. Abbreviations: NSCLC: non-small cell lung cancer; PLCO: the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; PC: principal component.

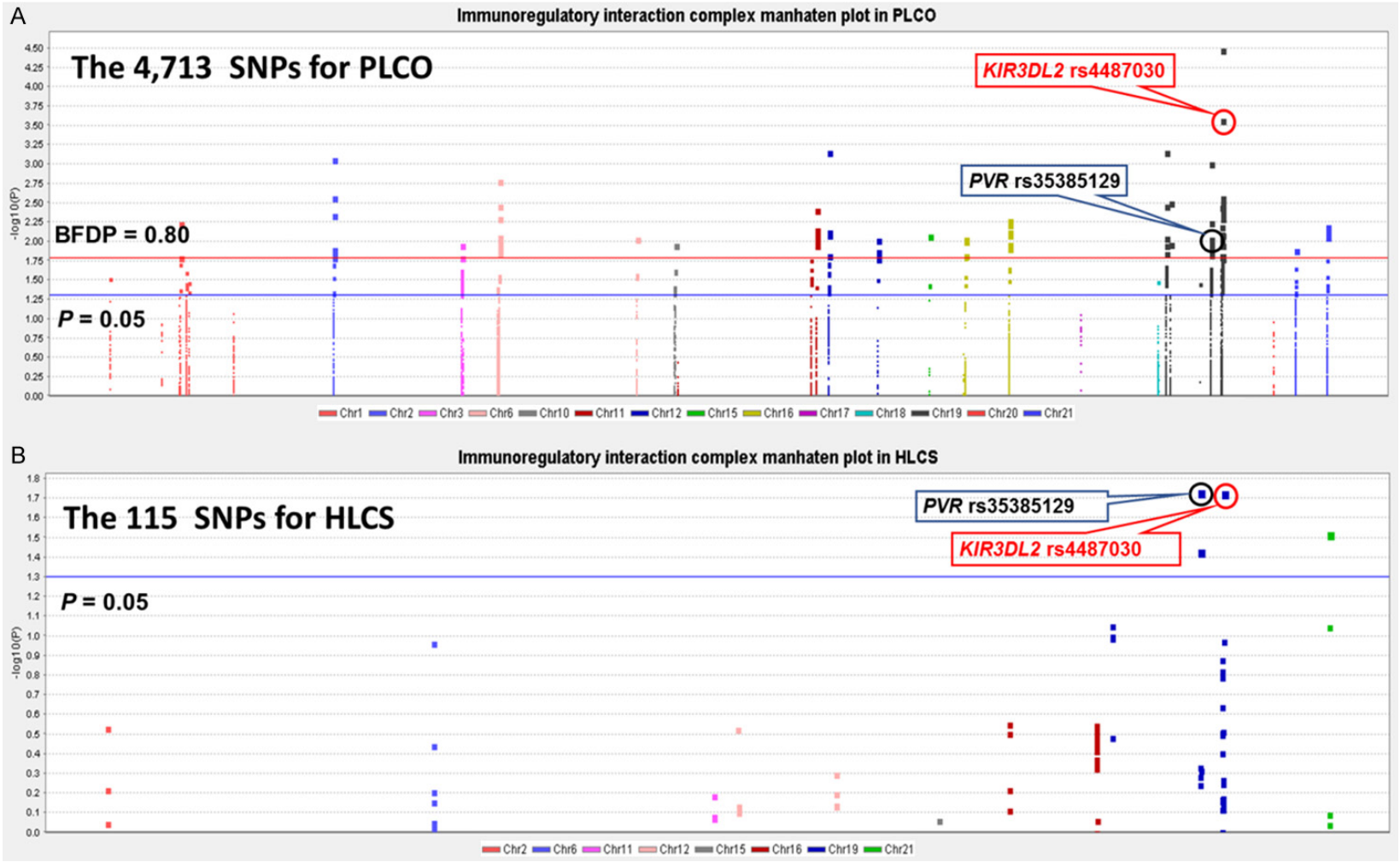


Figure S2. Manhattan plot of association between genotype data and overall survival of patients with NSCLC from the discovery and validation datasets in the PLCO trial and HLCS study, respectively. A. The statistical values across the autosomes of associations between 4,713 SNPs and overall survival in PLCO are plotted as -log10 *P* values. B. The statistical values across the autosomes of associations between 115 SNPs and overall survival in HLCS are plotted as -log10 *P* values. The blue horizontal line indicates P = 0.05 and the red line indicates BFDP = 0.80. Abbreviations: NSCLC, non-small cell lung cancer; SNPs, single nucleotide polymorphism; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening trial; OS, overall survival; BFDP, Bayesian false discovery probability.

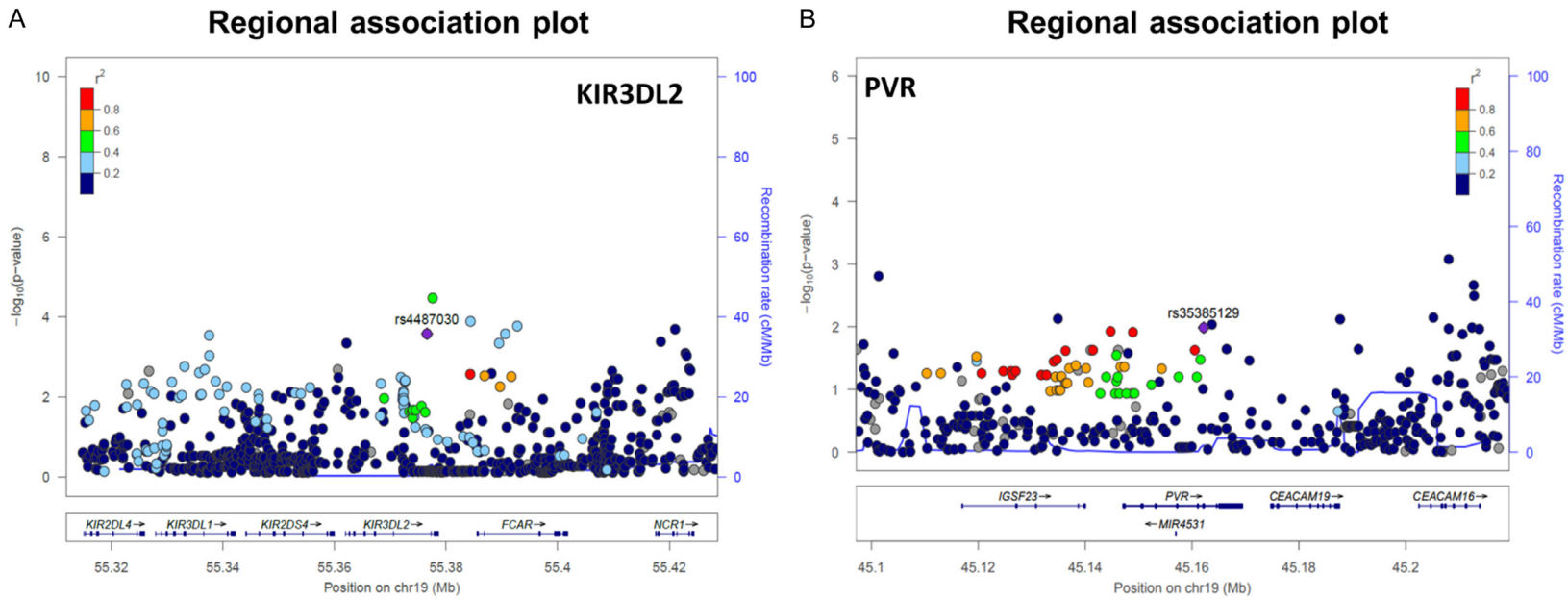


Figure S3. Regional association plots for the two independent SNPs in the immunoregulatory interaction pathway genes in the 1000 Genome Projects. Single nucleotide polymorphisms (SNPs) in the region of 50 kilobases up or downstream of (A) KIR3DL rs4487030 A>G and (B) PVR rs35385129 C>A. Data points are 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 S4. Stratified analysis for associations between the protective genotypes and survival of NSCLC in the PLCO trial[a]

Characteristics	0-1 protective genotype	2 protective genotypes	Multivariate Analysis[b] for OS			Multivariate Analysis[b] for DSS		
	Frequency	Frequency	HR (95% CI)	*P*	P_{inter}[b]	HR (95% CI)	*P*	P_{inter}[b]
Age (years)								
≤71	519	115	0.89 (0.68-1.16)	0.386		0.91 (0.69-1.20)	0.513	
>71	423	118	0.59 (0.45-0.78)	<0.001*	0.143	0.62 (0.46-0.82)	0.001*	0.304
Sex								
Male	547	148	0.71 (0.57-0.89)	0.003*		0.75 (0.59-0.96)	0.020*	
Female	395	85	0.84 (0.60-1.18)	0.320	0.785	0.83 (0.58-1.18)	0.291	0.970
Smoking status								
Never	101	13	0.88 (0.35-2.20)	0.777		0.94 (0.37-2.38)	0.897	
Current	335	82	0.98 (0.70-1.36)	0.886		1.12 (0.79-1.58)	0.530	
Former	506	138	0.62 (0.49-0.79)	<0.001*	0.096	0.62 (0.48-0.80)	<0.001*	0.045*
Histology								
Adeno	455	120	0.78 (0.59-1.03)	0.084		0.82 (0.61-1.09)	0.172	
Squamous	236	48	0.64 (0.42-0.97)	0.036*		0.65 (0.41-1.02)	0.062	
Others	251	65	0.73 (0.52-1.02)	0.062	0.500	0.75 (0.53-1.07)	0.111	0.408
Tumor stage								
I-IIIA	520	134	0.90 (0.67-1.21)	0.493		0.97 (0.70-1.34)	0.848	
IIIB-IV	422	99	0.68 (0.53-0.88)	0.003*	0.698	0.70 (0.54-0.91)	0.007	0.862
Chemotherapy								
No	509	129	0.61 (0.46-0.81)	<0.001*		0.63 (0.46-0.86)	0.003*	
Yes	433	104	0.84 (0.65-1.08)	0.173	0.047*	0.86 (0.67-1.12)	0.262	0.039*
Radiotherapy								
No	604	157	0.78 (0.61-1.00)	0.049*		0.83 (0.64-1.08)	0.160	
Yes	338	76	0.66 (0.49-0.88)	0.005*	0.852	0.66 (0.49-0.89)	0.007*	0.841
Surgery								
No	521	114	0.73 (0.58-0.92)	0.007*		0.75 (0.60-0.95)	0.016*	
Yes	421	119	0.73 (0.52-1.03)	0.070	0.554	0.76 (0.52-1.11)	0.153	0.585

[a]Adjusted for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3 and PC4; [b]P_{inter}: *P*-value for interaction analysis between characteristic and protective genotypes. *$P<0.05$, Abbreviations: HR: hazards ratio; CI: confidence interval.

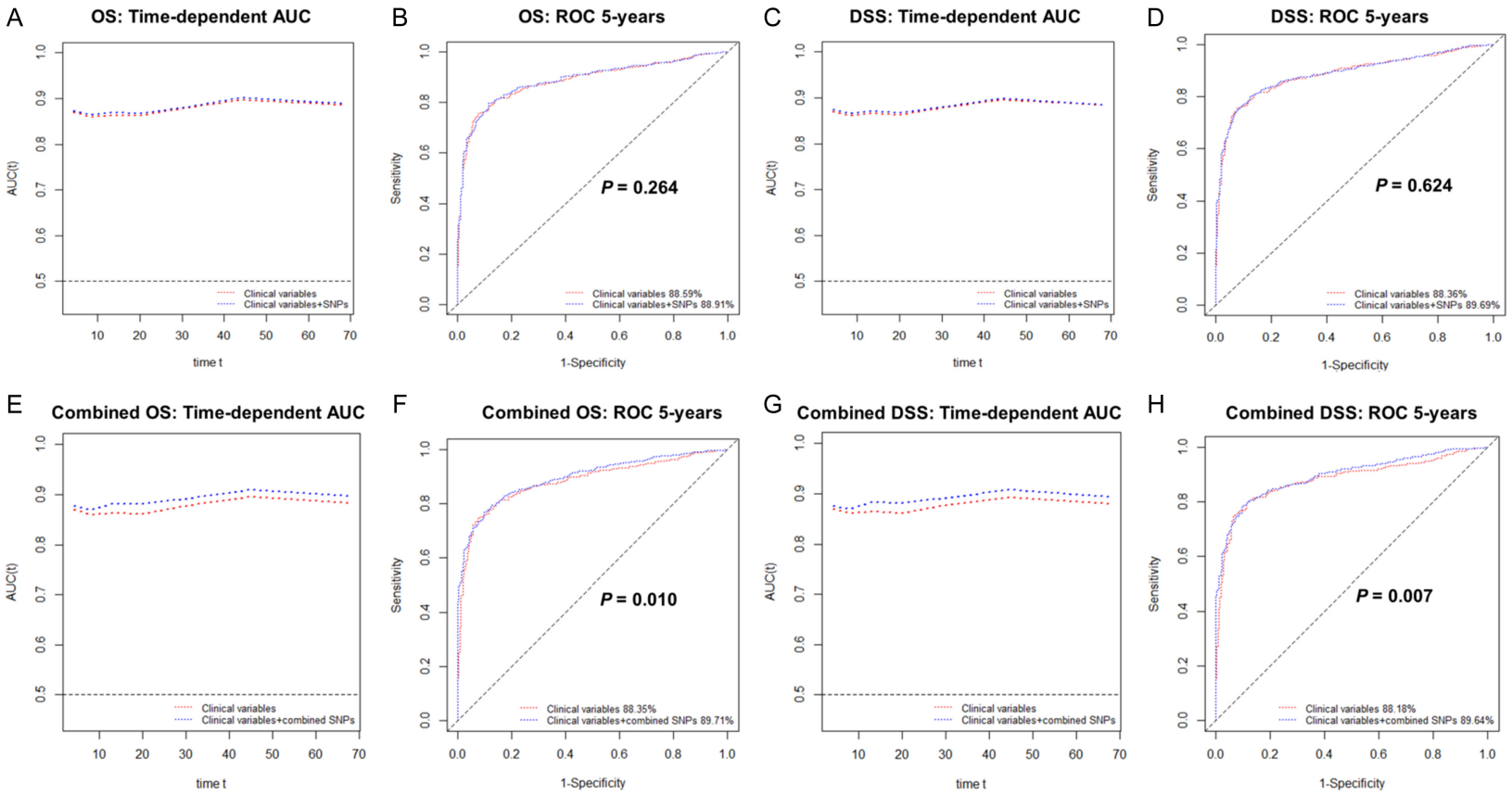


Figure S4. Five-year NSCLC survival prediction of the two SNPs and 15 previously published SNPs by ROC curve in PLCO dataset. Time-dependent AUC estimation to OS (A) and DSS (C) of the two SNPs: based on age, sex, smoking status, histology, tumor grade, tumor stage, chemotherapy, surgery, principal component and the risk genotypes of the two genes. 5-years NSCLC OS (B) and DSS (D) prediction by ROC curve of the two SNPs. Time-dependent AUC estimation to OS (E) and DSS (G) of the two genes combined 15 previously published SNPs. 5-years NSCLC OS (F) and DSS (H) prediction by ROC curve of combined 15 previously published SNPs. Abbreviations: ROC, receiver operating characteristic curve; AUC, area under curve; OS, overall survival; DSS, disease-specific survival; PLCO, The Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.

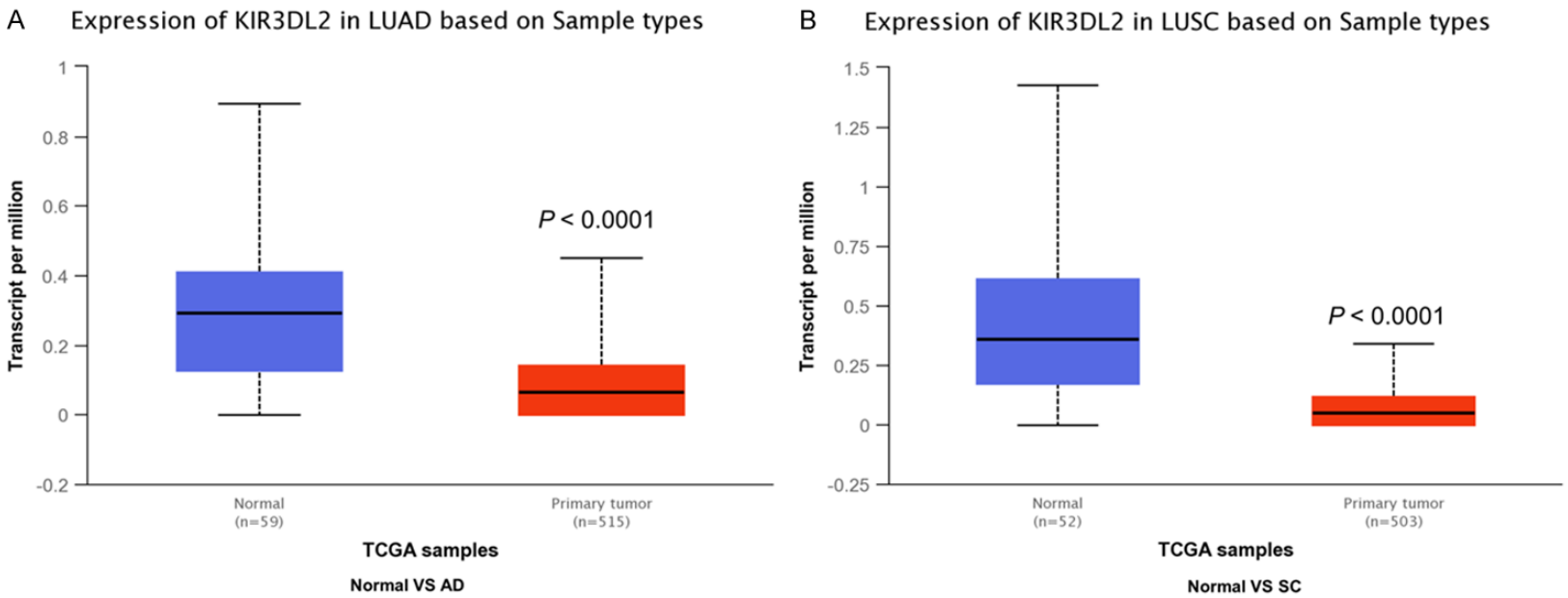
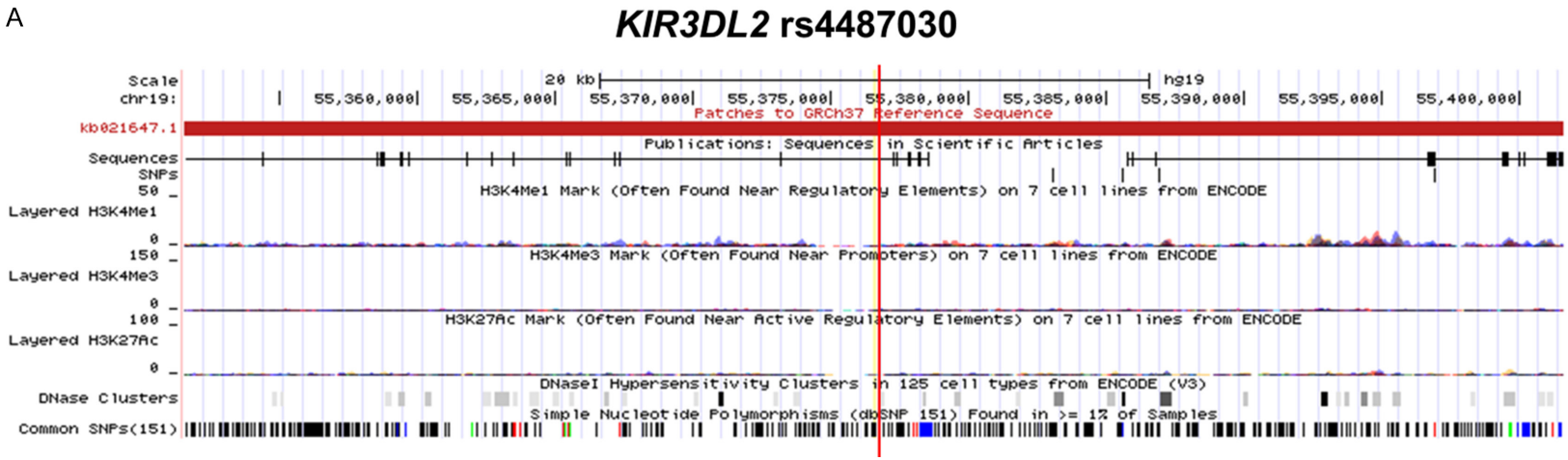


Figure S5. Differential mRNA expression analysis of the KIR3DL2 in the TCGA database. Higher mRNA of KIR3DL2 expression levels (A) were found in the non-paired normal tissues than tumor tissues in the LUSC and in the LUAD (B).

A



B

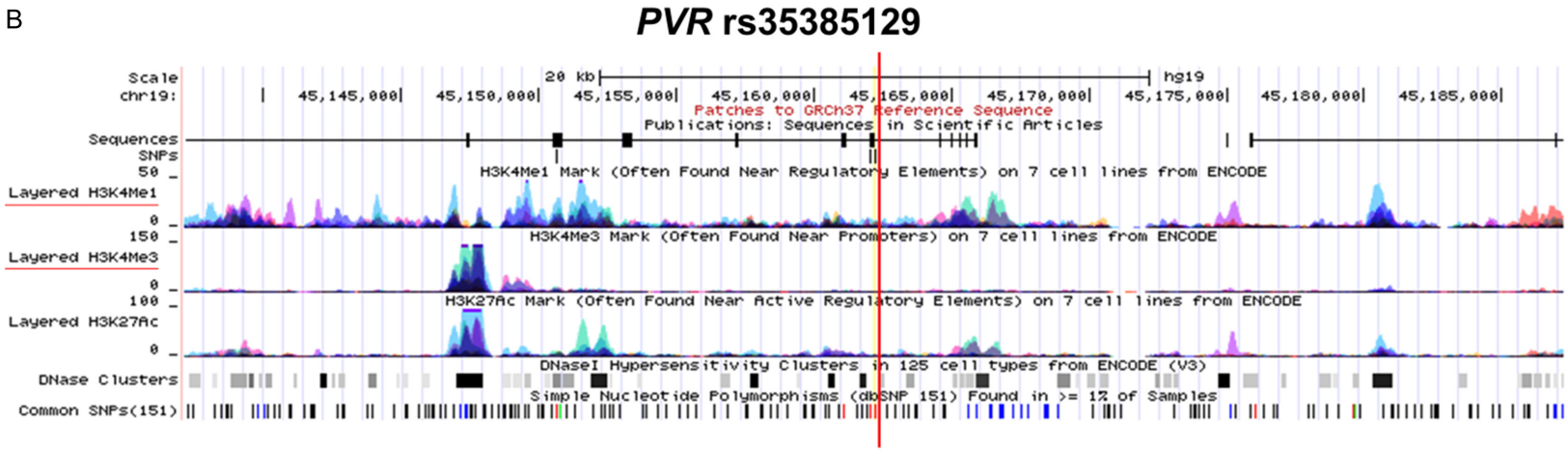
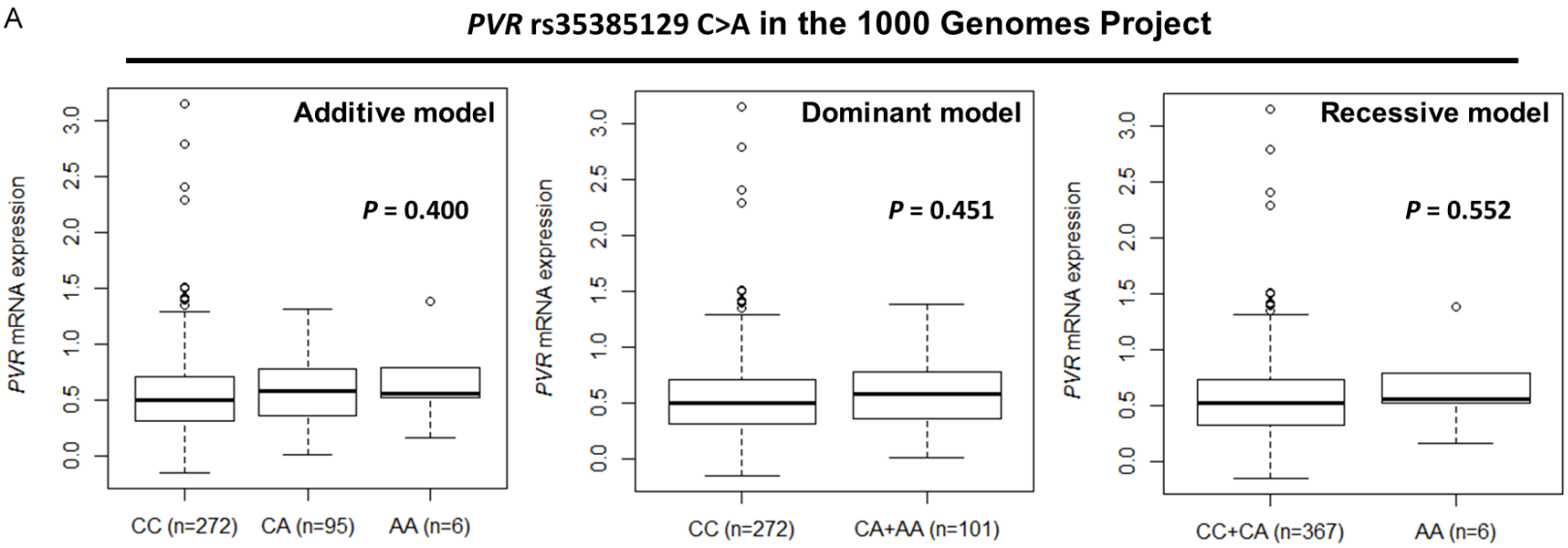
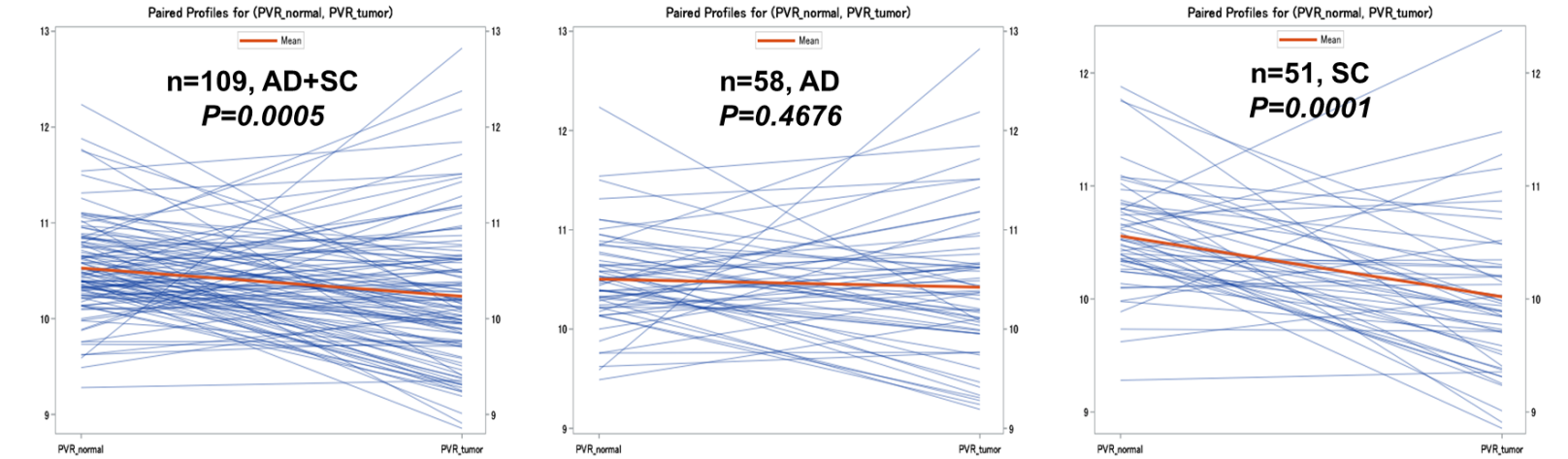


Figure S6. Expanded view of the ENCODE data for the rs4487030 and rs35385129. The H3K27Ac, H3K4Me1, and H3K4Me3 tracks showed the genome-wide levels of enrichment of acetylation of lysine 27, the mono-methylation of lysine 4, and tri-methylation of lysine 4 of the H3 histone protein, as determined by the ChIP-seq assays. These levels are thought to be associated with the promoter and enhancer regions. DNase clusters track showed Dnase hypersensitivity areas. Tnx factor track showed regions of transcription factor binding of DNA, as assayed by ChIP-seq experiments. Transcription showed target genes of transcription factors from transcription factor binding site profiles. Abbreviation: ChIP, Chromatin Immunoprecipitation Sequencing.

A


PVR rs35385129 C>A in the 1000 Genomes Project
Additive model
P = 0.400
PVR mRNA expression
CC (n=272)
CA (n=95)
AA (n=6)
Dominant model
P = 0.451
CC (n=272)
CA+AA (n=101)
Recessive model
P = 0.552
CC+CA (n=367)
AA (n=6)

B


Paired Test – PVR mRNA expression analysis in the TCGA dataset
Paired Profiles for (PVR_normal, PVR_tumor)
Mean
n=109, AD+SC
P=0.0005
n=58, AD
P=0.4676
n=51, SC
P=0.0001
PVR_normal
PVR_tumor

C

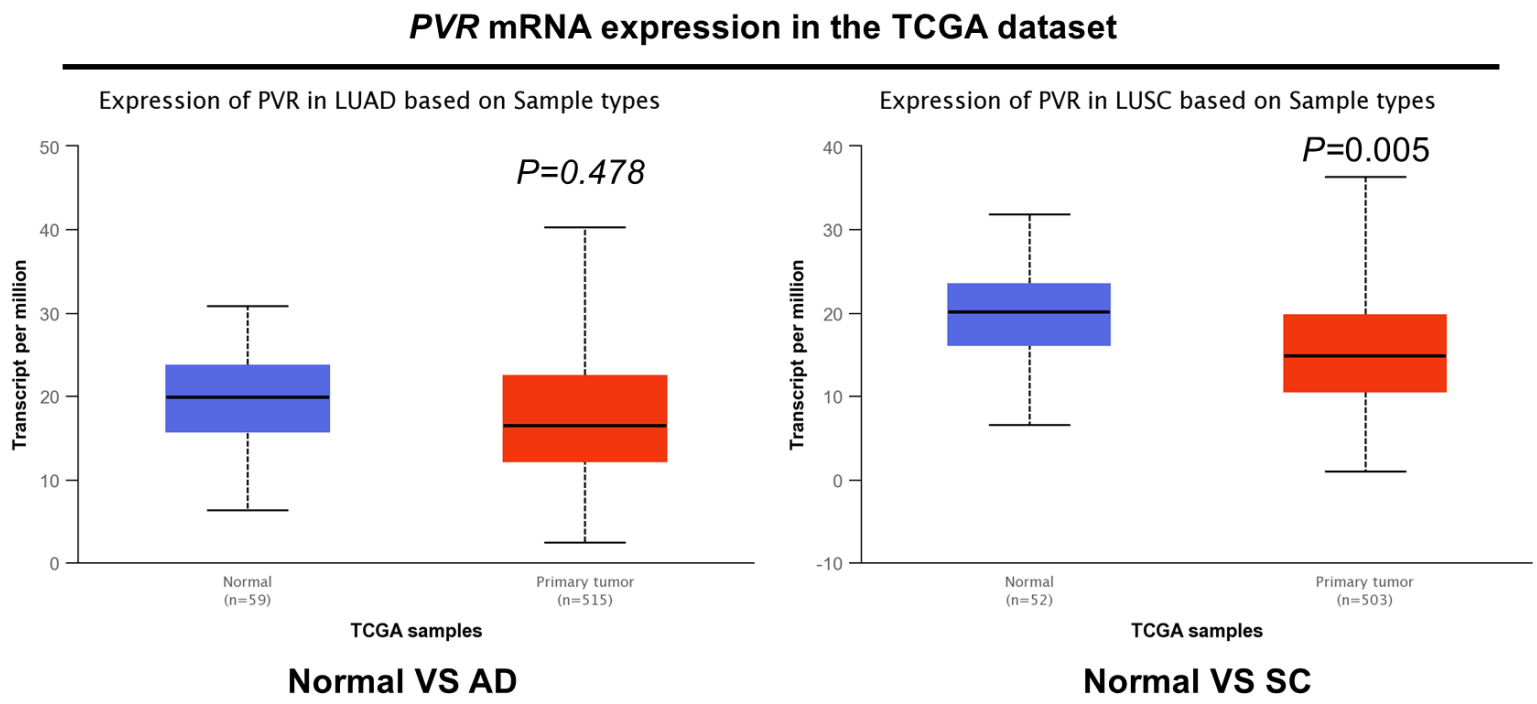


Figure S7. PVR rs35385129 genotypes corresponding with mRNA expression levels which is in differential expression in the TCGA database. The eQTL for PVR rs35385129 (A) in 373 Europeans from the 1000 Genomes Project; (B) Higher PVR mRNA expression levels were found in the adjacent normal tissues of 111 NSCLC and in the 51 paired LUSC tissues but no in LUAD; Higher PVR mRNA expression levels were found in the non-paired normal tissues (C) than tumor tissues in LUSC but not in the LUAD. Abbreviation: eQTL, expression quantitative trait loci analysis; TCGA, The Cancer Genome Atlas; NSCLC, non-small cell lung cancer; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

Table S5. Function prediction for rs4487030, rs35385129 and their in-high-LD SNPs

SNP	Gene	Chr	Allele	RegDB[1]	Haploreg v4.1[2]					
					Promoter histone marks	Enhancer histone marks	DNAse	Motifs changed	Selected eQTL hits	dbSNP func annot
rs4487030	*KIR3DL2*	19	A>G	No Data	--	--	--	CDP, p300	--	intronic
rs35385129	*PVR*	19	C>A	5	--	--	SKIN	Hoxa5, Msx-1	5 hits	missense
rs846886	6.6 kb 5' of *IGSF23*	19	A>G	3a	--	6 tissues	7 tissues		3 hits	
rs846893	*IGSF23*	19	A>G	No Data	--	4 tissues		4 altered motifs	4 hits	intronic
rs7247676	*IGSF23*	19	A>G	4	--	6 tissues	PANC, CRVX, SKIN	ATF3, CEBPD, SREBP	3 hits	intronic
rs846853	*IGSF23*	19	A>G	5	--	ESDR, GI, LIV			4 hits	intronic
rs846854	*IGSF23*	19	T>C	6	BLD	8 tissues		STAT, Smad3	4 hits	intronic
rs846855	*IGSF23*	19	G>A	4	BLD, BONE	11 tissues	5 tissues	AP-1, VDR	3 hits	intronic
rs846857	*IGSF23*	19	A>G	2a	BLD	14 tissues	MUS, MUS	8 altered motifs	4 hits	intronic
rs846861	*IGSF23*	19	T>C	No Data	--			NRSF, SETDB1, Zfx	4 hits	intronic
rs846863	*IGSF23*	19	T>G	No Data	--			Pax-6	4 hits	intronic
rs846866	*IGSF23*	19	A>C	5	--	4 tissues	ADRL, OVRY	Hic1, Znf143	4 hits	intronic
rs12978272	*IGSF23*	19	C>T	No Data	--	ADRL		Foxm1, SIX5, Znf143	4 hits	intronic
rs58967105	*IGSF23*	19	C>A	5	--	ESC, IPSC, LIV		Hand1, Smad	4 hits	intronic
rs66786749	940 bp 3' of *IGSF23*	19	AG>A	No Data	--	5 tissues		Smad4, TATA, ZBRK1	4 hits	
rs34537198	1.4 kb 3' of *IGSF23*	19	G>A	No Data	--	5 tissues		Pax-5	4 hits	
rs3760624	2.3 kb 5' of *PVR*	19	G>C	No Data	--	6 tissues		TATA	6 hits	
rs201547681	916 bp 5' of *PVR*	19	AATT>A	6	--	GI, LNG, LIV	GI	7 altered motifs	5 hits	
rs28411142	*PVR*	19	C>T	5	13 tissues	10 tissues	12 tissues	NRSF, Zec	5 hits	intronic
rs35739046	*PVR*	19	G>A	2b	--	12 tissues	27 tissues	7 altered motifs	5 hits	intronic

Abbreviations: SNP: single nucleotide polymorphism; LD: linkage disequilibrium; NSCLC: non-small cell lung cancer; Chr: chromosome; dbSNP func annot: dbSNP function annotation; [1]RegulomeDB: http://regulomedb.org/; [2]Haploreg: https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php.

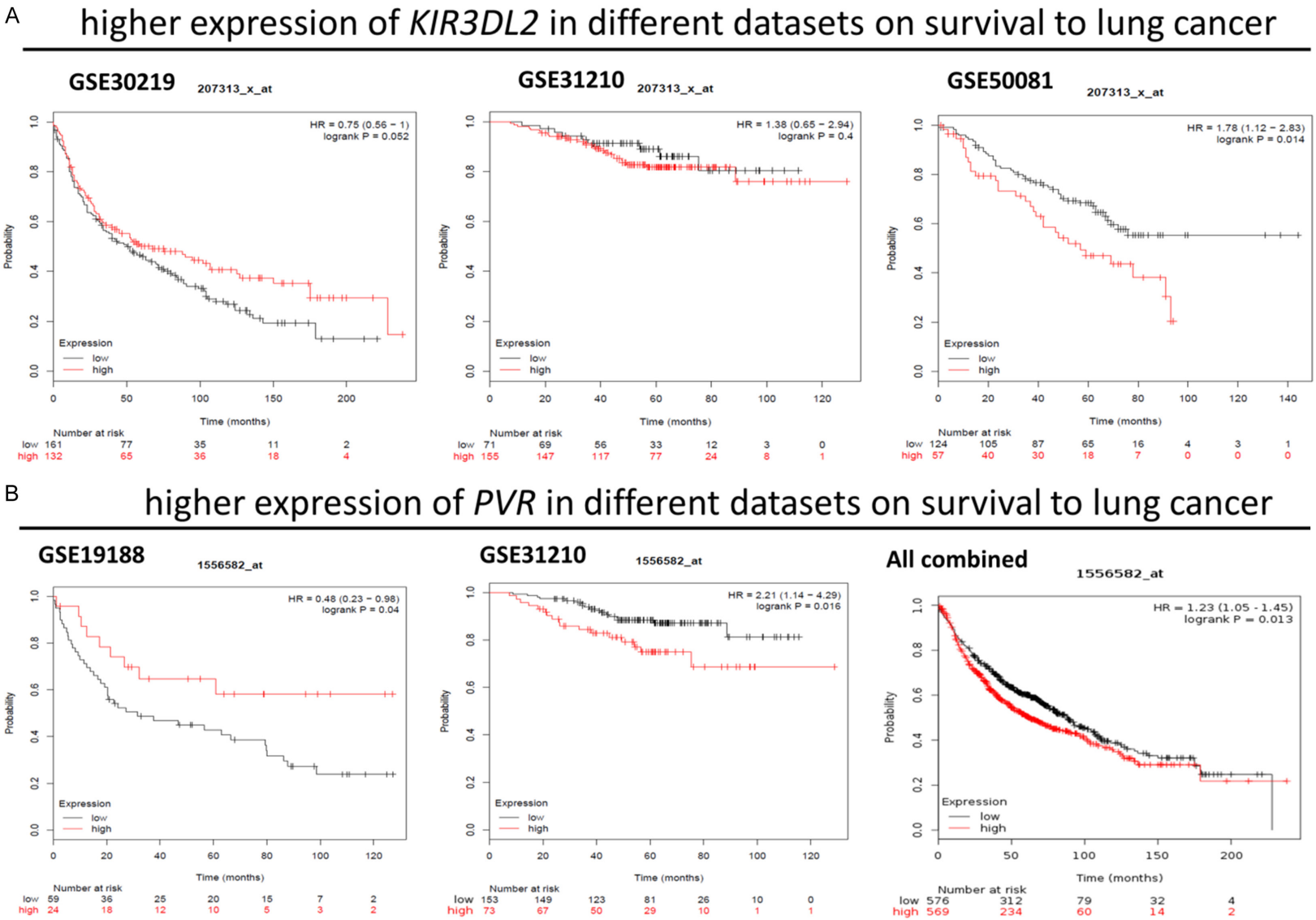


Figure S8. Effect of higher expression of KIR3DL2 or PVR in different datasets on survival to lung cancer.