Received: 8 January 2022 | Revised: 5 March 2022 | Accepted: 10 March 2022

DOI: 10.1002/mc.23400

RESEARCH ARTICLE

Molecular Carcinogenesis WILEY

Genetic variants in *DDO and PEX5L* in peroxisome-related pathways predict non-small cell lung cancer survival

Allan S. Chen[1,2] | Hongliang Liu[1,2] | Yufeng Wu[1,2] | Sheng Luo[3] | Edward F. Patz Jr.[1,4] | Carolyn Glass[1,5] | Li Su[6] | Mulong Du[6] | David C. Christiani[6,7] | Qingyi Wei[1,2,8,9]

[1]Duke Cancer Institute, Duke University Medical Center, Durham, North Carolina, USA

[2]Department of Population Health Sciences, Duke University School of Medicine, Durham, North Carolina, USA

[3]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina, USA

[4]Departments of Radiology, Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina, USA

[5]Department of Pathology, Duke University School of Medicine, Durham, North Carolina, USA

[6]Departments of Environmental Health and Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA

[7]Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA

[8]Department of Medicine, Duke University School of Medicine, Durham, North Carolina, USA

[9]Duke Global Health Institute, Duke University, Durham, North Carolina, USA

Correspondence
Qingyi Wei, Department of Population Health Sciences, Duke University School of Medicine, 905S LaSalle St, Durham, NC 27710, USA.
Email: qingyi.wei@duke.edu

Funding information
Duke Cancer Institute; V Foundation for Cancer Research; National Institutes of Health

Abstract

Peroxisomes play a role in lipid metabolism and regulation of reactive oxygen species, but its role in development and progression of non-small cell lung cancer (NSCLC) is not well understood. Here, we investigated the associations between 9708 single-nucleotide polymorphisms (SNPs) in 113 genes in the peroxisome-related pathways and survival of NSCLC patients from the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) and the Harvard Lung Cancer Susceptibility (HLCS) study. In 1185 NSCLC patients from the PLCO trial, we found that 213 SNPs were significantly associated with NSCLC overall survival (OS) ($p \leq 0.05$, Bayesian false discovery probability [BFDP] ≤ 0.80), of which eight SNPs were validated in the HLCS data set. In a multivariate Cox proportional hazards regression model, two independent SNPs (rs9384742 *DDO* and rs9825224 *PEX5L*) were significantly associated with NSCLC survival (hazards ratios [HR] of 1.17 with 95% CI [confidence interval] of 1.06–1.28 and 0.86 with 95% CI of 0.77–0.96, respectively). Patients with one or two protective genotypes had a significantly higher OS (HR: 0.787 [95% CI: 0.620–0.998] and 0.691 [95% CI: 0.543–0.879], respectively). Further expression quantitative trait loci analysis using whole blood and lung tissue showed that the minor allele of rs9384742 *DDO* was significantly associated with decreased messenger RNA (mRNA) expression levels and that *DDO* expression was also decreased in NSCLC tumor tissue. Additionally, high *PEX5L* expression levels were significantly associated with lower survival of NSCLC. Our data suggest that variants in these peroxisome-related genes may influence gene regulation and are potential predictors of NSCLC OS, once validated by additional studies.

KEYWORDS
bioinformatics, biostatistics, lung cancer, peroxisomes

1 | INTRODUCTION

Lung cancer is the leading cause of cancer deaths, making up an estimated 18.4% of total cancer deaths with 1.76 million deaths in 2018, and a 5-year survival rate of 19.4% in 2019.[1] Non-small cell lung cancer (NSCLC) makes up 85% of all lung cancers, with two primary subtypes being lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), comprising 50% and 40% of NSCLC, respectively.[2] Between 2010 and 2016, the overall 5-year survival rate of NSCLC was 25%, even with advances in

treatment, and thus, new strategies are needed to improve the outcomes.[3]

There are an estimated 10 million single-nucleotide polymorphisms (SNPs) located in the human genome, and over the last two decades, genome-wide association studies (GWAS) have identified many susceptibility loci for lung cancer.[4,5] However, more than 90% of SNPs are in noncoding regions, and their biological functions are unknown.[6] Additionally, tumor evolution and genetic heterogeneity further complicate genetic influences on cancer progression.[7] In the post-GWAS era, pathway-based approaches to utilizing genetic variation linked to specific biological pathways may provide a better understanding of the progression and treatment response of lung cancer.[8-9]

The peroxisome is a cellular organelle involved in a variety of functions, such as lipid metabolism, bile acid synthesis, plasmalogen synthesis, and maintaining homeostasis of reactive oxygen species (ROS).[11] While research into the role of peroxisomes in cancer is relatively new, metabolic reprogramming of tumors through peroxisomes is believed to contribute to cancer progression.[12] For example, high levels of ether phospholipids and plasmalogens, synthesized by peroxisomes, are found in cancer cells.[13] Arachidonic acid and docasahexaenoic acid are attached to plasmalogens; while arachidonic acid has been shown to suppress apoptosis and to promote cancer cell proliferation, docasahexaenoic acid induces apoptosis and reduces cancer cell proliferation.[13] In addition, alpha-methylacyl-CoA racemase and peroxisomal acyl-coenzyme A oxidase 1 are key enzymes in beta-oxidation of fatty acids, and they are also associated with a worse survival of various cancer types.[14,15] Peroxisomes are also essential for balancing oxidative stress by producing ROS, such as hydrogen peroxide H_2O_2, through byproducts of beta-oxidation and decreasing ROS through regulating levels of catalase, a ROS-decomposing enzyme.[14] Furthermore, ROS are known regulators of cancer, influencing proliferation, angiogenesis, immunity, and apoptosis.[14]

While peroxisomes appear to have an important role in cancer growth and metastasis, few studies have specifically investigated peroxisomes in lung cancer progression.[16] Zhang et al.[17] investigated 38 common genes in the peroxisome pathway and found three genes that have clinical significance in NSCLC; however, no genetic information was examined. By using existing genotyping data from two independent GWAS, we investigated associations between SNPs in peroxisome-related pathways and NSCLC survival.

2 | MATERIALS AND METHODS

2.1 | Study populations

The study population for the first-stage analysis was sourced from the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial initiated by the National Cancer Institute (NCI), which included 78,000 women and 77,000 men between the ages of 55 and 74, who were followed up between 1993 and 2001.[18] Background information was obtained for all patients, including age, sex, smoking status, and medical and family history.[18] The participants were randomly assigned to yearly screening for lung cancer for the first 3 years, and follow-up details of positive lung cancer screens were obtained, such as tumor stage, histology, chemotherapy, radiotherapy, and surgery.[19] Genotyping data were obtained from blood DNA samples genotyped by using Illumina HumanHap240Sv1.0, HumanHap300v1.1, and HumanHap550v3.0 (dbGaP accession: phs000093.v2.p2 and phs000336.v1.p1).[20,21] Of the samples, we used 1185 Caucasian NSCLC patients from the trial, including 487 women and 698 men, with varying histology, tumor stages, and treatments[22] (Table S1). The use of the PLCO data set was approved by the Internal Review Board of Duke University School of Medicine (#Pro00054575) and the dbGAP database administration (#6404, dbGaP accession: phs000336.v1.p1, and phs000093.v2.p2).

An additional 984 Caucasian NSCLC patients from Massachusetts General Hospital (MGH) as part of the Harvard Lung Cancer Susceptibility (HLCS) study were used for an independent validation of the PLCO data set[23](Table S2). Genotyping data for this group was obtained from blood DNA samples by using the Illumina Humanhap610-Quad array.[23] Imputation of additional SNPs was performed by using the MaCH software and data from the 1000 Genomes Project.[24,25] The prestudy was approved by the Institutional Review Board of the MGH and the Human Subjects Committee of the Harvard School of Public Health.

A flowchart detailing the present study is shown in Figure 1. We identified 113 genes in the peroxisome-related pathways through searching the keyword "peroxisome" in ontology gene sets of the Molecular Signatures Database v7.4[26,27] (Table S3). SNPs in these genes and the 2-kb flanking regions surrounding the genes were extracted from the genotyping data and imputation data, which was conducted by using Minimac4 based on the European data in the 1000 Genomes Project (Phase 3 release October 2014)[24,25] -(Figure S2). For quality control, SNPs were included only if they had a minor allelic frequency ≥ 0.05, a genotyping rate ≥0.95, Hardy –Weinberg equilibrium $P \geq 1 \times 10^{-5}$, and imputation $r^2 \geq 0.3$.[28,29]

We used multivariate Cox proportional hazards regression analysis through the GenABEL package in R 3.6.3 to estimate associations between SNPs and NSCLC overall survival (OS) in an additive genetic model for the PLCO data set.[30] We adjusted for clinical variables (age, sex, smoking status, tumor stage, histology, chemotherapy, radiotherapy, and surgery) and the top four principal components (PCs). Associations between the first 10 PCs and NSCLC OS are shown in Table S4. The primary endpoint in relation to the first diagnosis of lung cancer was either time of death or censored at time of the last follow-up, and we calculated the 95% confidence interval (CI) of the hazards ratio (HR) using the Cox regression model. To control for false positives in a multiple-hypothesis testing scenario, we used the threshold of Bayesian false-discovery probability (BFDP) < 0.800 to reduce the number of false positives as recommended.[31] We used a prior probability of 0.10 to detect an upper boundary HR of 3.0 for associations of minor alleles or variant genotypes of SNPs with $p < 0.05$. Significant SNPs were then

independently validated with the HLCS data set by using Cox regression model with adjustment for age, sex, smoking status, tumor stage, histology, chemotherapy, radiotherapy, surgery, and the first three PCs.

Meta-analysis on the results of the PLCO and HLCS data sets was performed through PLINK 1.9 using Cochran's *Q* test and I^2 for interstudy heterogeneity. Because little to no interstudy heterogeneity was found (Cochran's *Q* test $p > 0.10$, $I^2 < 0.25$ for all SNPs), we applied a fixed effects model for the combined analysis of these two data sets.

Linkage disequilibrium (LD) analysis was performed using Haploview and HaploReg v4.1 with information from the 1000 Genomes Project.[32] Manhattan plots were also generated by Haploview and regional association plots were constructed using Locus Zoom.[32,33] We evaluated functional annotation of SNPs using HaploReg v4.1, SNPinfo, and Regulome DB.[34-36]

We used SAS v9.4 software for stepwise multivariate Cox regression analysis with adjustment for clinical variables, the first four PCs, and 41 previously published SNPs from the same PLCO data set. Furthermore, we investigated associations of the combined genotypes of significant SNPs with the PLCO data set along with the stratified analysis by each clinical variable. We evaluated significant SNPs to a Kaplan–Meier (KM) model using the log-rank test to compare HRs. To further investigate predictive accuracy, we also constructed receiver operator characteristic (ROC) and time-dependent area under the curve (AUC) curves using the timeROC package of R 3.6.3.[37]

Finally, we used data from the Genotype-Tissue Expression (GTEx Analysis V7) project that included DNA and RNA sequencing data from 383 normal lung tissue and 369 whole blood samples for eQTL analysis to investigate associations between SNPs and messanger RNA (mRNA) expression levels.[38] We also used data from the 1000 Genomes Project that included DNA and RNA sequencing data from 373 Caucasian patients for further eQTL analysis.[39] We used the Kaplan-Meier (KM) method to investigate associations between dichotomized mRNA expression levels and OS of NSCLC in 1715 samples from 10 independent data sets.[40] Using The Cancer Genome Atlas (TCGA) database, we compared mRNA expression levels between tumor tissue in LUAD or LUSC and adjacent normal tissue.[41,42] We also used TCGA data with unpaired samples from UALCAN that included more mRNA expression data from LUAD, LUSC, and normal tissue samples.43 Using the cBioPortal for Cancer Genomics, we also investigated mutation rates of genes in lung cancer studies. 44,45

3 | RESULTS

We identified 9708 SNPs within 113 genes of peroxisome-related pathways and their 2-kb flanking regions in a single-locus multivariate Cox regression analysis, with adjustment for clinical variables and the first four PCs (Figure S3). After correcting for multiple-testing using BFDP, 213 of those SNPs remained significantly associated with NSCLC OS (BFDP < 0.80, $p < 0.05$). In the validation by the independent data set from the HLCS study, eight SNPs remained significantly associated with NSCLC OS. Further meta-analysis of the PLCO and HLCS studies revealed little to no interstudy heterogeneity for these eight SNPs, and the results of the combined analysis for the two data sets are shown in Table 1.

The LD analysis on the eight SNPs revealed that seven SNPs are in a 9-kb block of LD within the *PEX5L* gene, while the *DDO* rs9384742 SNP is independent (Figure S4). After adjustment for clinical variables and the first four PCs using a stepwise multivariate Cox regression model, two of these seven SNPs remained

TABLE 1 Associations of the eight validated SNPs with OS of NSCLC in the PLCO and HLCS studies.

				PLCO			HLCS			Combined		
SNP	Allele	Gene	MAF	HR (95% CI)[a]	*P*[a]	MAF	HR (95% CI)[b]	*P*[b]	P_{het}[c]	I^2	HR (95% CI)[d]	*P*[d]
rs9384742	C > T	*DDO*	0.47	1.17 (1.06–1.28)	0.002	0.48	1.13 (1.02–1.25)	0.022	0.305	0	1.15 (1.07–1.23)	0.0001
rs9825224	A > C	*PEX5L*	0.32	0.86 (0.77–0.96)	0.008	0.29	0.85 (0.76–0.96)	0.006	0.987	0	0.86 (0.79–0.93)	0.0001
rs9873162	T > A	*PEX5L*	0.32	0.86 (0.77–0.96)	0.009	0.28	0.86 (0.77–0.97)	0.011	0.515	0	0.86 (0.80–0.93)	0.0002
rs9855801	C > T	*PEX5L*	0.32	0.86 (0.77–0.96)	0.009	0.28	0.86 (0.77–0.97)	0.011	0.922	0	0.86 (0.80–0.93)	0.0002
rs10804880	C > T	*PEX5L*	0.31	0.86 (0.77–0.96)	0.008	0.28	0.88 (0.78–0.98)	0.026	0.366	0	0.87 (0.80–0.94)	0.0005
rs1534023	C > T	*PEX5L*	0.31	0.86 (0.78–0.96)	0.007	0.28	0.88 (0.79–0.99)	0.028	0.373	0	0.87 (0.80–0.94)	0.0005
rs1534024	C > T	*PEX5L*	0.31	0.86 (0.780–0.96)	0.007	0.28	0.88 (0.79–0.99)	0.029	0.322	0	0.87 (0.80–0.94)	0.0005
rs2140654	C > T	*PEX5L*	0.31	0.86 (0.77–0.96)	0.007	0.28	0.88 (0.79–0.99)	0.028	0.266	0	0.87 (0.80–0.94)	0.0005

Abbreviations: CI, confidence interval; HLCS, the Harvard Lung Cancer Susceptibility Study; HR, hazard ratio; MAF, minor allele frequency; NSCLC, non-small cell lung cancer; OS, overall survival; PLCO, the prostate, lung, colorectal and ovarian cancer screening trial; SNPs, single-nucleotide polymorphisms.

[a]Additive genetic model adjusting for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, and first four PCs.

[b]Additive genetic model adjusting for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, and first three PCs.

[c]P_{het}: Cochrane's *Q* test *p* value for study heterogeneity.

[d]Fixed-effects model for the meta-analysis.

significantly associated with NSCLC OS, including the *PEX5L* rs9825224 SNP ($p = 0.0052$) and the *DDO* rs9384742 SNP ($p = 0.0015$) (Table S5). These two SNPs were imputed in the PLCO data set with Minimac4 r^2 values (*PEX5L* rs9825224: 0.99239 and *DDO* rs9384742: 0.96611) and BFDP values (*PEX5L* rs9825224: 0.717 and *DDO* rs9384742: 0.782). *PEX5L* rs9825224 was genotyped in the HLCS study, while *DDO* rs9384742 was imputed with r^2 value: 0.983, with a BFDP value of 0.683 for *PEX5L* rs9825224: and 0.878 for *DDO* rs9384742. Further regional association plots for these two SNPs are shown in Figure S5. After addition of 41 previously published SNPs to the model, both SNPs remained significantly associated with NSCLC OS (*PEX5L* rs9825224 $p = 0.0018$; *DDO* rs9384742 $p = 0.0084$) (Table S5). Computational function prediction of these SNPs using HaploReg v4.1, SNPinfo and Regulome DB have revealed that *PEX5L* rs9825224 and *DDO* rs9384742 are located in an intronic region of their respective genes. Both SNPs are associated with enhancer and promoter regions in a variety of cell lines, likely to be located in DNase hypersensitivity regions, which may affect transcription factor binding. Further details of potential functional annotation are shown in Table S6.

As shown in Table 2, the *DDO* rs9384742 T minor allele was significantly associated with an increased risk of death for NSCLC

TABLE 2 Associations of *PEX5L* rs9825224, *DDO* rs9384742, and their combined genotypes with NSCLC survival.

		OS			DSS		
Genotype	**Frequency**	**Death (%)**	**HR (95% CI)[a]**	***p*[a]**	**Death (%)**	**HR (95% CI)[a]**	***p*[a]**
DDO rs9384742 C > T							
CC	327	219 (67.0)	1.00		193 (59.0)	1.00	
CT	581	381 (65.6)	1.236 (1.043–1.466)	0.015	342 (58.9)	1.241 (1.036–1.487)	0.015
TT	267	189 (70.79)	1.352 (1.110–1.648)	0.003	174 (65.2)	1.416 (1.150–1.743)	0.001
Trend				0.002			<0.001
CT/TT	848	570 (67.22)	1.273 (1.085–1.494)	0.003	516 (60.85)	1.296 (1.095–1.535)	0.003
PEX5L rs9825224 A > C							
AA	535	365 (68.2)	1.00		324 (60.6)	1.00	
AC	525	349 (66.5)	0.904 (0.780–1.047)	0.179	318 (60.6)	0.937 (0.802–1.095)	0.413
CC	115	70 (60.9)	0.683 (0.545–0.909)	0.007	67 (58.3)	0.741 (0.568–0.968)	0.028
Trend				0.008			0.041
AC/CC	640	419 (65.47)	0.863 (0.749–0.993)	0.040	385 (60.16)	0.897 (0.773–1.040)	0.151
NPA							
0	114	86 (75.44)	1.00		76 (66.67)	1.00	
1	397	262 (65.99)	0.843 (0.658–1.081)	0.1781	236 (59.45)	0.880 (0.676–1.146)	0.3426
2	449	303 (67.48)	0.736 (0.577–0.938)	0.0132	273 (60.80)	0.753 (0.582–0.974)	0.0307
3	188	125 (66.49)	0.633 (0.478–0.837)	0.0013	110 (58.51)	0.645 (0.479–0.868)	0.0038
4	37	22 (59.46)	0.505 (0.314–0.812)	0.0048	20 (54.05)	0.532 (0.323–0.876)	0.0132
Trend				<0.0001			0.0001
NPG							
0	114	86 (75.44)	1.00		76 (66.67)	1.00	
1	580	387 (66.72)	0.787 (0.620–0.998)	0.0477	348 (60.00)	0.813 (0.632–1.047)	0.1082
2	491	325 (66.19)	0.691 (0.543–0.879)	0.0026	291 (59.27)	0.710 (0.550–0.916)	0.0085
Trend				0.0030			0.0067
0	114	86 (75.44)	1.00		76 (66.67)		
1–2	1071	712 (66.48)	0.739 (0.589–0.927)	0.0090	639 (59.66)	0.761 (0.598–0.969)	0.0266

Note: NPA: Number of protective alleles calculated as number of following alleles (*DDO rs9384742*_C, *PEX5L rs9825224*_C).

NPG: Number of protective genotypes calculated as number of following genotypes: (DDO rs9384742 CT/CC, PEX5L rs9825224 CA/CC).

Abbreviations: CI, confidence interval; DSS, disease-specific survival; HR, hazard ratio; NPA, number of protective alleles; NPG: number of protective genotypes; NSCLC, non-small cell lung cancer; OS, overall survival.

[a]Multivariate Cox hazards regression adjusted for age, sex, smoking status, tumor stage, histology, chemotherapy, radiotherapy, surgery, first four PCs.

(OS: $P_{trend} = 0.002$ and DSS: $P_{trend} < 0.001$), while the *PEX5L* rs9825224 C minor allele exhibited a protective effect (*PEX5L* rs9825224 OS: $P_{trend} = 0.008$ and DSS: $P_{trend} = 0.041$). Patients with one or two minor alleles in a dominant model for *DDO* rs9384742 also had a significantly larger HR for OS and DSS of NSCLC than patients who were homozygous for the major allele, while patients with one or two minor alleles in a dominant model for *DDO* rs9384742 had a significantly smaller HR for OS only. We then looked at the number of protective alleles (NPA) (*PEX5L* rs9825224 C and *DDO* rs9384742 C) for each patient. Having a higher NPA was significantly associated with a better survival (OS: $P_{trend} < 0.0001$; DSS: $P_{trend} = 0.0001$). When looking at the number of protective genotypes (NPG) (*DDO* rs9384742 CT/CC, *PEX5L* rs9825224 CA/CC) from 0 to 2 protective genotypes for each patient, having a higher NPG was also associated with a better survival (OS: $P_{trend} = 0.0030$; DSS: $P_{trend} = 0.0067$). Dichotomizing NPG into two groups showed that patients with one or two protective genotypes had a significantly better survival than those with zero protective genotypes (OS HR = 0.739, 95% CI: 0.589–0.927, $p = 0.0090$; DSS HR = 0.761, 95% CI: 0.598–0.969, $p = 0.0266$).

We then stratified participants in the PLCO trial by clinical characteristic and examined associations of NPG for all variables. For OS and DSS of NSCLC, there were significant interactions between NPG and sex (OS: $P_{inter} = 0.0310$; DSS: $P_{inter} = 0.0414$), smoking status (OS: $P_{inter} = 0.0072$; DSS: $P_{inter} = 0.0307$), and chemotherapy (OS: $P_{inter} = 0.0160$; DSS: $P_{inter} = 0.0142$), but not for age, histology, tumor stage, radiotherapy, or surgery treatment (OS and DSS: $P_{inter} > 0.05$ for all). Having a higher NPG was only significantly associated with a better survival for patients aged >71 years, males, those who were former smokers, those with adenocarcinoma or other tumor histology, those with tumor stage of IIIB–IV, those without chemotherapy, those with and without radiotherapy, and those without surgery (Table S7).

Subsequently, we constructed KM curves for both SNPs individually as well as their combined effects measured by NPA and NPG (Figures 1 and 2). The log-rank test revealed significant differences in OS for NPA ($p = 0.0390$) and NPG in an additive model ($p = 0.0352$) (Figure 1C,D). When NPG was dichotomized, the log-rank test revealed significant differences for both OS and DSS ($p = 0.0097$ for OS and $p = 0.0238$ for DSS) (Figures 1E and 2E). These remained significant after adjustment for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, and surgery (Figures 1C–E and 2E). Additionally, DDO rs9384742 OS ($p = 0.0434$), NPA DSS ($p = 0.0073$), and NPG DSS ($p = 0.0033$) in additive models were significant after adjustment for covariates (Figures 1A and 2C,D). Furthermore, the association was significant ($p = 0.0498$), when *DDO* rs9384742 CT/TT was applied in a dominant model after adjustment for covariates (Figure SA).

To further assess the predictive ability of the two SNPs for OS and DSS, we constructed ROC curves and compared the time-dependent AUC curves. However, after adding the two SNPs to the model, the AUC was not significantly larger for OS or DSS of NSCLC at any time point (Figure S7).

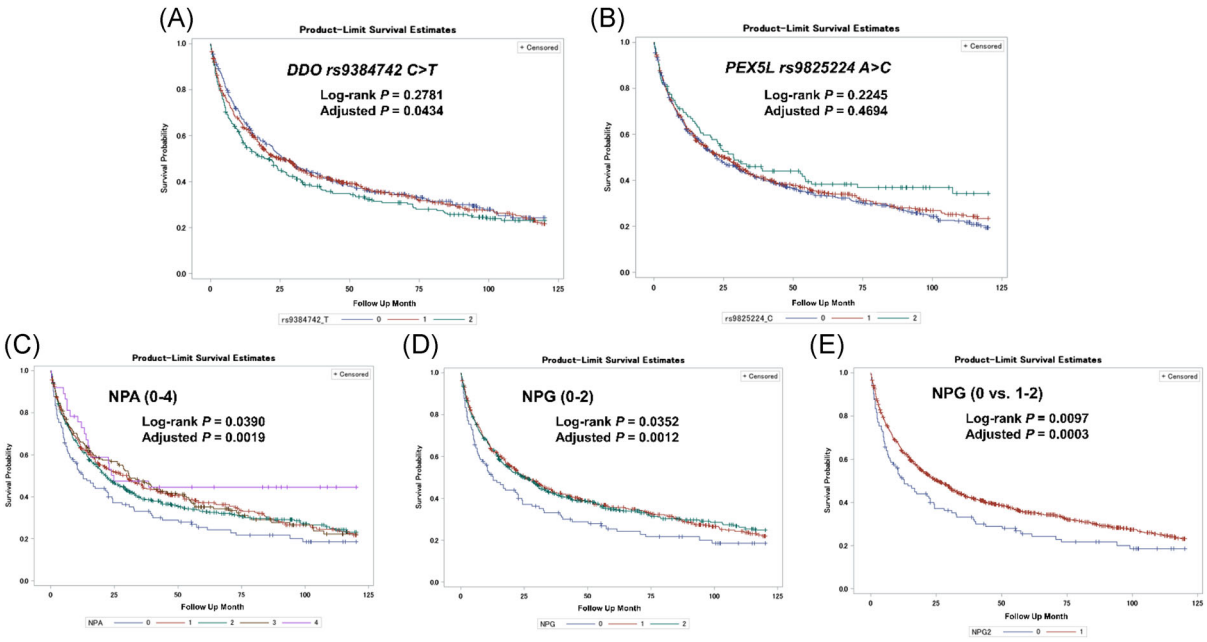

FIGURE 1 Kaplan–Meier survival curves for overall survival of non-small cell lung cancer *p*values from log-rank test and cox models with adjustment for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, and surgery are shown for (A) *DDO* rs9384742 in an additive model; (B) *PEX5L* rs9825224 in an additive model; (C) number of protective alleles (NPA): *DDO* rs9384742_C and *PEX5L* rs9825224_C; (D) two single-nucleotide polymorphisms combined with number of protective genotypes (NPG): *DDO* rs9384742 CT/CC and *PEX5L* rs9825224 CA/CC from 0 to 4; (E) 0 NPG versus 1–2 NPG. [Color figure can be viewed at wileyonlinelibrary.com]

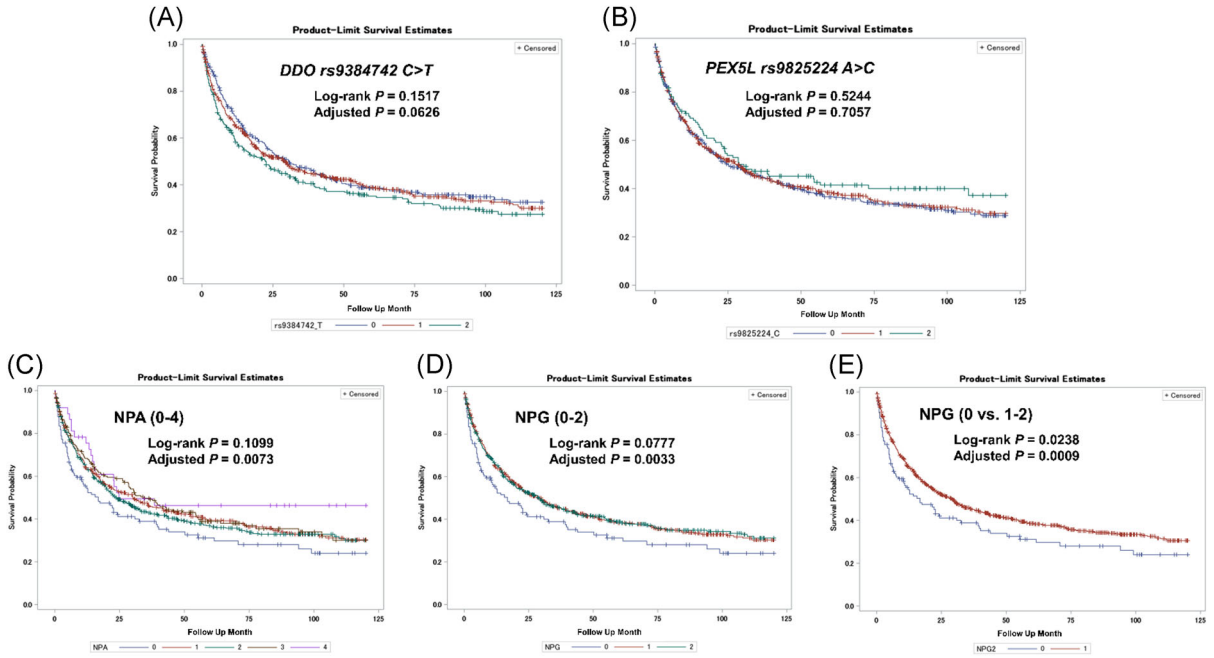

FIGURE 2 Kaplan-Meier survival curves for disease-specific survival of non-small cell lung cancer *p* values from log-rank test and cox models with adjustment for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, and surgery are shown for (A) *DDO* rs9384742 in an additive model; (B) *PEX5L* rs9825224 in an additive model; (C) number of protective alleles (NPA): *DDO* rs9384742_C and *PEX5L* rs9825224_C; (D) two single-nucleotide polymorphisms combined with number of protective genotypes (NPG): *DDO* rs9384742 CT/CC and *PEX5L* rs9825224 CA/CC from 0 to 4; (E) 0 NPG versus 1–2 NPG. [Color figure can be viewed at wileyonlinelibrary.com]

Using the eQTL analysis, we further investigated the effect of each SNP on mRNA expression levels in normal lung tissue and whole blood cells in an additive model using data from the GTEx project. Lower levels of *DDO* expression were associated with the rs9384742 T allele in lung tissue ($P = 1.1e{-}37$) (Figure 3A) and whole blood cells ($P = 4.4e{-}45$) (Figure 3B). The eQTL analysis using data from the 1000 Genomes Project showed the same significant association of lower *DDO* mRNA expression levels with the rs9384742 T allele in an additive model ($P = 5.13e{-}9$) (Figure SA), recessive model ($P = 2e{-}7$) (Figure SB), and dominant model ($P = 2.11e{-}5$) (Figure SC). *PEX5L* expression was not found to be significantly associated with rs9825224 in either of the data sets.

We then investigated whether expression levels of these genes were associated with OS of NSCLC using the Kaplan–Meier plotter. Higher levels of *PEX5L* expression were significantly associated with a worse survival (HR = 1.36, 95% CI: 1.15–1.61, log-rank $P = 2.4 \times 10^{-4}$) (Figure 3E), while expression of *DDO* was not significantly associated with OS (HR = 0.97, 95% CI: 0.85–1.1, log-rank $p = 0.6500$) (Figure 3C). We further investigated mRNA expression levels of the two genes in the paired tumor and adjacent normal lung tissue samples in LUAD and LUSC patients from the TCGA database. We found that *DDO* expression was decreased in LUSC and LUAD tumors ($p < 0.0001$ for both) (Figure SF–H). We repeated this analysis with more samples from the TCGA database through UALCAN and found that *DDO* expression level was significantly lower in both LUSC and LUAD tumors ($P = 1.62{-}e12$ for both) (Figure SD,E) and *PEX5L* expression was decreased in LUSC tumors ($p = 0.0040$) (Figure SE).

Finally, we investigated whether mutation rates of the two genes were significantly higher in tumors than in normal tissues, as they may affect survival more than the effect of a SNP. Using the cBioPortal for Cancer Genomics, we found that *PEX5L* had moderate mutation rates in some NSCLC data sets (12.5% in 16 patients in MSK 2015% and 8.57% in 35 patients in MSKCC 2015) and *DDO* had high mutation rates in some NSCLC data sets (35% in 20 patients in multi-institute 2017 and 12.6% in 16 patients in MSK 2015), which further suggest that *PEX5L* and *DDO* have important effects on survival of NSCLC (Figure S10).

4 | DISCUSSION

In the present study, we found that the *DDO* rs9384742 T allele was associated with a worse survival and that *PEX5L* rs9825224 C allele was associated with a better outcome of NSCLC in two independent cohorts of NSCLC patients. We also found that the *DDO* rs9384742 T variant was associated with a decreased mRNA expression level and that *DDO* rs9384742 is located in intronic regions of the gene associated with the enhancer, promoter, and DNAse hypersensitivity regions, providing further evidence that *DDO* rs9384742 may be involved in gene-regulatory roles. We further showed that expression level of *DDO* was decreased in both LUAD and LUSC tumor tissues compared with normal lung tissue, suggesting additional biological

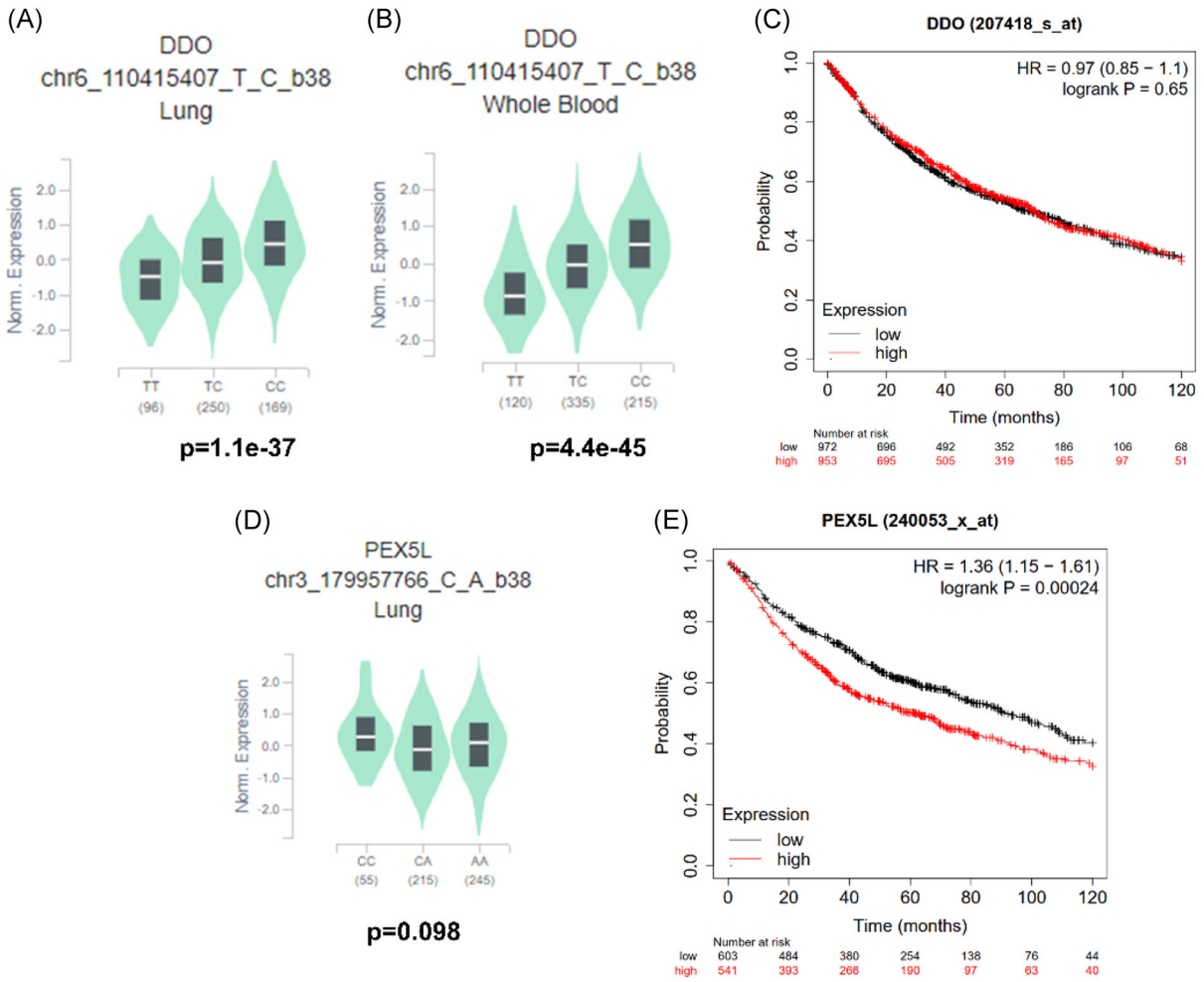

FIGURE 3 eQTL analysis of single-nucleotide polymorphisms (SNPs), SNP variants, and associations with messanger (mRNA) expression levels from GTEx in (A) *DDO* rs9384742 lung tissue; (B) *DDO* rs9384742 whole blood; (D) *PEX5L* rs9825224 lung tissue. Kaplan–Meier associations of dichotomized mRNA expression level and lung cancer overall survival with of (C) *DDO*; (E) *PEX5L*. [Color figure can be viewed at wileyonlinelibrary.com]

evidence for possible SNP functions. While *PEX5L* rs9825224 C is also located in an intronic region of the gene associated with the enhancer, promoter, and DNAse hypersensitivity regions, it was not associated with expression levels of *PEX5L* mRNA. However, high *PEX5L* expression levels were found to be significantly associated with a lower survival of NSCLC. Given that *PEX5L* is found to have moderately high mutation rates in NSCLC, it is likely that *PEX5L* is an important gene in the progression of NSCLC, which requires further mechanistic studies.

There are at least 36 known peroxins (PEX) proteins that form the import machinery of matrix proteins for peroxisomes.[14] Of these, PEX5 is shown to be essential for protein import into the peroxisomal matrix.[11] Peroxisome matrix proteins are synthesized outside the peroxisome and targeted by peroxisomal targeting sequence 1 or 2 (PTS1 and PTS2).[46] PEX5 recognizes PTS1 in the cytosol and docks at the peroxisomal docking/translocation machinery.[47,48] PEX5 has two isoforms, a long isoform PEX5L and a short isoform PEX5S, generated by alternative splicing.[49] PEX5L functions similarly to PEX5 with PTS1, but studies have additionally shown that PEX5L is necessary for PTS2-mediated import.[50,51]

To date, few studies have investigated these *PEX* genes in cancer. Because PEX5L is involved in peroxisome matrix protein import, the present study implies that this may be an important pathway for cancer progression. Because cancer cells are typically characterized by a high level of ROS, one hypothesis is that peroxisomes are hijacked to protect cancer cells from oxidative damage through enzymes such as catalase.[15] For example, a previous study showed that PEX5 was essential for the growth of hepatocellular carcinoma cells.[52] Another study showed that depletion of PEX5 increased cell sensitivity to oxidative stress and induced apoptosis.[53] Catalase is primarily imported by PTS1 and PEX5, and higher levels of ROS reduce PEX5L-mediated import of catalase into the peroxisome.[54-56] Therefore, peroxisomal control of cytosolic catalase levels through PEX5L may be a mechanism by which cancer cells cope with higher levels of oxidative stress throughout tumor growth. It is also possible that PEX5L influences cancer survival

through a modified immune response, as silencing PEX5 was shown to reduce macrophage phagocytic ability.[57] To date, few studies have investigated the role of peroxisomes in NSCLC, and none have examined the role of *PEX5L*.[13]

D-Aspartate oxidase (DDO) is a peroxisomal enzyme that selectively degrades D-Aspartate into oxaloacetate, hydrogen peroxide (H_2O_2), and ammonia, and overexpression of DDO has been shown to lead to cell death due to high levels of H_2O_2.[58] Thus, lower levels of DDO, leading to lower levels of ROS, may facilitate cancer growth. D-Aspartate is the D enantiomer of the more common amino acid L-Aspartate and has known some effects on nervous and endocrine systems through activating the *N*-methyl-D-aspartate receptor (NMDAR).[59] One study showed that NMDAR receptors were expressed by small-cell lung cancer (SCLC) lines and that NMDAR antagonist impaired SCLC cell growth in tumor xenografted mouse models.[60] Additional studies have shown that NMDAR antagonists reduce proliferation and migration of various tumor types in vitro and increase survival in vivo in mice with breast cancer, melanoma, and lung tumors, through the inhibition of MAP kinase, ERK, and AKT pathways.[61,62] While some studies have investigated D-Aspartate in lung cancer, there are no reported studies that have investigated the role of DDO in NSCLC.

DDO and PEX5L may work together to influence NSCLC progression, as DDO is located to the peroxisome through the PTS1 pathway (B). However, in the present study, it is unclear whether *PEX5L* rs9825224 and *DDO* rs9384742 affect survival jointly or independently. Thus, possible interactive mechanisms of these genes versus their independent contribution that influences NSCLC survival need to be further explored.

The role of peroxisomes in lung cancer is not well understood, and more mechanistic studies may reveal treatment targets in peroxisome-related pathways.[63] The present study demonstrates associations between two SNPs in two genes involved in peroxisome functions and survival of NSCLC. However, the present study has several limitations, including being conducted using a sample of only Caucasian patients, which may reduce generalizability to all populations. Additionally, even after correcting for multiple testing with BFDP, there is the possibility of false positives, because these multiple test corrections are relatively less stringent. Furthermore, although several clinical variables were available for us to analyze in the discovery PLCO data set, more detailed treatment and disease-progression information was not available in the validation HLCS data set. Finally, although our findings may provide some predictive biomarkers of NSCLC survival, the biological mechanisms behind how these SNPs may affect transcription regulation remain incompletely understood. Although additional and functional experiments should be performed, the present study suggests that the peroxisomal pathways may be a potential treatment target.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTION

Allan Chen, Carolyn Glass, Yufeng Wu, Sheng Luo, Edward F. Patz Jr, Li Su, Mulong Du, and Hongliang Liu participated in data collection and analysis. Allan Chen, Hongliang Liu, Sheng Luo, David C. Christiani, and Qingyi Wei participated in the conception and design of the study. All authors contributed to and edited the final version of the manuscript.

ACKNOWLEDGMENTS

The authors thank all the participants of the PLCO Cancer Screening Trial and the National Cancer Institute for providing the access to the data collected by the PLCO trial. However, 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 data sets (accession number phs000336.v1.p1 and phs000093.v2.p2). A list of contributing investigators and funding agencies for those studies can be found in the Supporting Information. This study was supported by the National Institute of Health [CA5U01CA209414]; the Duke Cancer Institute as part of the P30 Cancer Center Support Grant [NIH/NCI CA014236]; and the V Foundation for Cancer Research [D2017-19].

ORCID

Qingyi Wei http://orcid.org/0000-0002-3845-9445

REFERENCES

1. Bade BC, Dela Cruz CS. Lung cancer 2020: epidemiology, etiology, and prevention. *Clin Chest Med*. 2020;41(1):1-24.
2. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer*. 2014;14(8):535-546.
3. Nasim F, Sabath BF, Eapen GA. Lung cancer. *Med Clin North Am*. 2019;103(3):463-473.
4. Erichsen HC, Chanock SJ. SNPs in cancer research and treatment. *Br J Cancer*. 2004;90(4):747-751.
5. Bossé Y, Amos CI. A decade of GWAS results in lung cancer. *Cancer Epidemiol Biomarkers Prev*. 2018;27(4):363-379.
6. Gallagher MD, Chen-Plotkin AS. The post-GWAS era: from association to function. *Am J Hum Genet*. 2018;102(5):717-730.
7. Xue Y, Hou S, Ji H, Han X. Evolution from genetics to phenotype: reinterpretation of NSCLC plasticity, heterogeneity, and drug resistance. *Protein Cell*. 2017;8(3):178-190.
8. Marjoram P, Zubair A, Nuzhdin SV. Post-GWAS: where next? More samples, more SNPs or more biology? *Heredity*. 2014;112(1):79-88.
9. Tang D, Zhao YC, Qian D, et al. Novel genetic variants in HDAC2 and PPARGC1A of the CREB-binding protein pathway predict survival of non-small-cell lung cancer. *Mol Carcinog*. 2020;59(1):104-115.
10. Wu Y, Liu Z, Tang D, Liu H, Luo S, Stinchcombe TE, Glass C, Su L, Lin L, Christiani DC, Wang Q, Wei Q. Potentially functional variants of HBEGF and ITPR3 in GnRH signaling pathway genes predict survival of non-small cell lung cancer patients. *Transl Res*. 2021 Jul;233:92-103.
11. Okumoto K, Tamura S, Honsho M, Fujiki Y. Advances in experimental medicine and biology. In: Lizard G., ed. *Peroxisome: Metabolic Functions and Biogenesis, in Peroxisome Biology: Experimental Models,*

Peroxisomal Disorders and Neurological Diseases. Springer International Publishing; 2020:3-17.
12. Ghaffari P, Mardinoglu A, Nielsen J. Cancer metabolism: a modeling perspective. *Front Physiol*. 2015;6(382):1-9.
13. Kim J-A. Peroxisome metabolism in cancer. *Cells*. 2020;9(7):1692.
14. Islinger M, Voelkl A, Fahimi HD, Schrader M. The peroxisome: an update on mysteries 2.0. *Histochem Cell Biol*. 2018;150(5):443-471.
15. Dahabieh MS, Di Pietro E, Jangal M, et al. Peroxisomes and cancer: the role of a metabolic specialist in a disease of aberrant metabolism. *Biochim Biophysic Acta Rev Cancer*. 2018. 1870(1):103-121.
16. Svensson RU, Parker SJ, Eichner LJ, et al. Inhibition of acetyl-CoA carboxylase suppresses fatty acid synthesis and tumor growth of non-small-cell lung cancer in preclinical models. *Nat Med*. 2016; 22(10):1108-1119.
17. Zhang X, Yang H, Zhang J, Gao F, Dai L. HSD17B4, ACAA1, and PXMP4 in peroxisome pathway are down-regulated and have clinical significance in non-small cell lung cancer. *Front Genet*. 2020;11:273.
18. Prorok PC, Andriole GL, Bresalier RS, et al. Design of the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. *Controlled Clin Trials*. 2000;21(6), (suppl 1):273S-309S.
19. Oken MM, Marcus PM, Hu P, et al. Baseline chest radiograph for lung cancer detection in the randomized prostate, lung, colorectal and ovarian cancer screening trial. *J Natl Cancer Inst*. 2005;97(24): 1832-1839.
20. Mailman MD, Feolo M, Jin Y, et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*. 2007;39(10):1181-1186.
21. Tryka KA, Hao L, Sturcke A, et al. NCBI's database of genotypes and phenotypes: dbGaP. *Nucleic Acids Res*. 2014;42(D1):D975-D979.
22. Wang Y, Liu H, Ready NE, et al. Genetic variants in ABCG1 are associated with survival of nonsmall-cell lung cancer patients. *Int J Cancer*. 2016;138(11):2592-2601.
23. Zhai R, Yu X, Wei Y, Su L, Christiani DC. Smoking and smoking cessation in relation to the development of co-existing non-small cell lung cancer with chronic obstructive pulmonary disease. *Int J Cancer*. 2014;134(4):961-970.
24. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet*. 2009;10:387-406.
25. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010;34(8):816-834.
26. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34(3):267-273.
27. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005;102(43):15545-15550.
28. Coleman JRI, Euesden J, Patel H, Folarin AA, Newhouse S, Breen G. Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray. *Briefings Funct Genomics*. 2016;15(4):298-304.
29. Kreiner-Møller E, Medina-Gomez C, Uitterlinden AG, Rivadeneira F, Estrada K. Improving accuracy of rare variant imputation with a two-step imputation approach. *Eur J Hum Genet*. 2015;23(3):395-400.
30. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 2007; 23(10):1294-1296.
31. Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet*. 2007; 81(2):208-227.
32. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21(2): 263-265.
33. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010; 26(18):2336-2337.
34. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. 2012;40(Database issue):D930-D934.
35. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res*. 2009;37(Web Server issue):W600-W605.
36. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790-1797.
37. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC Med Res Methodol*. 2017;17(1):53.
38. GTEx Consortium. Human Genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348(6235):648-660.
39. Lappalainen T, Sammeth M, Friedländer MR, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013;501(7468):506-511.
40. Győrffy B, Surowiak P, Budczies J, Lánczky 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(12):e82241.
41. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014; 511(7511):543-550.
42. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012;489(7417):519-525.
43. Chandrashekar DS, Bashel B, Balasubramanya SAH, et al. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia*. 2017;19(8):649-658.
44. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401-404.
45. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1.
46. Girzalsky W, Saffian D, Erdmann R. Peroxisomal protein translocation. *Biochim Biophys Acta*. 2010;1803(6):724-731.
47. Berg JM, Gatto GJ, Geisbrecht BV, Gould SJ. Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. *Nat Struct Biol*. 2000;7(12):1091-1095.
48. Reguenga C, Oliveira MEM, Gouveia AMM, Sá-Miranda C, Azevedo JE. Characterization of the mammalian peroxisomal import machinery: Pex2p, Pex5p, Pex12p, and Pex14p are subunits of the same protein assembly. *J Biol Chem*. 2001;276(32):29935-29942.
49. Otera H, Harano T, Honsho M, et al. The mammalian peroxin Pex5pL, the longer isoform of the mobile peroxisome targeting signal (PTS) type 1 transporter, translocates the Pex7p·PTS2 protein complex into peroxisomes via its initial docking site, Pex14p*. *J Biol Chem*. 2000;275(28):21703-21714.
50. Kunze M, Malkani N, Maurer-Stroh S, Wiesinger C, Schmid JA, Berger J. Mechanistic insights into PTS2-mediated peroxisomal protein import: the co-receptor PEX5L drastically increases the interaction strength between the cargo protein and the receptor PEX7. The. *J Biol Chem*. 2015;290(8):4928-4940.
51. Zhang J, Tripathi DN, Jing J, et al. ATM functions at the peroxisome to induce pexophagy in response to ROS. *Nat Cell Biol*. 2015;17(10): 1259-1269.

52. Cai M, Sun X, Wang W, et al. Disruption of peroxisome function leads to metabolic stress, mTOR inhibition, and lethality in liver cancer cells. *Cancer Lett*. 2018;421:82-93.
53. Wang W, Xia ZJ, Farré JC, Subramani S. TRIM37, a novel E3 ligase for PEX5-mediated peroxisomal matrix protein import. *J Cell Biol*. 2017;216(9):2843-2858.
54. Okumoto K, El Shermely M, Natsui M, et al. The peroxisome counteracts oxidative stresses by suppressing catalase import via Pex14 phosphorylation. *eLife*. 2020;9:e55896.
55. Dahabieh MS, Ha Z, Di Pietro E, et al. Peroxisomes protect lymphoma cells from HDAC inhibitor-mediated apoptosis. *Cell Death Differ*. 2017;24(11):1912-1924.
56. Walton PA, Brees C, Lismont C, Apanasets O, Fransen M. The peroxisomal import receptor PEX5 functions as a stress sensor, retaining catalase in the cytosol in times of oxidative stress. *Biochim Biophys Acta Mol Cell Res*. 2017;1864(10):1833-1843.
57. Di Cara F, Sheshachalam A, Braverman NE, Rachubinski RA, Simmonds AJ. Peroxisome-mediated metabolism is required for immune response to microbial infection. *Immunity*. 2017;47(1):93-106. e7.
58. Katane M, Matsuda S, Saitoh Y, et al. Glyoxylate reductase/hydroxypyruvate reductase regulates the free d-aspartate level in mammalian cells. *J Cell Biochem*. 2021;122(11):1639-1652.
59. Pollegioni L, Molla G, Sacchi S, Murtas G. Human D-aspartate oxidase: a key player in D-aspartate metabolism. *Front Mol Biosci*. 2021;8:689719.
60. North WG, Gao G, Jensen A, Memoli VA, Du J. NMDA receptors are expressed by small-cell lung cancer and are potential targets for effective treatment. *Clin Pharmacol*. 2010;2:31-40.
61. Ribeiro MP, Custódio JB, Santos AE. Ionotropic glutamate receptor antagonists and cancer therapy: time to think out of the box? *Cancer Chemother Pharmacol*. 2017;79(2):219-225.
62. Rzeski W, Ikonomidou C, Turski L. Glutamate antagonists limit tumor growth. *Biochem Pharmacol*. 2002;64(8):1195-1200.
63. Díaz P, Sandoval-Bórquez A, Bravo-Sagua R, Quest A, Lavandero S. Perspectives on organelle interaction, protein dysregulation, and cancer disease. *Front Cell Dev Biol*. 2021;9:613336.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Chen AS, Liu H, Wu Y, et al. Genetic variants in *DDO and PEX5L* in peroxisome-related pathways predict non-small cell lung cancer survival. *Molecular Carcinogenesis*. 2022;1-10. doi:10.1002/mc.23400