HHS Public Access

Author manuscript

Int J Cancer. Author manuscript; available in PMC 2021 May 04.

Published in final edited form as:

Int J Cancer. 2020 July 15; 147(2): 392–403. doi:10.1002/ijc.32739.

Novel genetic variants in *KIF16B* and *NEDD4L* in the endosome-related genes are associated with non-small cell lung cancer survival

Sen Yang[#1,2,3], **Dongfang Tang**[#2,3], **Yu Chen Zhao**[2,3], **Hongliang Liu**[2,3], **Sheng Luo**[4], **Thomas E. Stinchcombe**[2,5], **Carolyn Glass**[2,6], **Li Su**[7], **Sipeng Shen**[7], **David C. Christiani**[7,8], **Qiming Wang**[1,**], **Qingyi Wei**[2,3,5,**]

[1]Department of Internal Medicine, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China.

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

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

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

[5]Department of Medicine, Duke University Medical Center, Durham, NC 27710, USA

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

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

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

[#] These authors contributed equally to this work.

Abstract

The endosome is a membrane-bound organ inside most eukaryotic cells, playing an important role in adaptive immunity by delivering endocytosed antigens to both MHC class I and II pathways. Here, by analyzing two published genome-wide association studies (GWASs), we evaluated associations between genetic variants in the endosome-related gene-set and survival of patients with non-small cell lung cancer (NSCLC). The discovery included 44,112 (3,478 genotyped and 40,634 imputed) single-nucleotide polymorphisms (SNPs) in 220 genes in a single locus analysis for their associations with survival of 1,185 NSCLC patients from the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. After validation of the 821 survival-associated significant SNPs in additional 984 NSCLC patients from the Harvard Lung Cancer Susceptibility

[**]Correspondence to: Qiming Wang, Department of Internal Medicine, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China; qimingwang1006@126.com; or Qingyi Wei, Duke Cancer Institute, Duke University Medical Center and Department of Population Health Sciences, Duke University School of Medicine, 905 S LaSalle Street, Durham, NC 27710, USA, Tel.: (919) 660-0562, qingyi.wei@duke.edu.

Conflict of interest

The authors state no conflict of interest.

study, 14 SNPs remained significant. The final multivariate stepwise Cox proportional hazards regression model in the PLCO datasets identified three potentially functional and independent SNPs (*KIF16B* rs1555195 C>T, *NEDD4L* rs11660748 A>G and rs73440898 A>G) with an adjusted hazards ratio (HR) of 0.86 [95% confidence interval (CI)=0.79-0.94, P=0.0007], 1.31 (1.16-1.47, P=6.0×10^{-5}) and 1.27 (1.12-1.44, P=0.0001) for overall survival (OS), respectively. Combined analysis of the adverse genotypes of these three SNPs revealed a trend in the genotype-survival association (P_{trend}<0.0001 for OS and P_{trend}<0.0001 for disease-specific survival). Furthermore, the survival-associated *KIF16B* rs1555195T allele was significantly associated with decreased mRNA expression levels of *KIF16B* in both lung tissues and blood cells. Therefore, genetic variants of the endosome-related genes may be biomarker for NSCLC survival, possibly through modulating the expression of corresponding genes.

Keywords

Non-small cell lung cancer; endosome pathway; genome-wide association study; single-nucleotide polymorphism; survival

Introduction

Lung cancer is one of the most common malignancies, with the highest cancer-related mortality worldwide [1]. In the United States, it is estimated that there will be approximately 228,150 new cases and 142,670 deaths from lung cancer in 2019 [2]. The most common histological type of lung cancer is non-small cell lung cancer (NSCLC), accounting for approximately 85% of all lung cancer patients [3]. Although targeted therapy and immunotherapy have made remarkably improved outcomes in patients with NSCLC and facilitated the development of personalized cancer treatment [4], the prognosis of NSCLC patients remain heterogeneous, suggesting that genetic factors may play an important role in treatment response and efficacy. Moreover, genetic factors, such as single nucleotide polymorphisms (SNPs), have been shown to have an effect on prognosis of lung cancer patients [5-7]. Therefore, identifying the roles of these genetic factors in development and progression of lung cancer may lead to better personalized management and treatment of NSCLC patients.

To date, few novel and functional SNPs have been identified to be associated with prognosis of lung cancer patients in genome-wide association studies (GWASs). This is because a hypothesis-free GWAS has always focused on the top or most-significant SNPs/genes with a stringent P value after correction of multiple tests for numerous SNPs, and most of the identified top SNPs lack of functional annotations. Recently, the biological pathway-based approach, as a hypothesis-driven method in the post GWAS era, has been applied to the reanalysis of published GWAS datasets to test the cumulative effect of potentially functional SNPs across multiple genes in the same biological pathway. As a result, much fewer SNPs in candidate genes of a significant biological pathway were included in the analysis to avoid the nuisance of multiple tests, which improves the study power to detect statistically significant and biologically important associations for additional functional analysis.

The endosome is a membrane-bound compartment inside eukaryotic cells and plays an important role in the endocytosis of exogenous antigens [8]. Classical antigen-presentation studies have showed that major histocompatibility complex (MHC) class I molecules present peptides derived from proteins synthesized within the cell, whereas MHC class II molecules present exogenous proteins from outside of the cell and the microenvironment. Emerging evidence indicates that dendritic cells have a specialized capacity of processing exogenous antigens into the MHC class I pathway. This function, known as a cross-presentation, helps dendritic cells to activate the anti-tumor activity of cytotoxic T lymphocyte (CTL) [9]; thus, the endocytosed antigens from the outside are delivered to both MHC class I and MHC class II pathways through the functioning endosome.

In recent years, the role of the immune system in cancer development and progression has been recognized widely [10-12]. Immunotherapy is now established as the “fourth pillar” of cancer treatment alongside surgery, radiation, and chemotherapy [13]. Immunotherapy alone in patients with a high level of PD-L1 expression or in combination with chemotherapy is now the standard first-line therapy for patients with metastatic NSCLC [14-16]. For patients with stage-III NSCLC treated with chemotherapy and radiation, additional immunotherapy is the current standard of care. However, many patients do not benefit from immunotherapy, and there is an urgent need to identify tumor- and patient-related predictive biomarkers of immunotherapy. Such observations may be due to the killing effect of the immune system in tumor cells being highly dependent on the activation of CTL and CD4+ helper T cells (Th cells). CTL and Th cells are activated by the complex of internalized tumor antigens bonded to MHC class I and MHC class II protein molecules located on the surface of cancer cells and dendritic cells, respectively [17]. Therefore, we hypothesize that genetic variants of the genes involved in the endosome-related pathway in the process of anti-tumor immune response are associated with NSCLC survival. We tested this hypothesis by using genotyping data of two independently published NSCLC GWAS datasets.

Materials and Methods

Study populations

We used a GWAS dataset from the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial as the discovery, which is a randomized control study conducted by the National Cancer Institute (NCI) [16]. The PLCO trial included 77,500 men and 77,500 women aged 55-74 years, who were enrolled between the year of 1993 and 2011 from 10 medical centers in the United States; the participants were randomized to either the intervention arm that received a trial screening or the control arm that received standard care instead [18]. All participants provided their blood samples and personal information including smoking status, histologic diagnosis, tumor stage, treatment method and family history at enrollment and were followed up for at least 13 years after the enrollment [19]. After excluding two individuals who had no follow-up information, a total of 1,185 NSCLC patients were eligible for survival analysis. 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) [20,21]. These 1,185 NSCLC patients with both complete follow-up information and genotype data were

used for survival analysis. The institutional review boards of each participating institution had approved the PLCO trial and use of its data, and all the participants had provided a written informed consent permitting the research represented here.

We used another GWAS dataset of 984 histology-confirmed Caucasian NSCLC patients from the Harvard Lung Cancer Susceptibility (HLCS) Study which began in 1992 as the validation. In the HLCS study, the whole blood samples and personal information were collected after diagnosis, and DNA from the blood samples were 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 was for imputation with the Mach3 software based on the sequencing data from the 1,000 Genomes Project[22].

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=1185) and the HLCS study (n=984) is presented in Supplementary Table 1.

Gene and SNP selection

The genes involved in the endosome-related pathway were selected by the Molecular Signatures Database (http://software.broadinstitute.org/gsea/msigdb/index.jsp) with the keyword “endosome”. After the removal of 44 duplicated genes and six genes in the X chromosome, 220 genes remained as candidate genes for further analysis (Supplementary Table 2). These genes were used for imputation with IMPUTE2 and the 1,000 Genomes Project data (phase 3), in which SNPs within their ±2kb flanking regions (SNPs located in the 2-kb upstream and downstream of a gene were considered having potential effects on gene transcription) were extracted with the following criteria: imputation info score ≥ 0.8 (Supplementary Figure 1), genotyping rate ≥ 95%, minor allelic frequency (MAF) ≥ 5%, and Hardy-Weinberg equilibrium (HWE) ≥ 1×10^{-5}. As a result, 3,478 genotyped SNPs were selected from the PLCO GWAS dataset (dbGaP accession: phs000093.v2.p2 and phs000336.v1.p1) and 40,634 SNPs were imputed.

Statistical analyses

We used multivariate Cox proportional hazards regression analysis to assess associations between each SNP and NSCLC survival (in an additive genetic model) in the PLCO dataset, with adjustment for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery and the first four principal components by using the GenABEL package of R software [23]. We used the recommended Bayesian false discovery probability (BFDP) method with a cut-off value of 0.80 for multiple testing correction to lower the probability of potentially false positive results [24]. We assigned a prior probability of 0.05 to detect an HR of 3.0 for an association with variant genotypes or minor alleles of the SNPs with $P<0.05$. After that, we validated these chosen SNPs by using the HLCS GWAS dataset. Next, we performed an inverse variance weighted meta-analysis to combine the results of both discovery and validation datasets. In the 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 SNP. After that these potential independent SNPs was adjusted for previously published SNPs.

Then, we used the combined genotypes to evaluate the cumulative effects of the identified SNPs and the Kaplan-Meier curve to estimate the 10-year survival probability associated with the genotypes. We also assessed possible interactions with a Chi-square-based Q-test between subgroups in the stratified analysis, and P<0.05 was considered statistically significant. We then performed the receiver operating characteristic (ROC) curve and time-dependent area under the curve (AUC) with timeROC package of R software (version 3.5.0) to illustrate the prediction accuracy of the model integrating 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 linear regression using the R software. The mRNA expression data of genes were obtained from two sources: 373 European individuals included in the 1,000 Genomes Project and 369 whole blood samples and 383 normal lung tissue included in the genotype-tissue expression (GTEx) project [26,27]. Then, bioinformatics functional prediction for the identified SNPs were performed with SNPinfo [28], RegulomeDB [29] (http://www.regulomedb.org) and HaploReg [30] (http://archive.broadinstitute.org/mammals/haploreg/haploreg.php). Finally, the differences in mRNA expression levels were examined in 109 pairs of lung cancer tissues and adjacent normal tissues from the Cancer Genome Atlas (TCGA) dataset by using a paired *t* test model. Kaplan-Meier 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) [31]. All statistical analyses were performed with the SAS software (version 9.4; SAS Institute, Cary, NC, USA) unless otherwise indicated.

Data availability

The datasets used for the analyses described in the present study were obtained from dbGaP (http://www.ncbi.nlm.nih.gov/gap) through dbGaP accession number phs000336.v1.p1 and phs000093.v2.p2.

Results

Associations between SNPs in the endosome-related pathway genes and NSCLC survival

The workflow chart of the present study is shown in Figure 1. The basic characteristics of 1,185 NSCLC patients from the PLCO trial and 984 NSCLC patients from the HLCS study have been described elsewhere [32]. In the discovery PLCO genotype dataset, a single-locus multivariate Cox regression analysis was performed for the selected 44,112 SNPs with adjustment for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery and the first four principal components (Supplementary Table 3). For multiple testing correction, none of the SNPs passed Bonferroni Correction (P>0.05) or false

discovery rate (>0.20). This is likely due to the high LD among the SNPs generated by imputation. Besides, our purpose of using this pre-screening was to identify "candidates for functional SNPs for further analysis. Therefore, we used the BFDP method as recommended by the authors of the method[24]. After multiple testing correction by BFDP ≤ 0.80, 821 SNPs were identified to be significantly associated with NSCLC OS (P<0.05). All the significant SNPs identified from the PLCO trial were further validated by the HLCS genotype dataset, and 14 SNPs remained significant. Subsequently, we performed meta-analysis of these 14 newly identified SNPs in both PLCO and HLCS datasets and found that a better survival was associated with the *KIF16B* rs1555195 C>T (P=0.0007), but a poor survival was associated with the other 13 SNPs, without heterogeneity between the two studies (Table 1).

Independent SNPs associated with NSCLC survival in the PLCO dataset

To identify independent of the other 13 SNPs, we performed a multivariate stepwise Cox regression analysis with adjustment for demographic and clinical variables and the first four principal components in the PLCO dataset, and we used the Schwarz Bayesian Criterion (SBC)[33] for model selection to identify independent SNPs associated with NSCLC survival.

When all the 14 validated SNPs were added to the model, only three SNPs were left and significantly associated with survival. After that, in the same model, we also adjusted for other 15 previously reported significant SNPs, and these three SNPs remained significantly associated with survival (Table 2). The results of selected SNPs are summarized in a Manhattan plot (Supplementary Figure 2) and the regional association plot of each of these three SNPs is shown in Supplementary Figure 3.

In the PLCO dataset with available covariates for complete adjustment, patients with the rs1555195T allele had a decreased risk of death [P_{trend}=0.003 for OS and P_{trend}=0.003 for disease-specific survival (DSS)], while patients with the rs11660748G allele and rs73440898G allele had an increased risk of death (P_{trend}<0.0001 for OS and P_{trend}=0.0003 for DSS; P_{trend}=0.001 for OS and P_{trend}=0.015 for DSS; respectively) (Table 3). Compared with the reference genotype in a dominant genetic model, *KIF16B* rs1555195 CT+TT genotypes were associated with a better survival (HR=0.81, 95% CI=0.71-0.94, P=0.005 for OS and HR=0.80, 95% CI=0.69-0.94, P=0.005 for DSS), while *NEDD4L* rs11660748 AG +GG genotypes and rs73440898 AG+GG had a worse survival (HR=1.37, 95% CI=1.16-1.63 and P=0.0003 for OS and HR=1.37, 95% CI=1.14-1.65 and P=0.0006 for DSS; and HR=1.32, 95% CI=1.11-1.58, P=0.002 for OS and HR=1.25, 95% CI=1.03-1.51, P=0.022 for DSS; respectively) (Table 3).

Haplotype analysis of two SNPs in *NEDD4L* and NSCLC survival in PLCO

Since rs73440898 and rs11660748 were both in *NEDD4L*, we performed haplotype analysis assess the relation between different haplotypes and survival. As shown in Table 4, there three SNPs is shown in Supplementary Figure 3. were four *NEDD4L* haplotypes (A-A, A-G, G-A and G-G) of the rs73440898 and rs11660748 loci, with a frequency of 82.40%, 8.18%, 7.16%, and 2.26%, respectively, and a significant NSCLC death-risk was associated with the G haplotypes (HR=1.32, 1.27 and 1.46 for OS, respectively; and HR=1.32. 1.20 and 1.43 for DSS, respectively, compared with the A-A haplotype) in a G-allele dose-dependent manner

(P_{trend} < 0.0001 for OS and P_{trend} = 0.001 for DSS). In the dichotomized analysis, patients who had 1-2 death-risk alleles had an unfavorable survival, compared with those with the A-A haplotype (HR = 1.32, 95% CI = 1.16-1.50; P<0.0001 for OS and HR = 1.28, 95% CI = 1.12-1.47; P=0.0004 for DSS). These results are consistent with the observed death-risk associated with the *NEDD4L* rs11660748G and rs73440898G alleles.

Combined effects of the three independent SNPs in the PLCO dataset

We used the PLCO dataset to assess the combined effect of the three independent SNPs on NSCLC OS and DSS. First, we combined the unfavorable genotypes (i.e., *KIF16B* rs1555195 CC, *NEDD4L* rs73440898 AG+GG, *NEDD4L* rs11660748 AG+GG) into a genetic score as the number of unfavorable genotypes (NUGs). As shown in Table 3, the increased genetic score of the NUGs was associated with a worse effect on death in the multivariate analysis in the PLCO dataset (P_{trend}<0.0001 for OS and P_{trend}<0.0001 for DSS). Then, we dichotomized all the patients into a low-unfavorable group (0-1 scores) and a high-unfavorable group (2-3 scores). Compared with the low-unfavorable score group, patients in the high-unfavorable score group had a significantly worse survival (HR=1.58, 95% CI=1.33-1.87, P<0.0001 for OS and HR=1.48, 95% CI=1.23-1.78, P<0.0001 for DSS). Kaplan-Meier survival curves were presented to depict the associations between unfavorable genotypes and NSCLC OS and DSS (Figure 2a, 2b, 2c and 2d).

Stratified analysis for associations between NUGs and NSCLC survival

We performed stratified analysis to evaluate whether the combined effect of unfavorable genotypes on NSCLC OS and DSS was modified by age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy and surgery in the PLCO dataset. As a result, no significant interactions were found when it was performed on both NSCLC OS AND DSS (P_{inter}>0.05 for all factors, Supplementary Table 4).

The ROC curves and time dependent AUC

We assessed the predictive value of the three SNPs with time-dependent AUC and ROC curves at the 60th month (or the fifth year) and 12th (or the first year) month in the PLCO dataset. Compared with the covariates model, the time-dependent AUC plot with the independent unfavorable genotypes did not improve prediction performance of the model at the 60th month (Supplementary Figure 4). However, when we performed the time-dependent AUC and ROC curves at the 12th month in the PLCO dataset. The prediction performance of the model was improved significantly. The AUCs changed from 85.84% to 86.61% (P = 0.006) for OS and from 86.16% to 86.78% (P = 0.025) for DSS (Figure 2e, 2f).

The eQTL analyses

We performed the eQTL analysis to explore the correlations between genotypes of the three independent SNPs and their corresponding mRNA expression levels by using the RNA-Seq data of lymphoblastoid cell lines from 373 European descendants available in the 1,000 Genomes Project and the data of 369 whole blood samples and 383 normal lung tissue from the GTEx project. In the 1,000 Genomes Project, all three SNPs showed no significant correlation with their corresponding mRNA expression levels (Supplementary Figure 5) [26].

Then, we performed eQTL by using the expression data of the lung and whole blood from the GTEx project. We found that the *KIF16B* rs1555195T allele was associated with lower expression levels of *KIF16B* in both lung normal tissues and whole blood cells (P=0.0009 and P=0.005, respectively; Figure 2g and 2h). For the *NEDD4L* rs11660748G and rs73440898G alleles, they were not significantly correlated with their corresponding mRNA expression levels (Supplementary Figure 6) [27]. At last, we performed functional prediction for these three with the online tools of SNPinfo [28], RegulomeDB [29], and Haploreg [30] to predict their bioinformatics function. As a result, all the three SNPs had no function based on the SNPinfo, but have some bioinformatics function based on RegulomeDB and Haploreg. For examples rs1555195 has an effect on enhancer histone marks, DNAse and motifs while rs11660748 and rs73440898 have an effect on enhancer histone marks and motifs (Supplementary Table 5 and Supplementary Figure 7).

Differential mRNA expression analysis

We assessed mRNA expression levels of the two genes in 109 pairs of NSCLC tumor and adjacent normal tissue samples available in the TCGA database. As shown in Supplementary Figure 8a, 8b and 8c, compared with adjacent normal tissues, the Mrna expression levels of *KIF16B* were no difference in all tumor tissue samples (P = 0.449) but lower in lung adenocarcinoma (LUAD) (P = 0.002) and higher in lung squamous cell carcinoma (LUSC) (P = 0.076). The higher expression levels of *KIF16B* mRNA were associated with a better survival in LUAD patients (Supplementary Figure 8e) but a worse survival in LUSC patients (Supplementary Figure 8f). Compared with adjacent normal tissues, mRNA expression levels of *NEDD4L* were lower in all tumor tissue samples as well as in LUAD and LUSC samples (P < 0.0001, P < 0.0001 and P < 0.0001, respectively) (Supplementary Figure 9a, 9b and 9c). The higher expression levels of *NEDD4L* mRNA were associated with a better survival in LUAD patients but again a worse survival in LUSC patients (Supplementary Figure 9e and 9f).

Discussion

In the present study, we assessed associations between SNPs in the endosome-related gene-set and NSCLC survival by using available genotyping data from two published GWAS datasets. We identified and validated three independent SNPs (i.e., *KIF16B* rs1555195, *NEDD4L* rs11660748 and rs73440898) that were significantly associated with NSCLC survival in Caucasian populations. In subsequent eQTL analysis for functional genotype-mRNA expression correlation, we found that the *KIF16B* rs1555195T allele was associated with lower mRNA expression levels in normal lung tissues and whole blood cells. Based on the TCGA database, *KIF16B* appears to be a potential oncogene, and we also found that the rs1555195T allele was associated with a lower risk of death and a lower mRNA expression level of *KIF16B*. However, this conclusion is consistent with the observation in LUSC but LUAD, and this discrepancy is likely due to small numbers of tumor samples included in the analysis or a difference at the transcriptomic level between LUSC and LUAD [34]; and other possible reasons may be differences in the molecular mechanisms of carcinogenesis [35-37] or therapies for these two tumors [38].

Both rs11660748G and rs73440898G alleles in *NEDD4L* were found to be associated with a higher risk of death. However, we did not find eQTL evidence to support the relationship between the two SNPs and the mRNA expression of *NEDD4L*. According to the results from the differential mRNA expression analysis, *NEDD4L* is more likely to be a suppresser gene in LUAD, but also possibly an oncogene in LUSC, considering that a higher expression of *NEDD4L* was associated with a better survival in LUAD patients but a worse survival in LUSC patients. This differentiation may be due to the difference in tumor types as above-mentioned for *KIF16B*. Additional functional investigations are needed to further explore the differences between these two types of NSCLC.

KIF16B, located on chromosome 20, encodes a member of the superfamily of kinesin proteins (KIF), which drives a variety of microtubule-dependent motility events [39]. A key feature of *KIF16B* is the PX domain at the C terminus that could target the motor at early endosomes by binding to PI(3)P, and through that, *KIF16B* could transport early endosomes to the plus end of microtubules in a process regulated by the small GTPase Rab5 and its effector [40]. *KIF16B* overexpression could relocate early endosomes to the cell periphery and inhibit the transport to the degradative pathway [41]. Conversely, expression of dominant-negative mutants or ablation of *KIF16B* by RNAi caused the clustering of early endosomes to the perinuclear region, delayed receptor recycling to the plasma membrane, and accelerated degradation [41]. These suggest that *KIF16B*, by regulating the plus end motility of early endosomes, modulates the intracellular localization of early endosomes and the balance between receptor recycling and degradation [41]. Overall, *KIF16B* expression affects the presentation of intracellular antigens by alternating early endosome location. However, few studies about *KIF16B* and lung cancer have been reported. One study reported that downregulation of *KIF16B* was found to be associated with brain metastasis in LUAD [42]. *NEDD4L*, located on chromosome 18, encodes a ubiquitin ligase belonging to the NEDD4 family of E3 HECT domain ubiquitin ligases [43,44]. *NEDD4L* proteins are known to be involved in regulating many membrane proteins via ubiquitination and endocytosis [45]. *NEDD4L* binds through its WW domains to the PY motifs of the epithelial Na+ channel (ENaC), leading to ENaC ubiquitylation, endocytosis to endosomes and multivesicular bodies, and degradation [44]. Overall, *NEDD4L* expression affects the presentation of intracellular antigens by alternating endosome forming and degradation. Few studies about *NEDD4L* and lung cancer have been reported. For example, one study found that in NSCLC patients with low *NEDD4L* expression, their prognoses were significantly poorer than those with high *NEDD4L* expression [46]. It was found that miR-93 could promote TGF-β-induced epithelial-to-mesenchymal transition through downregulation of *NEDD4L* in lung cancer cells [47] and that *NEDD4L* acted as a tumor suppressor gene in NSCLC and targeting EZH2 could upregulate *NEDD4L* expression [48]. There are no reports about the role of genetic variants of *NEDD4L* in the survival of NSCLC patients.

Although few studies about the relationship between *KIF16B* or *NEDD4L* and lung cancer have been reported, the relevant correlation between the endosome and immunotherapy for lung cancer have been well studied. For example, exogenous antigens including tumor antigens are taken up by antigen-presenting cells (APCs) and are degraded in endosome/lysosomes [49]. They are subsequently degraded antigenic peptide and are bound to MHC class II molecules. These antigenic peptide/MHC class II complexes are presented to CD4-

positive T cells, which engenders helper T cell-based humoral immune responses. A part of the exogenous antigen is also carried onto MHC class I molecules via transferring from the endosome to cytosol or in an early endosomes to engender CTL-based cellular immune responses. This presentation process of exogenous antigen is known as “cross-presentation” [50]. Therefore, the delivery of antigen into APCs in the body and the control of intracellular distribution of antigen in these cells for the induction of antigen-specific CTLs are crucially important to achieve cancer immunotherapy.

There are several limitations in the present study. Firstly, although several genetic variants backed up with *in silico* functional evidence in the endosome-related genes were found to be associated with NSCLS survival, the exact molecular mechanisms of these SNPs underlying the observed associations are still unclear. Secondly, both discovery and validation datasets were from Caucasian populations; therefore, our results may not be generalizable to other ethnic populations. Thirdly, though some clinical factors were available in the analysis for the PLCO but not HLCS datasets, there are still some information, such as the performance status, nutritional status and specific treatments such as immunotherapy, that was not available for further adjustment. However, our findings provided new insights for additional functional studies to further support these genetic variants of endosome pathway genes as promising predictors of survival in NSCLC patients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

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. We 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

This work was supported by the National Institute of Health [CA090578, CA074386, CA092824, U01CA209414, and R56AG062302]; the Duke Cancer institute as part of the P30 Cancer Center Support Grant [NIH/NCI CA014236]; the V Foundation for Cancer Research [D2017-19].

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

DSS	disease-specific 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
EAF	effect allele frequency
HR	hazards ratio
CI	confidence interval
AUC	area under the receiver operating characteristic curve
KIF16B	kinesin family member 16B
NEDD4L	neural precursor cell expressed developmentally downregulated gene 4-like
APC	antigen-presenting cell

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015; 65: 87–108 [PubMed: 25651787]
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019; 69: 7–34 [PubMed: 30620402]
3. Govindan R, Page N, Morgensztern D, Read W, Tierney R, Vlahiotis A, Spitznagel EL, Piccirillo J. Changing epidemiology of small-cell lung cancer in the United States over the last 30 years: analysis of the surveillance, epidemiologic, and end results database. J Clin Oncol 2006; 24: 4539–44 [PubMed: 17008692]
4. Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. Transl Lung Cancer Res 2016; 5: 288–300 [PubMed: 27413711]
5. Jia M, Zhu M, Zhou F, Wang M, Sun M, Yang Y, Wang X, Wang J, Jin L, Xiang J, Zhang Y, Chang J, Wei Q. 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: 807–817 [PubMed: 27861856]
6. Zienolddiny S, Skaug V. Single nucleotide polymorphisms as susceptibility, prognostic, and therapeutic markers of nonsmall cell lung cancer. Lung Cancer (Auckl) 2012; 3: 1–14 [PubMed: 28210120]
7. Stenzel-Bembenek A, Sagan D, Guz M, Stepulak A. [Single nucleotide polymorphisms in lung cancer patients and cisplatin treatment]. Postepy Hig Med Dosw (Online) 2014; 68: 1361–73 [PubMed: 25531699]
8. Stoorvogel W, Strous GJ, Geuze HJ, Oorschot V, Schwartz AL. Late endosomes derive from early endosomes by maturation. Cell 1991; 65: 417–27 [PubMed: 1850321]
9. Heath WR, Carbone FR. Cross-presentation in viral immunity and self-tolerance. Nat Rev Immunol 2001; 1: 126–34 [PubMed: 11905820]

10. Liu Y, Zeng G. Cancer and innate immune system interactions: translational potentials for cancer immunotherapy. J Immunother 2012; 35: 299–308 [PubMed: 22495387]
11. Ostrand-Rosenberg S Immune surveillance: a balance between protumor and antitumor immunity. Curr Opin Genet Dev 2008; 18: 11–8 [PubMed: 18308558]
12. Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. Science 2011; 331: 1565–70 [PubMed: 21436444]
13. Ryu R, Ward KE. Atezolizumab for the First-Line Treatment of Non-small Cell Lung Cancer (NSCLC): Current Status and Future Prospects. Front Oncol 2018; 8: 277 [PubMed: 30087855]
14. Paz-Ares L, Luft A, Vicente D, Tafreshi A, Gumus M, Mazieres J, Hermes B, Cay Senler F, Csoszi T, Fulop A, Rodriguez-Cid J, Wilson J, Sugawara S, Kato T, Lee KH, Cheng Y, Novello S, Halmos B, Li X, Lubiniecki GM, Piperdi B, Kowalski DM, Investigators K-. Pembrolizumab plus Chemotherapy for Squamous Non-Small-Cell Lung Cancer. N Engl J Med 2018; 379: 2040–2051 [PubMed: 30280635]
15. 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, Garassino MC, Investigators K-. Pembrolizumab plus Chemotherapy in Metastatic Non-Small-Cell Lung Cancer. N Engl J Med 2018; 378: 2078–2092 [PubMed: 29658856]
16. 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, Brahmer JR, Investigators K-. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. N Engl J Med 2016; 375: 1823–1833 [PubMed: 27718847]
17. Kobayashi KS, van den Elsen PJ. NLRC5: a key regulator of MHC class I-dependent immune responses. Nat Rev Immunol 2012; 12: 813–20 [PubMed: 23175229]
18. 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, Berg CD, Team PP. Lung cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. J Natl Cancer Inst 2010; 102: 722–31 [PubMed: 20442215]
19. Oken MM, Marcus PM, Hu P, Beck TM, Hocking W, Kvale PA, Cordes J, Riley TL, Winslow SD, Peace S, Levin DL, Prorok PC, Gohagan JK, Team PP. Baseline chest radiograph for lung cancer detection in the randomized Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. J Natl Cancer Inst 2005; 97: 1832–9 [PubMed: 16368945]
20. Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, Lee M, Popova N, Sharopova N, Kimura M, Feolo M. NCBI's Database of Genotypes and Phenotypes: dbGaP. Nucleic Acids Res 2014; 42: D975–9 [PubMed: 24297256]
21. 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, Sherry ST. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet 2007; 39: 1181–6 [PubMed: 17898773]
22. Zhai R, Yu X, Wei Y, Su L, Christiani DC. Smoking and smoking cessation in relation to the development of co-existing non-small cell lung cancer with chronic obstructive pulmonary disease. Int J Cancer 2014; 134:961–70 [PubMed: 23921845]
23. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics 2007; 23: 1294–6 [PubMed: 17384015]
24. Wakefield J A Bayesian measure of the probability of false discovery in genetic epidemiology studies. Am J Hum Genet 2007; 81: 208–27 [PubMed: 17668372]
25. Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. Stat Med 2006; 25: 3474–86 [PubMed: 16220486]
26. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T, Ferreira PG, Barann M, Wieland T, 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, Dermitzakis ET. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013; 501: 506–11 [PubMed: 24037378]

27. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015; 348: 648–60 [PubMed: 25954001]
28. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res 2009; 37: W600–5 [PubMed: 19417063]
29. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012; 22: 1790–7 [PubMed: 22955989]
30. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res 2016; 44: D877–81 [PubMed: 26657631]
31. Gyorffy B, Surowiak P, Budczies J, Lanczky A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. PLoS One 2013;8: e82241 [PubMed: 24367507]
32. Wang Y, Liu H, Ready NE, Su L, Wei Y, Christiani DC, Wei Q. Genetic variants in ABCG1 are associated with survival of nonsmall-cell lung cancer patients. Int J Cancer 2016; 138: 2592–601 [PubMed: 26757251]
33. Schwarz G Estimating the Dimension of a Model. The Annals of Statistics 1978; 6: 461–464
34. Relli V, Trerotola M, Guerra E, Alberti S. Abandoning the Notion of Non-Small Cell Lung Cancer. Trends Mol Med 2019;
35. Chang JT, Lee YM, Huang RS. The impact of the Cancer Genome Atlas on lung cancer. Transl Res 2015; 166: 568–85 [PubMed: 26318634]
36. Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. Nature 2012; 489: 519–25 [PubMed: 22960745]
37. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014; 511: 543–50 [PubMed: 25079552]
38. Rekhtman N, Paik PK, Arcila ME, Tafe LJ, Oxnard GR, Moreira AL, Travis WD, Zakowski MF, Kris MG, Ladanyi M. Clarifying the spectrum of driver oncogene mutations in biomarker-verified squamous carcinoma of lung: lack of EGFR/KRAS and presence of PIK3CA/AKT1 mutations. Clin Cancer Res 2012; 18: 1167–76 [PubMed: 22228640]
39. Goldstein LS. Molecular motors: from one motor many tails to one motor many tales. Trends Cell Biol 2001; 11: 477–82 [PubMed: 11719052]
40. Blatner NR, Wilson MI, Lei C, Hong W, Murray D, Williams RL, Cho W. The structural basis of novel endosome anchoring activity of KIF16B kinesin. EMBO J 2007; 26: 3709–19 [PubMed: 17641687]
41. Hoepfner S, Severin F, Cabezas A, Habermann B, Runge A, Gillooly D, Stenmark H, Zerial M. Modulation of receptor recycling and degradation by the endosomal kinesin KIF16B. Cell 2005; 121: 437–50 [PubMed: 15882625]
42. Singh M, Venugopal C, Tokar T, McFarlane N, Subapanditha MK, Qazi M, Bakhshinyan D, Vora P, Murty NK, Jurisica I, Singh SK. Therapeutic Targeting of the Premetastatic Stage in Human Lung-to-Brain Metastasis. Cancer Res 2018; 78: 5124–5134 [PubMed: 29986997]
43. Yang B, Kumar S. Nedd4 and Nedd4-2: closely related ubiquitin-protein ligases with distinct physiological functions. Cell Death Differ 2010; 17: 68–77 [PubMed: 19557014]
44. Rotin D, Kumar S. Physiological functions of the HECT family of ubiquitin ligases. Nat Rev Mol Cell Biol 2009; 10: 398–409 [PubMed: 19436320]
45. Harvey KF, Kumar S. Nedd4-like proteins: an emerging family of ubiquitin-protein ligases implicated in diverse cellular functions. Trends Cell Biol 1999; 9: 166–9 [PubMed: 10322449]

46. Sakashita H, Inoue H, Akamine S, Ishida T, Inase N, Shirao K, Mori M, Mimori K. Identification of the NEDD4L gene as a prognostic marker by integrated microarray analysis of copy number and gene expression profiling in non-small cell lung cancer. Ann Surg Oncol 2013; 20 Suppl 3: S590–8 [PubMed: 23812770]

47. Qu MH, Han C, Srivastava AK, Cui T, Zou N, Gao ZQ, Wang QE. miR-93 promotes TGF-beta-induced epithelial-to-mesenchymal transition through downregulation of NEDD4L in lung cancer cells. Tumour Biol 2016; 37: 5645–51 [PubMed: 26581907]

48. Wang X, Duan J, Fu W, Yin Z, Sheng J, Lei Z, Wang H. Decreased expression of NEDD4L contributes to NSCLC progression and metastasis. Biochem Biophys Res Commun 2019; 513: 398–404 [PubMed: 30967264]

49. Mellman I, Steinman RM. Dendritic cells: specialized and regulated antigen processing machines. Cell 2001; 106: 255–8 [PubMed: 11509172]

50. Joffre OP, Segura E, Savina A, Amigorena S. Cross-presentation by dendritic cells. Nat Rev Immunol 2012; 12: 557–69 [PubMed: 22790179]

Novelty and Impact:

We looked through genotyping datasets from two genome-wide association studies for genes involved in the endosome-related pathway and their associations with NSCLC survival. We found that three genetic variants of two genes were associated with survival of non-small cell lung cancer. The survival-associated variant T genotypes of rs1555195 were also associated with mRNA expression levels of the *KIF16B* gene. These variants could be useful predictors of NSCLC survival, and further functional studies could uncover the roles of these genes in the development of lung cancer.

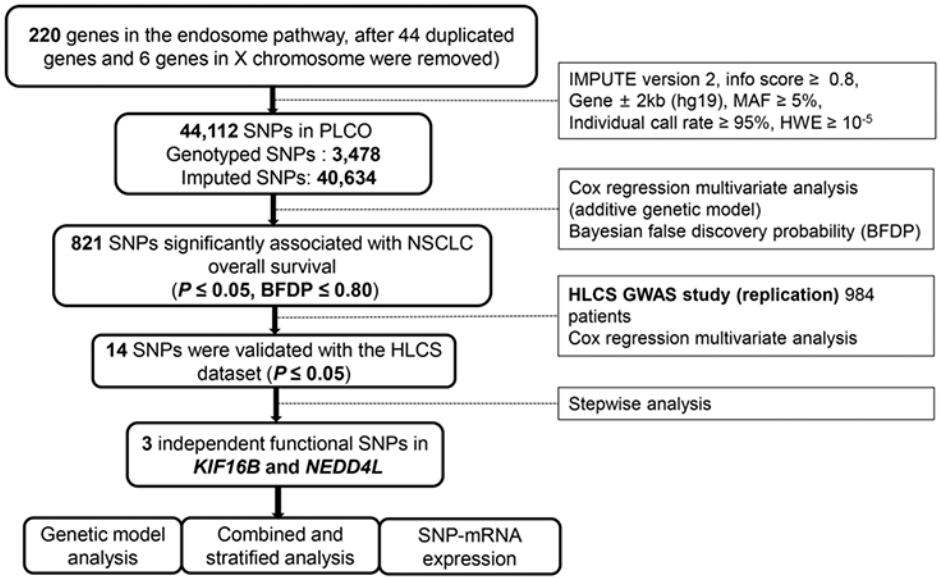

Figure 1.

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; KIF16B, kinesin family member 16B; NEDD4L, neural precursor cell expressed developmentally downregulated gene 4-like.

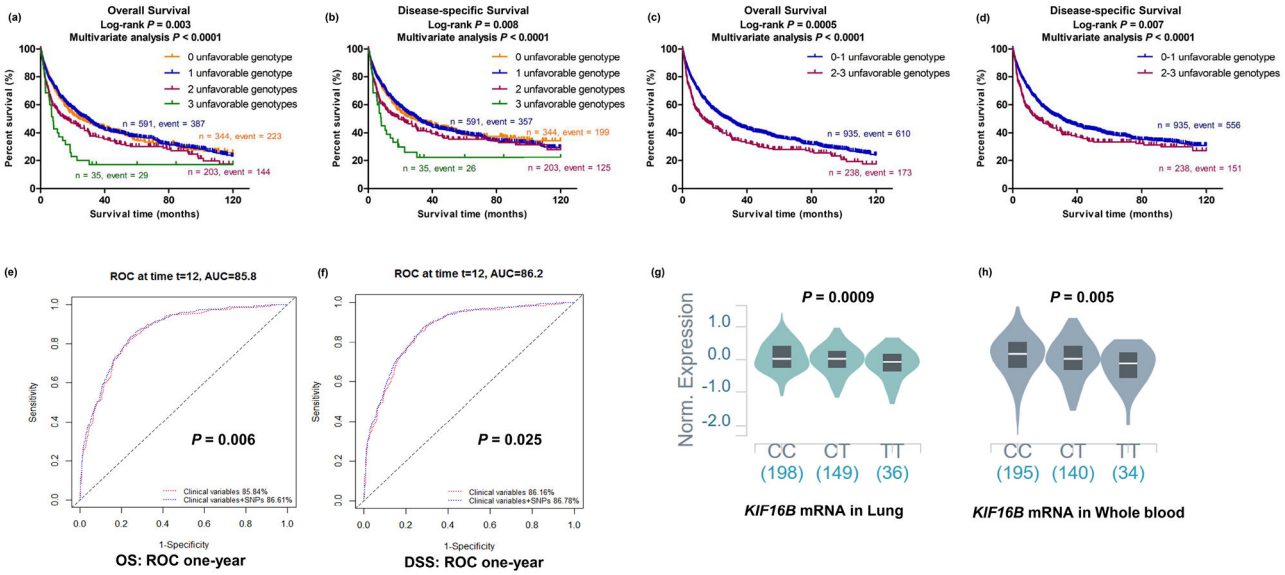

Figure 2.

Prediction of 10-year survival with combined unfavorable genotypes and eQTL for KIF16B rs1555195. Kaplan-Meier survival curves for the 10-year OS in the PLCO dataset for (a) the combined unfavorable genotypes and (b) dichotomized groups of the NUGs; Kaplan-Meier survival curves for the 10-year DSS in the PLCO dataset for (c) the combined unfavorable genotypes and (d) dichotomized groups of the NUGs. One-year NSCLC OS prediction by ROC curve (e) and one-year NSCLC DSS prediction by ROC curve (f). The correlation of rs1555195 genotypes and corresponding mRNA expression levels in the GTEx Project was significant in(g) normal lung tissue (P = 0.0009) and (h) whole blood cells (P = 0.005). Abbreviations: eQTL, expression quantitative trait loci; OS, overall survival; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening trial; NUG, number of unfavorable genotypes; DSS, disease-specific survival; KIF16B, kinesin family member 16B; ROC, receiver operating characteristic curve.

Table 1.

The 14 validated and survival-associated significant SNPs in two previously published NSCLC GWAS datasets

SNP	Allele[a]	Gene	PLCO (n=1185)					HLCS (n=984)			Combined-analysis			
			FDR	BFDP	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]
rs1555195	**C>T**	***KIF16B***	**0.29**	**0.66**	**0.26**	**0.84 (0.74-0.94)**	**0.003**	**0.28**	**0.89 (0.79-1.00)**	**0.049**	**0.520**	**0**	**0.86 (0.79-0.94)**	**0.0007**
rs71355689	T>C	*NEDD4L*	0.29	0.23	0.12	1.32 (1.14-1.54)	0.0003	0.12	1.38 (1.16-1.63)	0.0003	0.722	0	1.34 (1.20-1.50)	2.9E-7
rs11660199	G>A	*NEDD4L*	0.29	0.17	0.12	1.33 (1.15-1.55)	0.0002	0.12	1.36 (1.15-1.61)	0.0003	0.834	0	1.34 (1.20-1.50)	2.2E-7
rs11665627	T>C	*NEDD4L*	0.29	0.33	0.18	1.25 (1.09-1.42)	0.0009	0.18	1.25 (1.08-1.44)	0.002	0.991	0	1.25 (1.13-1.38)	7.1E-6
rs9957736	G>A	*NEDD4L*	0.37	0.75	0.17	1.21 (1.05-1.38)	0.007	0.17	1.24 (1.07-1.44)	0.004	0.792	0	1.23 (1.11-1.35)	7.2E-5
rs60605848	C>G	*NEDD4L*	0.29	0.56	0.17	1.23 (1.08-1.40)	0.002	0.18	1.23 (1.06-1.42)	0.005	0.999	0	1.23 (1.12-1.35)	2.5E-5
rs12960902	T>A	*NEDD4L*	0.29	0.66	0.17	1.22 (1.07-1.39)	0.003	0.17	1.23 (1.06-1.42)	0.005	0.951	0	1.22 (1.11-1.35)	4.6E-5
rs60418930	G>A	*NEDD4L*	0.29	0.56	0.17	1.23 (1.08-1.40)	0.002	0.17	1.22 (1.06-1.41)	0.006	0.961	0	1.23 (1.11-1.35)	3.1E-5
rs59402591	A>G	*NEDD4L*	0.29	0.56	0.17	1.23 (1.08-1.40)	0.002	0.17	1.22 (1.06-1.41)	0.006	0.960	0	1.23 (1.11-1.35)	3.2E-5
rs17064520	C>T	*NEDD4L*	0.29	0.56	0.17	1.23 (1.08-1.40)	0.002	0.17	1.22 (1.06-1.41)	0.006	0.960	0	1.23 (1.11-1.35)	3.2E-5
rs11660748	A>G	*NEDD4L*	0.29	0.05	0.11	1.36 (1.17-1.58)	8.2E-5	0.11	1.23 (1.03-1.48)	0.024	0.409	0	1.31 (1.16-1.47)	6.0E-5
rs73440898	A>G	*NEDD4L*	0.29	0.50	0.10	1.31 (1.11-1.55)	0.002	0.11	1.23 (1.02-1.47)	0.030	0.596	0	1.27 (1.12-1.44)	0.0001
rs7576673	A>G	*ERBB4*	0.34	0.75	0.18	1.21 (1.06-1.38)	0.006	0.19	1.17 (1.02-1.34)	0.026	0.734	0	1.19 (1.08-1.31)	0.0003
rs10932385	C>G	*ERBB4*	0.32	0.75	0.18	1.21 (1.06-1.38)	0.005	0.18	1.17 (1.02-1.34)	0.027	0.728	0	1.19 (1.08-1.31)	0.0003

Abbreviations: 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. FDR, false discovery rate; BFDP, Bayesian false discovery probability; EAF, effect allele frequency; HR, hazards ratio; CI, confidence interval

[a] Reference>effect allele

[b] Obtained from an additive genetic model with adjustment for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3, and PC4

[c] Obtained from an additive genetic model with adjustment 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 fixed-effects model.

Table 2.

Three indenpenden SNPs from multivariate Cox regression analysis of selected functional variables and previous published SNPs in the PLCO GWAS dataset

Variables	Category	Frequency	HR (95% CI) [a]	*P* [a]	HR (95% CI) [b]	*P* [b]
Age	Continuous	1185	1.03 (1.02-1.05)	<0.0001	1.04 (1.02-1.05)	<0.0001
Sex	Male	698	1.00		1.00	
	Female	487	0.78 (0.67-0.91)	0.001	0.78 (0.66-0.91)	0.002
Smoking status	Never	115	1.00		1.00	
	Current	423	1.69 (1.26-2.26)	0.0004	1.94 (1.44-2.62)	<0.0001
	Former	647	1.65 (1.26-2.18)	0.0003	1.89 (1.42-2.51)	<0.0001
Histology	AD	577	1.00		1.00	
	SC	285	1.14 (0.95-1.38)	0.163	1.20 (0.99-1.46)	0.064
	Others	323	1.32 (1.11-1.56)	0.002	1.37 (1.14-1.63)	0.0006
Stage	I-IIIA	655	1.00		1.00	
	IIIB-IV	528	2.82 (2.32-3.43)	<0.0001	3.00 (2.46-3.66)	<0.0001
Chemotherapy	No	639	1.00		1.00	
	Yes	538	0.58 (0.49-0.69)	<0.0001	0.58 (0.48-0.70)	<0.0001
Radiotherapy	No	762	1.00		1.00	
	Yes	415	0.97 (0.82-1.14)	0.724	0.97 (0.82-1.15)	0.738
Surgery	No	637	1.00		1.00	
	Yes	540	0.21 (0.16-0.27)	<0.0001	0.19 (0.15-0.25)	<0.0001
KIF16B rs1555195 C>T	CC/CT/TT	640/466/79	0.83 (0.74-0.94)	0.002	0.86 (0.76-0.97)	0.017
NEDD4L rs11660748 A>G	AA/AG/GG	937/229/19	1.34 (1.15-1.56)	0.0002	1.26 (1.07-1.47)	0.005
NEDD4L rs73440898 A>G	AA/AG/GG	959/216/8	1.28 (1.08-1.51)	0.004	1.28 (1.07-1.52)	0.006

Abbreviations: SNP: single-nucleotide polymorphism; PLCO, Prostate, Lung, Colorectal and Ovarian cancer screening trial; GWAS, genome-wide association study; HR: hazards ratio; CI: confidence interval

[a] Stepwise analysis included age, sex, smoking status, tumor stage, histology, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3, PC4 and SNPs.

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

Table 3.

Associations between three significantly independent SNPs and 10-year survival of NSCLC in the PLCO Trial

Genotype	Frequency	OS[a]			DSS[a]		
		Death (%)	HR (95% CI)	*P*	Death (%)	HR (95% CI)	*P*
KIF16B rs1555195 C>T[b]							
CC	636	432 (67.92)	1.00		391 (61.48)	1.00	
CT	460	306 (66.52)	0.83 (0.72-0.97)	0.016	277 (60.22)	0.82 (0.70-0.96)	0.016
TT	79	46 (58.23)	0.71 (0.53-0.97)	0.031	40 (50.63)	0.70 (0.50-0.96)	0.029
Trend test				0.003			0.003
Dominant							
CC	636	432 (67.92)	1.00		391 (61.48)	1.00	
CT+TT	539	352 (65.31)	0.81 (0.71-0.94)	0.005	317 (58.81)	0.80 (0.69-0.94)	0.005
NEDD4L rs11660748 A>G[c]							
AA	929	609 (65.55)	1.00		550 (59.20)	1.00	
AG	227	159 (70.04)	1.33 (1.11-1.60)	0.002	144 (63.44)	1.34 (1.11-1.61)	0.003
GG	19	16 (84.21)	1.98 (1.20-3.29)	0.008	14 (73.68)	1.89 (1.10–3.24)	0.021
Trend test				<0.0001			0.0003
Dominant							
AA	929	609 (65.55)	1.00		550 (59.20)	1.00	
AG+GG	246	175 (71.14)	1.37 (1.16-1.63)	0.0003	158 (64.23)	1.37 (1.14-1.65)	0.0006
NEDD4L rs73440898 A>G[d]							
AA	952	628 (65.97)	1.00		571 (59.98)	1.00	
AG	213	149 (69.95)	1.30 (1.09-1.56)	0.004	131 (61.50)	1.23 (1.02-1.49)	0.034
GG	8	6 (75.00)	2.05 (0.91-4.61)	0.083	5 (62.50)	1.91 (0.79-4.63)	0.154
Trend test				0.001			0.015
Dominant							
AA	952	628 (65.97)	1.00		571 (59.98)	1.00	
AG+GG	221	155 (70.14)	1.32 (1.11-1.58)	0.002	136 (61.54)	1.25 (1.03-1.51)	0.022
NUG[e,f]							
0	344	223 (64.83)	1.00		199 (57.85)	1.00	

Genotype	Frequency	OS[a]			DSS[a]		
		Death (%)	HR (95% CI)	*P*	Death (%)	HR (95% CI)	*P*
1	591	387 (65.48)	1.19 (1.01-1.41)	0.099	357 (60.41)	1.24 (1.04-1.49)	0.016
2	203	144 (70.94)	1.71 (1.38-2.12)	<0.0001	125 (61.58)	1.64 (1.30-2.06)	<0.0001
3	35	29 (82.86)	2.07 (1.40-3.06)	0.0003	26 (74.29)	2.05 (1.35-3.09)	0.0007
Trend test				<0.0001			<0.0001
0-1	935	610 (65.24)	1.00		556 (59.47)	1.00	
2-3	238	173 (72.69)	1.58 (1.33-1.87)	<0.0001	151 (63.45)	1.48 (1.23-1.78)	<0.0001

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; NUG: number of unfavorable genotypes.

[a]Adjusted for age, sex, smoking status, histology, tumor stage, chemotherapy, surgery, and principal components.

[b]10 missing date were excluded

[c]10 missing date were excluded

[d]12 missing date were excluded

[e]12 missing date were excluded

[f]Unfavorable genotypes were *KIF16B* rs1555195 CC, *NEDD4L* rs73440898 AG+GG, *NEDD4L* rs11660748 AG+GG.

Table 4.

Haplotype analysis of association between two SNPs in *NEDD4L* and NSCLC 10-year survival in PLCO

Haplotypes[a, b]	Haplotype frequency		Multivariate Analysis[c] for OS		Multivariate Analysis[c] for DSS	
	N	%	HR (95% CI)	*P*	HR (95% CI)	*P*
A-A	1933	82.40	1.00		1.00	
A-G	192	8.18	1.32 (1.10-1.58)	0.002	1.32 (1.09-1.59)	0.004
G-A	168	7.16	1.27 (1.05-1.54)	0.013	1.20 (0.98-1.47)	0.081
G-G	53	2.26	1.46 (1.06-2.02)	0.022	1.43 (1.01-2.01)	0.043
Trend test				<0.0001		0.001
A-A	1933	82.40	1.00		1.00	
(A-G)+(G-A)+(G-G)	413	17.60	1.32 (1.16-1.50)	<0.0001	1.28 (1.12-1.47)	0.0004

Abbreviations: SNP, single nucleotide polymorphism; NSCLC, non-small cell lung cancer; PLCO, the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; OS, overall survival; DSS, disease-specific survival; HR, hazards ratio; CI, confidence interval.

[a]The alleles in the haplotype were ranked in the SNP order of rs rs73440898A>G and rs11660748A>G

[b]12 missing date were excluded

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