life

MDPI

Article

Subtype-Specific Prognostic Associations of *CD300LF* Expression in Lung Adenocarcinoma and Lung Squamous Cell Carcinoma

Chaeyun Kim [1], Hyowon Hong [2,*] and Jae-Ho Lee [2,*]

[1] Medical Course, School of Medicine, Keimyung University, Daegu 42601, Republic of Korea; leminela@kmu.kr
[2] Department of Anatomy, School of Medicine, Keimyung University, Daegu 42601, Republic of Korea
* Correspondence: cnsl428@dsmc.or.kr (H.H.); anato82@dsmc.or.kr (J.-H.L.)

Abstract

CD300LF (CD300f) is an immunoregulatory receptor predominantly expressed on myeloid cells, but its prognostic significance in non-small cell lung cancer remains unclear. This study investigated whether the associations of *CD300LF* expression with survival and immune-related transcriptional profiles differ between lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). RNA-sequencing and curated survival data from The Cancer Genome Atlas (TCGA) were analyzed for 502 LUAD and 494 LUSC patients. *CD300LF* was modeled as a continuous variable in multivariable Cox regression, and a *CD300LF*-by-histology interaction was tested in a pooled model. Sensitivity analyses adjusted for driver mutations, myeloid markers, and ESTIMATE immune and stromal scores. Gene set enrichment analysis (GSEA) used phenotype permutation. Five Gene Expression Omnibus (GEO) LUAD cohorts ($n = 965$) were analyzed by random-effects meta-analysis with Hartung–Knapp adjustment. In LUAD, higher *CD300LF* expression was associated with longer overall survival after adjustment for clinical covariates (hazard ratio [HR] per standard deviation [SD], 0.855; 95% confidence interval [CI], 0.737–0.991), whereas no association was observed in LUSC (HR, 1.089; 95% CI, 0.945–1.255; interaction $p = 0.017$). In LUAD, the association was confined to the first 24 months after diagnosis. *CD300LF* correlated strongly with myeloid markers in both subtypes. Immune and inflammatory gene sets remained enriched under phenotype permutation, whereas proliferation-related gene sets were largely not significant. In GEO cohorts, the unadjusted pooled HR per SD was 0.825 (95% CI, 0.684–0.996), but the covariate-adjusted pooled estimate was not significant. *CD300LF* expression marks a myeloid-rich, immune-active tumor microenvironment and is associated with favorable early survival in LUAD but not in LUSC. External support is modest, and its prognostic contribution beyond overall immune infiltration remains to be established.

Keywords: *CD300LF*; lung adenocarcinoma; lung squamous cell carcinoma; tumor microenvironment; myeloid cells; overall survival; gene set enrichment analysis

Check for updates

Academic Editor: Jian-Hua Mao

Received: 28 October 2026
Revised: 19 November 2026
Accepted: 25 November 2026
Published: 26 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Lung cancer is the most frequently diagnosed cancer and the leading cause of cancer death worldwide, accounting for almost 2.5 million new cases (12.4% of all cancers) and an estimated 1.8 million deaths (18.7% of cancer deaths) in 2022 [1]. Lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), the two major histological subtypes of non-small cell lung cancer, differ substantially in their genomic alterations and molecular features [2,3]. A biomarker may therefore carry different prognostic meaning in each subtype, and subtype-specific markers that complement pathologic stage are needed.

The tumor microenvironment, particularly its immune composition and activation state, plays an important role in cancer progression and patient survival. A pan-cancer analysis of The Cancer Genome Atlas (TCGA) identified distinct immune subtypes, including an interferon-γ-dominant subtype, that were associated with clinical outcomes [4]. In lung cancer, tumor-associated macrophages exhibit heterogeneous and context-dependent prognostic associations. M1 macrophages and macrophages located within tumor islets have been associated with favorable survival, whereas M2 macrophages and stromal macrophages have been associated with poorer outcomes [5]. These findings suggest that the expression of molecules associated with myeloid cells may provide insights into the immune state of the tumor microenvironment and its relationship with prognosis.

CD300LF, also known as CD300f or CLM-1, is an immunoregulatory receptor encoded on chromosome 17q25 and expressed predominantly in myeloid cells, including monocytes and macrophages [6,7]. Its cytoplasmic domain contains both inhibitory motifs and phosphoinositide 3-kinase (PI3K)-binding sites, enabling context-dependent signaling that may promote or suppress cellular responses [6,7]. CD300f recognizes lipids, including phosphatidylserine and ceramide, and participates in the clearance of apoptotic cells through efferocytosis. Disruption of CD300f-mediated signaling has also been associated with dysregulated inflammatory and immune responses [6]. In addition, CD300f has attracted interest as a potential therapeutic target in acute myeloid leukemia with monocytic differentiation [7]. A pan-cancer analysis of CD300 family genes further reported that *CD300LF* expression was lower in LUAD and LUSC than in normal lung tissue, and that its prognostic associations differed in direction across cancer types [8]. Survival analyses of *CD300LF* mRNA expression in TCGA LUAD are also publicly available through the Human Protein Atlas pathology resource [9]. However, such single-cohort associations do not establish whether the prognostic relevance of *CD300LF* differs between histological subtypes, nor whether it reflects *CD300LF* itself or the overall abundance of immune cells in the tumor.

Despite these observations, whether the prognostic association of *CD300LF* differs between LUAD and LUSC has not been formally tested. Furthermore, because *CD300LF* is expressed mainly by myeloid cells, it is unclear whether its association with outcome is independent of myeloid and immune infiltration. In this study, we analyzed TCGA LUAD and LUSC cohorts to examine the associations of *CD300LF* expression with clinicopathological characteristics and overall survival, modeling *CD300LF* as a continuous variable and testing a *CD300LF*-by-histology interaction. We examined whether the association persisted after adjustment for driver mutations and for myeloid and immune infiltration, and characterized *CD300LF*-associated pathways using correlation analysis and gene set enrichment analysis (GSEA) with phenotype permutation. Finally, we evaluated the consistency of the prognostic association in LUAD across independent Gene Expression Omnibus (GEO) cohorts with covariate-adjusted models, and examined the cellular source of *CD300LF* in public single-cell transcriptomic data.

2. Materials and Methods

2.1. TCGA Data

Level 3 RNA-seq data (Illumina HiSeq RNASeqV2; Illumina, San Diego, CA, USA, gene-level RSEM normalized counts, log2 (x + 1)-transformed) for TCGA LUAD and LUSC from the Broad Institute GDAC Firehose (standard data run 28 January 2016) were obtained through the UCSC Xena TCGA hub [10]. Overall survival (OS) was taken from the TCGA Pan-Cancer Clinical Data Resource (TCGA-CDR) [11], clinical variables from the Firehose clinical matrix, and somatic mutation status of *EGFR*, *KRAS*, and *STK11* from the MC3 gene-level non-silent mutation calls [12]. Primary tumor samples were identified by TCGA barcode (sample type code 01), and one sample per patient was retained, yielding

515 LUAD and 502 LUSC primary tumors; all of these were used for correlation analysis and GSEA. Patients with available OS and follow-up longer than 0 days (LUAD $n = 502$; LUSC $n = 494$) formed the survival cohorts, and complete-case multivariable analyses included 471 LUAD (168 deaths) and 473 LUSC (207 deaths) patients (Table S1). Smoking status was classified as never (lifelong non-smoker) or ever smoker, and OS was converted from days to months (days/30.44).

2.2. Clinicopathological and Survival Analysis

Patients were divided into high- and low-*CD300LF*-expression groups according to the median expression value within each histological subtype for descriptive comparisons and Kaplan–Meier analysis. Categorical variables were compared using the chi-square test, or Fisher's exact test when any expected count was below 5, and age was compared using the Mann–Whitney U test; p values were adjusted for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) within each subtype. Overall survival was compared by Kaplan–Meier analysis with the log-rank test. In Cox proportional hazards regression, *CD300LF* was modeled primarily as a continuous variable standardized within each subtype (hazard ratio [HR] per 1 standard deviation [SD] increase), with the median split as a sensitivity analysis. Multivariable models included age (continuous, per 10 years), sex, smoking status, and pathologic stage (I–IV) and were fitted to complete cases. The proportional hazards assumption was evaluated using scaled Schoenfeld residuals; when it was violated, time-varying effects were estimated by splitting follow-up at 24 and 60 months. To test whether the association differed between subtypes, LUAD and LUSC were pooled in a Cox model stratified by histology that included a *CD300LF*-by-histology interaction term, evaluated by the likelihood ratio test. Sensitivity analyses additionally adjusted for *EGFR*, *KRAS*, and *STK11* mutation status (LUAD); residual tumor status; and markers of myeloid or immune infiltration, namely *TYROBP*, *CD68*, a myeloid score (the standardized mean of *TYROBP*, *CD68*, *CD163*, *CD14*, and *ITGAM*), and ESTIMATE immune and stromal scores [13] calculated with the tidyestimate implementation [14]. The added prognostic value of *CD300LF* beyond each of these markers was assessed by the likelihood ratio test and Harrell's concordance index (C-index).

2.3. Correlation Analysis

Correlations between *CD300LF* and 14 selected genes were assessed in primary tumors (LUAD $n = 515$; LUSC $n = 502$) using log2-transformed expression and Pearson's correlation coefficient, with Spearman's rank correlation as a sensitivity analysis. p values were adjusted for multiple testing across the 14 genes within each subtype using the Benjamini–Hochberg FDR. The genes represented myeloid and antigen-presentation markers (*CD68*, *TYROBP*, *HLA-DRA*), inflammatory and interferon signaling (*CXCL10*, *STAT1*, *TNF*, *IL6*), proliferation (*CDK1*, *CCNB1*, *CDKN2A*, *MYC*), mesenchymal state (*VIM*), immune checkpoint (*CD274*), and metabolic regulation (*PPARG*).

2.4. Gene Set Enrichment Analysis

GSEA was performed in primary tumor samples only (LUAD $n = 515$, LUSC $n = 502$; one sample per patient), using *CD300LF* expression as a continuous phenotype with Pearson correlation as the ranking metric [15]. Gene sets were taken from the MSigDB Hallmark collection (version 7.5.1) [16]. Enrichment scores were calculated with the weighted ($p = 1$) Kolmogorov–Smirnov-like statistic, and significance was assessed with 1000 phenotype (sample-label) permutations, which preserve the gene–gene correlation structure, using a gene set size of 15–500. Gene sets with an FDR $q < 0.05$ were considered significant.

2.5. External Validation

GEO cohorts were eligible if they included patients with LUAD, provided overall survival data, and were profiled on a platform containing a probe targeting *CD300LF* (NM_139018). Five cohorts met these criteria: GSE72094 ($n = 398$) [17], GSE31210 ($n = 204$, after excluding patients flagged by the submitters for incomplete resection or adjuvant therapy), GSE36471 ($n = 115$), GSE13213 ($n = 117$), and GSE42127 ($n = 131$, adenocarcinoma subset with follow-up longer than 0 days). GSE68465 and GSE14814 were excluded because their platform (Affymetrix HG-U133A, GPL96; Affymetrix, Santa Clara, CA, USA) lacks a *CD300LF* probe according to the GEO platform annotation [18]. Probes were mapped to gene symbols using the platform annotation files; expression values on a linear scale (GSE31210) were log2-transformed. When more than one probe targeted *CD300LF* (GSE72094, GSE13213, and GSE36471), the per-sample mean of the probes was used, and use of the highest-expressed probe was examined as a sensitivity analysis. Survival times in days were converted to months. In each cohort, *CD300LF* expression was standardized to a z-score, and Cox regression estimated the HR per 1 SD increase, both unadjusted and adjusted for age, sex, and pathologic stage. Additional cohort-specific models adjusted for *KRAS*, *EGFR*, *STK11*, and *TP53* mutation status (GSE72094 and GSE36471), *KRAS* and *EGFR* mutation status (GSE31210 and GSE13213), tumor cell content (GSE36471), and adjuvant chemotherapy (GSE42127). Cohort estimates were pooled using a DerSimonian–Laird random-effects model [19], with and without the Hartung–Knapp–Sidik–Jonkman adjustment [20], and heterogeneity was assessed with Cochran's Q and I^2. Robustness was examined by leave-one-out analysis.

2.6. Single-Cell Expression Analysis

The cellular source of *CD300LF* was examined in two public single-cell RNA-sequencing datasets processed and annotated by TISCH2 [21]: GSE131907, comprising 58 samples from 44 patients with LUAD, including primary tumors, normal lung, lymph nodes, pleural effusions, and brain metastases [22], and GSE148071, comprising tumors from 42 patients with advanced NSCLC [23]. Mean *CD300LF* expression in each annotated cell type was obtained from TISCH2.

2.7. Statistical Analysis

TCGA clinicopathological comparisons, Cox regression, ESTIMATE scoring, GEO survival analyses, and meta-analyses were performed in R version 4.3.3 (R Foundation for Statistical Computing, Vienna, Austria) using the survival package (version 3.5-8). Correlation analysis and GSEA were performed in Python version 3.13 (Python Software Foundation, Wilmington, DE, USA) with pandas, NumPy, SciPy, lifelines, and matplotlib; enrichment scores from the GSEA implementation were verified against gseapy (version 1.3.2). Two-sided $p < 0.05$ was considered significant.

3. Results

3.1. Clinicopathological Features

In LUAD, high *CD300LF* expression was more frequent in patients older than 65 years (57.3% vs. 46.3%; FDR q = 0.043), never-smokers (18.9% vs. 10.6%; q = 0.043), and patients with T1 tumors (39.8% vs. 27.6%; q = 0.043) (Table 1). The association with pathologic stage did not reach significance ($p = 0.054$), and no associations were observed with N stage, M stage, or sex. In LUSC, high *CD300LF* expression was more frequent in women (31.6% vs. 20.2%; q = 0.032), whereas no significant associations were observed with age, smoking status, pathologic stage, or TNM classification.

Table 1. Clinicopathological characteristics according to *CD300LF* expression in TCGA LUAD and LUSC.

Variable	LUAD High (n = 251)	LUAD Low (n = 251)	p	q	LUSC High (n = 247)	LUSC Low (n = 247)	p	q
Age (years) (n = 492; 489)								
≤65	105 (42.7)	132 (53.7)	0.015	0.043	99 (40.6)	90 (36.7)	0.383	0.537
>65	141 (57.3)	114 (46.3)			145 (59.4)	155 (63.3)		
Sex (n = 502; 494)								
Female	144 (57.4)	127 (50.6)	0.128	0.146	78 (31.6)	50 (20.2)	0.004	0.032
Male	107 (42.6)	124 (49.4)			169 (68.4)	197 (79.8)		
Smoking (n = 488; 482)								
Never	46 (18.9)	26 (10.6)	0.010	0.043	12 (5.0)	6 (2.5)	0.144	0.537
Ever	197 (81.1)	219 (89.4)			228 (95.0)	236 (97.5)		
Pathologic stage (n = 494; 490)								
I	148 (60.2)	121 (48.8)	0.054	0.103	129 (52.9)	113 (45.9)	0.278	0.537
II	55 (22.4)	65 (26.2)			69 (28.3)	89 (36.2)		
III	31 (12.6)	49 (19.8)			43 (17.6)	40 (16.3)		
IV	12 (4.9)	13 (5.2)			3 (1.2)	4 (1.6)		
T stage (n = 499; 494)								
T1	99 (39.8)	69 (27.6)	0.016	0.043	64 (25.9)	50 (20.2)	0.403	0.537
T2	126 (50.6)	142 (56.8)			137 (55.5)	150 (60.7)		
T3	16 (6.4)	29 (11.6)			33 (13.4)	37 (15.0)		
T4	8 (3.2)	10 (4.0)			13 (5.3)	10 (4.0)		
N stage (n = 491; 488)								
N0	169 (69.3)	156 (63.2)	0.101	0.135	160 (66.1)	156 (63.4)	0.386	0.537
N1	48 (19.7)	47 (19.0)			57 (23.6)	70 (28.5)		
N2–N3	27 (11.1)	44 (17.8)			25 (10.3)	20 (8.1)		
M stage (n = 358; 413)								
M0	167 (93.8)	167 (92.8)	0.693	0.693	208 (98.6)	198 (98.0)	0.719	0.757
M1	11 (6.2)	13 (7.2)			3 (1.4)	4 (2.0)		

Values are n (%). High and low groups were defined by the median *CD300LF* expression within each subtype. The number of patients with available data is given for each variable; MX and NX were treated as missing. p values by chi-square test, or Fisher's exact test when any expected count was <5; q, Benjamini–Hochberg FDR within each subtype.

3.2. Prognostic Value in TCGA

In LUAD, patients with high *CD300LF* expression had longer overall survival than those with low expression (log-rank $p = 0.003$; Figure 1A). In multivariable Cox analysis adjusted for age, sex, smoking status, and pathologic stage, higher *CD300LF* expression was associated with lower mortality (HR per SD, 0.855; 95% CI, 0.737–0.991; $p = 0.037$; 471 patients, 168 deaths; Table 2); with the median split, the HR for low versus high expression was 1.381 (95% CI, 1.009–1.890; $p = 0.044$). In LUSC, *CD300LF* expression was not associated with survival in Kaplan–Meier (log-rank $p = 0.816$; Figure 1B) or multivariable analysis (HR per SD, 1.089; 95% CI, 0.945–1.255; $p = 0.237$; 473 patients, 207 deaths). In the pooled model stratified by histology, the *CD300LF*-by-histology interaction was significant (interaction HR, 1.275; 95% CI, 1.045–1.555; likelihood ratio $p = 0.017$), indicating that the association between *CD300LF* and survival differed between LUAD and LUSC.

In LUAD, the proportional hazards assumption was violated for *CD300LF* (Schoenfeld residual test, $p = 0.013$), consistent with the convergence and crossing of the Kaplan–Meier curves after approximately 60 months. In time-split analysis, higher *CD300LF* expression was associated with lower mortality during the first 24 months after diagnosis (HR per SD, 0.736; 95% CI, 0.611–0.887; $p = 0.001$; 100 deaths), but not between 24 and 60 months (HR, 1.040; 95% CI, 0.815–1.328) or after 60 months (HR, 1.025; 95% CI, 0.595–1.765; 7 deaths) (Table S3). No violation was observed in LUSC ($p = 0.569$).

In LUAD, adjustment for *EGFR*, *KRAS*, and *STK11* mutation status left the estimate essentially unchanged (HR per SD, 0.860; 95% CI, 0.737–1.005; $p = 0.058$; 466 patients), and adjustment for residual tumor status, available for 332 patients, gave a similar estimate (HR, 0.869; 95% CI, 0.736–1.026; $p = 0.097$) (Table S2). *CD300LF* was strongly correlated with *TYROBP* (r = 0.83), *CD68* (r = 0.76), the myeloid score (r = 0.81), and the ESTIMATE

immune score (r = 0.79). Adjustment for *TYROBP*, *CD68*, or the myeloid score did not attenuate the association (HRs per SD, 0.684, 0.712, and 0.671, respectively; all $p < 0.01$), and *CD300LF* improved model fit beyond each marker (likelihood ratio $p = 0.005$–0.009; C-index 0.679 to 0.694 for the myeloid score). After adjustment for the ESTIMATE immune score, the point estimate was unchanged (HR, 0.858) but the confidence interval widened to include 1 (95% CI, 0.667–1.103; $p = 0.23$), and *CD300LF* did not significantly improve model fit (likelihood ratio $p = 0.23$). In LUSC, none of these adjustments revealed an association (Table S2).

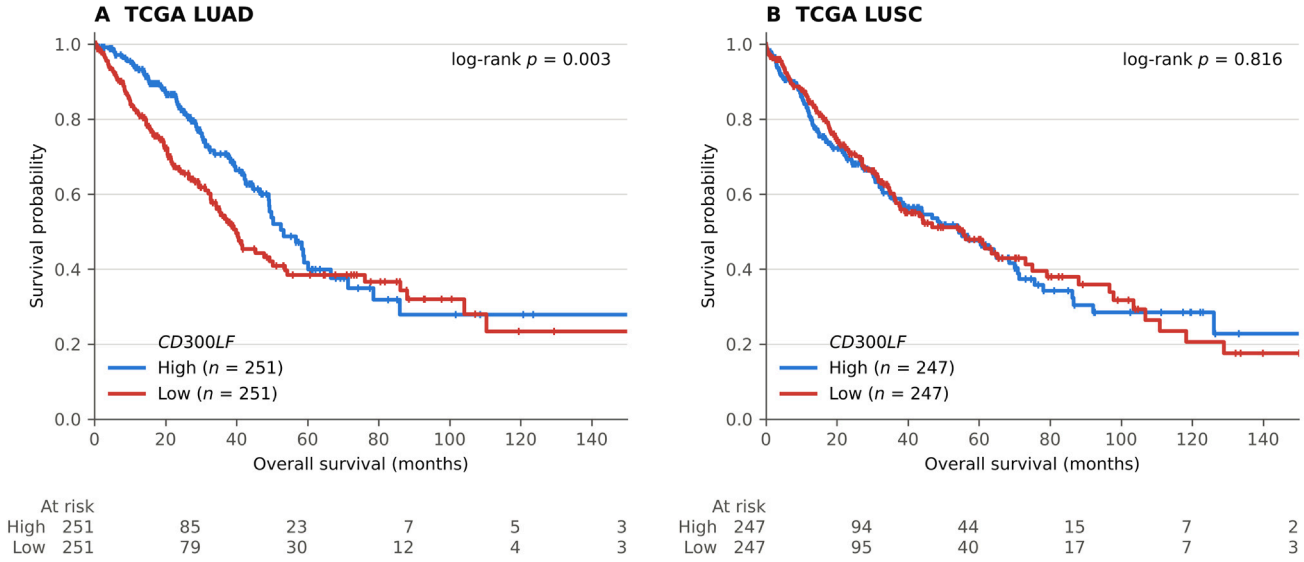

Figure 1. Kaplan–Meier curves for overall survival according to *CD300LF* expression (median split) in TCGA LUAD (**A**) and LUSC (**B**). *p* values by log-rank test.

Table 2. Multivariable Cox regression analysis of overall survival in TCGA LUAD and LUSC, with *CD300LF* modeled as a continuous variable.

Variable	Category	LUAD HR (95% CI)	*p*	LUSC HR (95% CI)	*p*
CD300LF (continuous)	per 1 SD increase	0.855 (0.737–0.991)	0.037	1.089 (0.945–1.255)	0.237
Age	per 10 years	1.122 (0.957–1.314)	0.156	1.262 (1.062–1.499)	0.008
Sex	Male vs. female	0.969 (0.701–1.341)	0.850	1.247 (0.898–1.733)	0.188
Smoking status	Ever vs. never	0.938 (0.604–1.456)	0.775	0.484 (0.209–1.119)	0.089
Pathologic stage	I	Reference		Reference	
	II	2.344 (1.598–3.438)	<0.001	1.147 (0.826–1.593)	0.413
	III	3.071 (2.063–4.572)	<0.001	1.662 (1.159–2.383)	0.006
	IV	3.539 (2.013–6.222)	<0.001	3.114 (1.253–7.740)	0.014
CD300LF (median split)	Low vs. high	1.381 (1.009–1.890)	0.044	0.901 (0.682–1.191)	0.463

Complete-case analysis: LUAD, 471 patients and 168 deaths; LUSC, 473 patients and 207 deaths. A separate model with *CD300LF* dichotomized at the median and the same covariates (covariate estimates in Table S2). In the pooled model stratified by histology, the *CD300LF*-by-histology interaction HR was 1.275 (95% CI, 1.045–1.555; likelihood ratio $p = 0.017$). HR, hazard ratio; CI, confidence interval; SD, standard deviation.

3.3. Co-Expression with Immune and Proliferation Genes

CD300LF showed strong positive correlations with myeloid and antigen-presentation markers in both subtypes, most strongly with *TYROBP* (LUAD r = 0.832; LUSC r = 0.923), *CD68* (r = 0.765; r = 0.759), and *HLA-DRA* (r = 0.768; r = 0.891) (all FDR q < 0.001; Table 3). Positive correlations were also observed with *VIM* (r = 0.458; r = 0.606), *CD274* (r = 0.463; r = 0.363), *CXCL10* (r = 0.395; r = 0.588), and *STAT1* (r = 0.273; r = 0.472). Correlations with proliferation genes were weak: in LUAD, *CCNB1* (r = −0.215) and *CDK1* (r = −0.144)

were weakly negatively correlated, and *MYC* was not correlated (r = −0.001), whereas in LUSC, *MYC* was weakly negatively correlated (r = −0.175) and *CDK1* and *CCNB1* showed negligible correlations (r = −0.085 and −0.092). Spearman correlations gave similar results (Table S4).

Table 3. Correlations between *CD300LF* and selected genes in TCGA LUAD (n = 515) and LUSC (n = 502) primary tumors.

Category	Gene	LUAD r	q	LUSC r	q
Myeloid/antigen presentation	*TYROBP*	0.832	<0.001	0.923	<0.001
	CD68	0.765	<0.001	0.759	<0.001
	HLA-DRA	0.768	<0.001	0.891	<0.001
Inflammation/interferon	*CXCL10*	0.395	<0.001	0.588	<0.001
	STAT1	0.273	<0.001	0.472	<0.001
	TNF	0.431	<0.001	0.337	<0.001
	IL6	0.147	0.001	0.203	<0.001
Proliferation	*CDK1*	−0.144	0.001	−0.085	0.063
	CCNB1	−0.215	<0.001	−0.092	0.045
	CDKN2A	0.037	0.438	0.070	0.117
	MYC	−0.001	0.977	−0.175	<0.001
Mesenchymal	*VIM*	0.458	<0.001	0.606	<0.001
Immune checkpoint	*CD274*	0.463	<0.001	0.363	<0.001
Metabolic	*PPARG*	0.431	<0.001	0.112	0.015

r, Pearson correlation coefficient on log2-transformed expression; q, Benjamini–Hochberg FDR across the 14 genes within each subtype.

3.4. Pathways Associated with CD300LF Expression

With phenotype permutation, 14 Hallmark gene sets were positively enriched in each subtype at FDR q < 0.05 (Figure 2; Tables 4 and S5). Eleven were shared between subtypes, including allograft rejection, complement, IL2–STAT5 signaling, inflammatory response, IL6–JAK–STAT3 signaling, interferon-γ response, and interferon-α response (LUAD q ≤ 0.017; LUSC q ≤ 0.004). In contrast, gene sets inversely associated with *CD300LF* were largely not significant: G2M checkpoint (LUAD q = 0.095; LUSC q = 0.128), E2F targets (q = 0.085; q = 0.162), and MYC targets V2 (q = 0.074; q = 0.209); only MYC targets V1 reached significance in LUAD (q = 0.042).

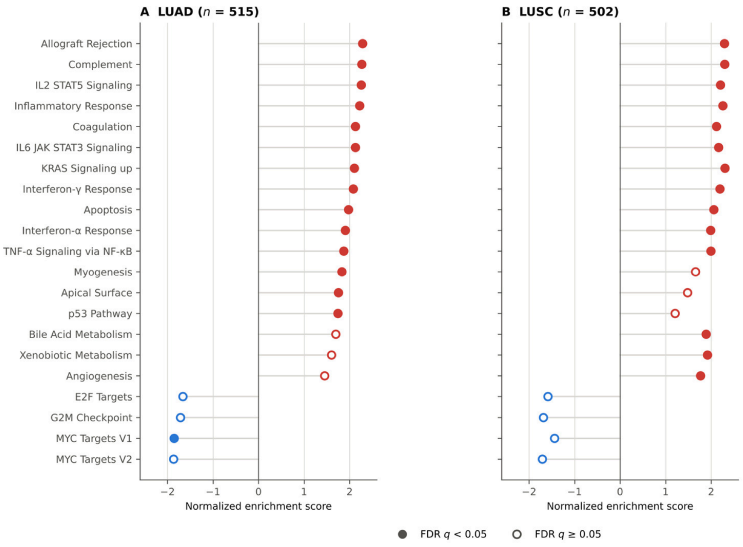

Figure 2. Hallmark gene sets correlated with *CD300LF* expression in TCGA LUAD (**A**) (n = 515) and LUSC (**B**) (n = 502 primary tumors) (continuous phenotype, Pearson ranking, 1000 phenotype

permutations). Gene sets with FDR q < 0.05 in either subtype and the proliferation-related gene sets are shown. Filled circles, FDR q < 0.05; open circles, FDR q ≥ 0.05. Red, positive enrichment; blue, negative enrichment. NES, normalized enrichment score.

Table 4. Hallmark gene sets associated with *CD300LF* expression in TCGA LUAD and LUSC primary tumors (phenotype permutation).

Hallmark Gene Set	LUAD NES	FDR q	LUSC NES	FDR q
Positively associated with *CD300LF*				
Allograft rejection	2.28	0.001	2.29	<0.001
Interferon-γ response	2.08	<0.001	2.19	<0.001
Inflammatory response	2.22	<0.001	2.26	<0.001
IL6–JAK–STAT3 signaling	2.12	<0.001	2.16	<0.001
Complement	2.27	<0.001	2.30	<0.001
Interferon-α response	1.90	0.012	1.99	0.004
KRAS signaling up	2.10	<0.001	2.30	<0.001
IL2–STAT5 signaling	2.25	<0.001	2.21	<0.001
Coagulation	2.13	<0.001	2.12	<0.001
TNF-α signaling via NF-κB	1.87	0.017	1.99	0.004
Apoptosis	1.98	0.005	2.06	0.001
Epithelial–mesenchymal transition	1.50	0.139	1.73	0.050
Negatively associated with *CD300LF*				
G2M checkpoint	−1.72	0.095	−1.68	0.128
E2F targets	−1.66	0.085	−1.59	0.162
MYC targets V1	−1.86	0.042	−1.44	0.251
MYC targets V2	−1.87	0.074	−1.71	0.209
Unfolded protein response	−1.67	0.106	−0.84	0.647
Mitotic spindle	−1.13	0.578	−1.16	0.582
Glycolysis	−1.13	0.515	−0.85	0.708
mTORC1 signaling	−1.18	0.590	−0.86	0.789

NES, normalized enrichment score; FDR, false discovery rate (1000 phenotype permutations). Positive NES indicates enrichment among genes positively correlated with *CD300LF*. Gene sets reported in the original submission are shown; all 50 Hallmark gene sets are listed in Table S5.

Epithelial–mesenchymal transition was not significant in LUAD (q = 0.139) and did not reach significance in LUSC (q = 0.050). Notably, enrichment of the KRAS signaling up gene set should be interpreted cautiously, because this gene set contains immune- and inflammation-related genes; its enrichment therefore indicates coordinated expression of genes within the set rather than direct evidence of *KRAS* pathway activation.

3.5. External Validation in LUAD

Across five independent GEO cohorts comprising 965 patients (299 deaths; Table S7), higher *CD300LF* expression was associated with lower mortality in four cohorts, with HRs per SD ranging from 0.732 to 0.917; GSE31210 showed a non-significant association in the opposite direction (HR, 1.085; 95% CI, 0.761–1.548) (Figure 3). GSE72094 (HR, 0.732; 95% CI, 0.617–0.867; $p < 0.001$) and GSE42127 (HR, 0.733; 95% CI, 0.558–0.963; $p = 0.026$) showed individually significant associations. The DerSimonian–Laird pooled HR was 0.825 (95% CI, 0.723–0.942; $p = 0.004$; $I^2 = 28.1\%$; Cochran's Q $p = 0.234$), and the association remained significant with the Hartung–Knapp adjustment (95% CI, 0.684–0.996; $p = 0.047$). After adjustment for age, sex, and pathologic stage, the pooled HR was attenuated to 0.871 and was not significant (DerSimonian–Laird 95% CI, 0.756–1.004, $p = 0.057$; Hartung–Knapp 95% CI, 0.712–1.067, $p = 0.131$). In GSE72094, the association persisted after adjustment for pathologic stage and *KRAS*, *EGFR*, *STK11*, and *TP53* mutation status (HR, 0.751; 95% CI, 0.626–0.902; $p = 0.002$) (Table S8). In leave-one-out analyses, the pooled HR remained below 1.0 in all iterations (0.789–0.872), but statistical significance was lost when GSE72094

was excluded (DerSimonian–Laird $p = 0.060$) and, with the Hartung–Knapp adjustment, in all iterations except that excluding GSE31210 (Figure 4).

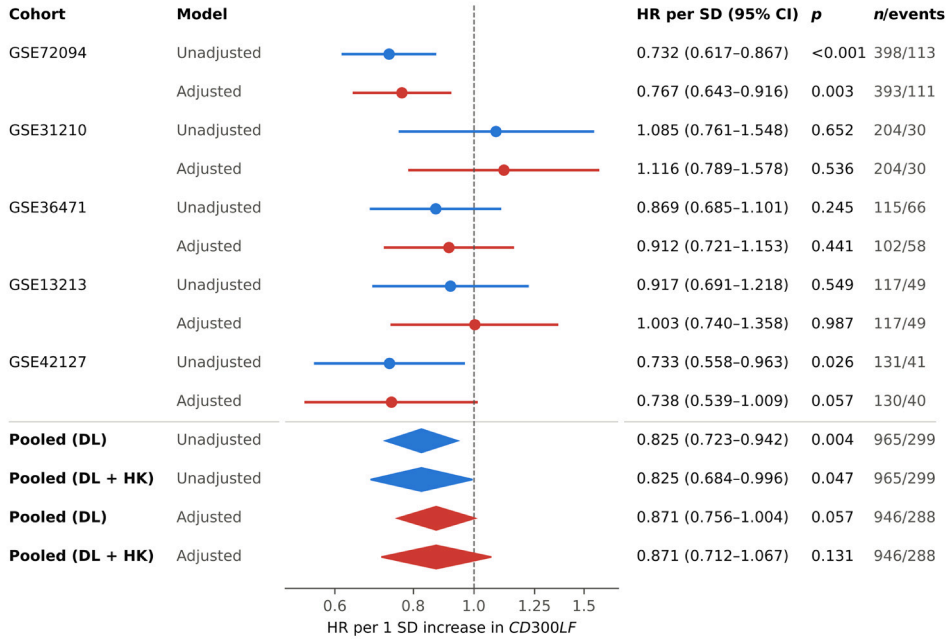

Figure 3. Forest plot of HRs for overall survival per 1 SD increase in *CD300LF* expression in five GEO LUAD cohorts, unadjusted and adjusted for age, sex, and pathologic stage, with pooled estimates from the DerSimonian–Laird (DL) random-effects model with and without the Hartung–Knapp (HK) adjustment. Circles and horizontal lines indicate cohort-level HRs and 95% CIs, and diamonds indicate pooled estimates; blue, unadjusted; red, adjusted for age, sex, and pathologic stage. The dashed vertical line indicates HR = 1. CI, confidence interval.

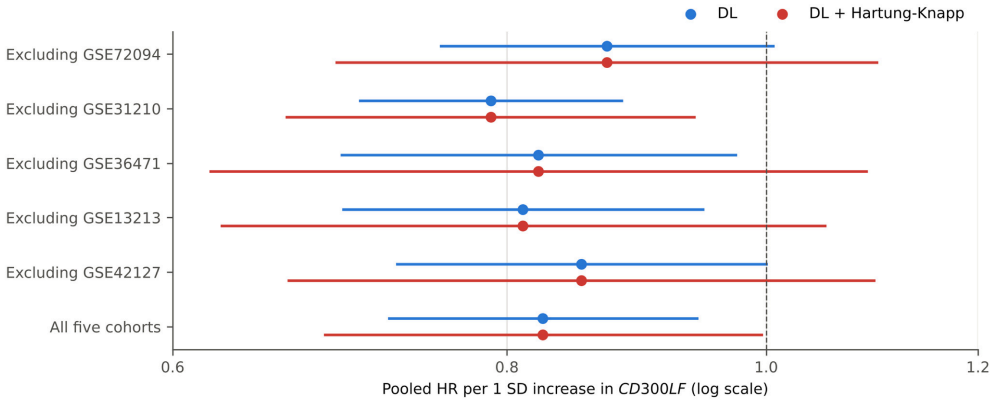

Figure 4. Leave-one-out sensitivity analysis of the pooled unadjusted HR using the DerSimonian–Laird model with and without the Hartung–Knapp adjustment. Each row shows the pooled estimate after excluding the indicated cohort. Circles and horizontal lines indicate pooled HRs and 95% CIs; blue, DerSimonian–Laird; red, DerSimonian–Laird with the Hartung–Knapp adjustment. The dashed vertical line indicates HR = 1.

3.6. Cellular Source of CD300LF

In public single-cell data, *CD300LF* expression was largely restricted to myeloid cells (Figure S1). In GSE131907, mean expression was highest in M2-like macrophages (0.46), conventional type 2 dendritic cells (0.26), monocytes (0.17), plasmacytoid dendritic cells (0.11), and mast cells (0.07), whereas epithelial cells (0.004), T cells, B cells, fibroblasts, and

endothelial cells showed minimal expression (≤0.014). In GSE148071, expression in the monocyte/macrophage population (0.14) far exceeded that in malignant cells (0.0008).

4. Discussion

In this study, higher *CD300LF* expression was associated with longer overall survival in LUAD but not in LUSC, and a formal interaction test supported a difference between the two subtypes. In LUAD, the association was concentrated within the first 2 years after diagnosis. In both subtypes, *CD300LF* expression was strongly associated with a myeloid-enriched, interferon- and inflammation-related transcriptional profile, whereas its inverse relationship with proliferation-related programs did not withstand phenotype permutation. The direction of the prognostic association was consistent with lower mortality in four of five independent GEO cohorts, but the pooled evidence was modest and was attenuated after adjustment for clinical covariates.

The strong correlations of *CD300LF* with *TYROBP*, *CD68*, and *HLA-DRA* indicate that, in bulk tumor tissue, its expression is closely linked to myeloid-cell abundance and antigen-presentation-related activity, consistent with its known expression pattern [6,7]. Public single-cell data confirmed that *CD300LF* is expressed almost exclusively by myeloid cells, particularly macrophages and conventional dendritic cells, with negligible expression in malignant epithelial cells; bulk *CD300LF* expression therefore reflects the tumor microenvironment rather than tumor cells. The co-enrichment of interferon-γ, allograft rejection, and other inflammatory and immune-related gene sets suggests that *CD300LF* expression is associated with an immunologically active tumor microenvironment. This pattern resembles the interferon-γ-dominant immune states associated with favorable outcomes in pan-cancer analyses [4] and the favorable survival associations reported for M1-like and intratumoral macrophages in lung cancer [5]. In LUAD, adjustment for *TYROBP*, *CD68*, or a composite myeloid score did not attenuate the association, suggesting that *CD300LF* is not simply a surrogate for the number of myeloid cells and may instead mark a particular myeloid state. However, after adjustment for the ESTIMATE immune score, the confidence interval widened to include 1, reflecting the strong collinearity between *CD300LF* and overall immune infiltration. We therefore regard *CD300LF* as a marker of an immune-active myeloid microenvironment rather than an established independent prognostic determinant. Survival analyses of *CD300LF* expression in TCGA LUAD are also publicly available through the Human Protein Atlas [9]; the present study extends such single-cohort observations by formally testing the difference between subtypes, adjusting for driver mutations and immune infiltration, and examining external cohorts. The prognostic meaning of *CD300LF* is context-dependent: Xu et al. reported that high *CD300LF* expression was associated with poor survival in acute myeloid leukemia, lower-grade glioma, and uveal melanoma, and with better survival only in skin cutaneous melanoma [8].

In the original analysis based on gene-set permutation, cell-cycle and *MYC* target gene sets appeared negatively enriched in *CD300LF*-high tumors. With phenotype permutation, which preserves the gene–gene correlation structure, these gene sets were no longer significant except for MYC targets V1 in LUAD, and correlations with individual proliferation genes were weak. Because lower proliferation-related signals in bulk tumors can also arise from dilution by non-malignant cells, these data do not support reduced proliferation of malignant cells in *CD300LF*-high tumors. The association of high *CD300LF* expression with T1 tumors raises the possibility that part of its prognostic association reflects tumor extent; however, the association persisted after adjustment for pathologic stage. The external cohort GSE72094 is relevant in this context because it is annotated for *KRAS*, *EGFR*, *STK11*, and *TP53* mutations [17], and the association persisted after adjustment for these mutations and stage in that cohort. Nevertheless, the attenuation of the pooled estimate after covariate

adjustment and the loss of significance in several leave-one-out analyses indicate that the external evidence should be interpreted cautiously.

The time-dependent pattern in LUAD, with an association confined to the first 2 years after diagnosis, explains the crossing of the Kaplan–Meier curves and suggests that the *CD300LF*-associated immune context may be related to early events, such as early recurrence or death, rather than to long-term outcome. Because few deaths occurred after 60 months, the late estimates are imprecise, and this pattern requires confirmation in independent cohorts.

The lack of a prognostic association in LUSC, despite an even stronger correlation with myeloid-associated markers (*TYROBP*, r = 0.92), together with the absence of any added value beyond myeloid markers, suggests that in LUSC *CD300LF* expression behaves largely as a marker of myeloid abundance. Differences in the immune landscape between LUAD and LUSC [24] may modify the relationship between myeloid-associated expression patterns and clinical outcomes. Although *CD300LF* was more strongly correlated with *VIM* in LUSC, enrichment of the epithelial–mesenchymal transition gene set did not reach significance under phenotype permutation, and the available data do not explain the absence of a survival association in LUSC. This question requires cell-type-resolved and functional approaches.

Mechanistically, CD300f recognizes phosphatidylserine and ceramide and participates in the regulation of efferocytosis, with activating or inhibitory effects depending on cellular context [6]. These functions provide a biological rationale for investigating CD300f in relation to myeloid-cell activity and the tumor microenvironment [25]. However, whether CD300f actively shapes antitumor immunity in lung cancer or primarily reflects the abundance or state of particular myeloid-cell populations cannot be determined from transcriptomic data.

This study has several limitations. First, the survival analyses were retrospective and based on bulk mRNA expression. The single-cell analysis used cell-type-averaged expression pooled across patients and, in GSE131907, across tissue sources, and therefore could not address patient-level variation in *CD300LF* expression per myeloid cell. Second, because *CD300LF* is strongly collinear with overall immune infiltration, its prognostic contribution beyond the immune content of the tumor could not be established. Third, external validation was limited: only two cohorts showed individually significant associations, the covariate-adjusted pooled estimate was not significant, and the pooled association depended substantially on GSE72094. Cohorts profiled on platforms lacking a *CD300LF* probe could not be included, and differences in measurement platforms may affect cross-cohort comparability. Thus, the external analyses support a generally consistent direction of association but do not establish a robust, cohort-independent prognostic effect. Fourth, the association in LUAD was time-dependent, and late follow-up was sparse. Fifth, information on residual tumor and adjuvant treatment was incomplete in TCGA, and adjustment for residual tumor substantially reduced the sample size. Finally, independent validation using prespecified cohorts, protein-level assessment by immunohistochemistry, and functional studies are needed to clarify the biological significance and potential clinical utility of *CD300LF*.

5. Conclusions

CD300LF expression is associated with a myeloid-enriched, interferon-active tumor microenvironment in both LUAD and LUSC, and with favorable overall survival, particularly early survival, in LUAD but not in LUSC, with a significant difference between the subtypes. External GEO analyses showed a generally consistent direction of association, but the pooled association was modest and was attenuated after covariate adjustment, and the prognostic contribution of *CD300LF* beyond overall immune infiltration could not be

established. *CD300LF* warrants further evaluation as a LUAD-specific marker of the myeloid microenvironment in prespecified cohorts and at the protein level.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/life16122006/s1, Figure S1: *CD300LF* expression across cell types in public single-cell RNA-sequencing data; Table S1: Sample selection in TCGA LUAD and LUSC; Table S2: Sensitivity analyses of the association between *CD300LF* expression and overall survival in TCGA, added prognostic value of *CD300LF* beyond myeloid and immune markers, and pooled *CD300LF*-by-histology interaction models; Table S3: Time-varying association of *CD300LF* with overall survival in TCGA LUAD; Table S4: Correlations between *CD300LF* and selected genes; Table S5: GSEA of all 50 Hallmark gene sets with phenotype permutation; Table S6: Sensitivity GSEA comparing *CD300LF*-high and -low tumors; Table S7: Characteristics of the GEO LUAD validation cohorts; Table S8: Cox models for *CD300LF* in each GEO cohort, random-effects meta-analysis, and leave-one-out analysis.

Author Contributions: Conceptualization, C.K. and J.-H.L.; methodology, C.K.; formal analysis, C.K.; investigation, C.K.; data curation, C.K.; visualization, C.K.; writing—original draft preparation, C.K.; writing—review and editing, H.H. and J.-H.L.; supervision, H.H. and J.-H.L.; project administration, H.H. and J.-H.L. C.K. confirmed the authenticity of all the raw data. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study used publicly available, de-identified data; the original studies were approved by their respective institutional review boards.

Informed Consent Statement: Patient consent was waived because this study used only publicly available, de-identified data. Informed consent was obtained from all participants by the original studies that generated the TCGA and GEO datasets.

Data Availability Statement: The datasets analyzed in this study are publicly available. TCGA-LUAD and TCGA-LUSC RNA-sequencing data (Broad Institute GDAC Firehose), clinical information, and MC3 mutation calls were obtained from the UCSC Xena TCGA hub, and survival endpoints from the TCGA Pan-Cancer Clinical Data Resource (TCGA-CDR). Independent lung adenocarcinoma cohorts were obtained from the Gene Expression Omnibus (GEO) under accession numbers GSE31210, GSE72094, GSE36471, GSE13213, and GSE42127, and single-cell data (GSE131907 and GSE148071) from TISCH2. No new datasets were generated. The analysis code is available at https://github.com/chaeyunkim-research/TCGA-CD300LF-lung-cancer (accessed on 24 November 2026).

Acknowledgments: During the preparation of this manuscript, the authors used Claude (Opus 5.5; Anthropic, San Francisco, CA, USA) to assist with writing analysis code, generating figures, and drafting and editing the text. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bray, F.; Laversanne, M.; Sung, H.; Ferlay, J.; Siegel, R.L.; Soerjomataram, I.; Jemal, A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2024**, *74*, 229–263. [CrossRef] [PubMed]
2. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **2014**, *511*, 543–550. [CrossRef] [PubMed]
3. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **2012**, *489*, 519–525. [CrossRef] [PubMed]
4. Thorsson, V.; Gibbs, D.L.; Brown, S.D.; Wolf, D.; Bortone, D.S.; Ou Yang, T.H.; Porta-Pardo, E.; Gao, G.F.; Plaisier, C.L.; Eddy, J.A.; et al. The immune landscape of cancer. *Immunity* **2018**, *48*, 812–830.e14. [CrossRef] [PubMed]
5. Wu, P.; Wu, D.; Zhao, L.; Huang, L.; Chen, G.; Shen, G.; Huang, J.; Chai, Y. Inverse role of distinct subsets and distribution of macrophage in lung cancer prognosis: A meta-analysis. *Oncotarget* **2016**, *7*, 67323–67335. [CrossRef] [PubMed]

6. Voss, O.H.; Tian, L.; Murakami, Y.; Coligan, J.E.; Krzewski, K. Emerging role of CD300 receptors in regulating myeloid cell efferocytosis. *Mol. Cell. Oncol.* **2015**, *2*, e964625. [CrossRef] [PubMed]
7. Abadir, E.; Gasiorowski, R.E.; Lai, K.; Kupresanin, F.; Romano, A.; Silveira, P.A.; Lo, T.; Fromm, P.D.; Kennerson, M.L.; Iland, H.J.; et al. CD300f epitopes are specific targets for acute myeloid leukemia with monocytic differentiation. *Mol. Oncol.* **2019**, *13*, 2107–2120. [CrossRef] [PubMed]
8. Xu, Z.J.; Jin, Y.; Zhang, X.L.; Xia, P.; Wen, X.; Ma, J.; Lin, J.; Qian, J. Pan-cancer analysis identifies CD300 molecules as potential immune regulators and promising therapeutic targets in acute myeloid leukemia. *Cancer Med.* **2023**, *12*, 789–807. [CrossRef] [PubMed]
9. Uhlen, M.; Zhang, C.; Lee, S.; Sjöstedt, E.; Fagerberg, L.; Bidkhori, G.; Benfeitas, R.; Arif, M.; Liu, Z.; Edfors, F.; et al. A pathology atlas of the human cancer transcriptome. *Science* **2017**, *357*, eaan2507. [CrossRef] [PubMed]
10. Goldman, M.J.; Craft, B.; Hastie, M.; Repečka, K.; McDade, F.; Kamath, A.; Banerjee, A.; Luo, Y.; Rogers, D.; Brooks, A.N.; et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* **2020**, *38*, 675–678. [CrossRef] [PubMed]
11. Liu, J.; Lichtenberg, T.; Hoadley, K.A.; Poisson, L.M.; Lazar, A.J.; Cherniack, A.D.; Kovatich, A.J.; Benz, C.C.; Levine, D.A.; Lee, A.V.; et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **2018**, *173*, 400–416.e11. [CrossRef] [PubMed]
12. Ellrott, K.; Bailey, M.H.; Saksena, G.; Covington, K.R.; Kandoth, C.; Stewart, C.; Hess, J.; Ma, S.; Chiotti, K.E.; McLellan, M.; et al. Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Syst.* **2018**, *6*, 271–281.e7. [CrossRef] [PubMed]
13. Yoshihara, K.; Shahmoradgoli, M.; Martínez, E.; Vegesna, R.; Kim, H.; Torres-Garcia, W.; Treviño, V.; Shen, H.; Laird, P.W.; Levine, D.A.; et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **2013**, *4*, 2612. [CrossRef] [PubMed]
14. Aragaki, K. *tidyestimate: A Tidy Implementation of 'ESTIMATE'*, R Package Version 1.1.1; R Core Team: Vienna, Austria, 2023. Available online: https://CRAN.R-project.org/package=tidyestimate (accessed on 16 November 2026).
15. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [CrossRef] [PubMed]
16. Liberzon, A.; Birger, C.; Thorvaldsdóttir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **2015**, *1*, 417–425. [CrossRef] [PubMed]
17. Schabath, M.B.; Welsh, E.A.; Fulp, W.J.; Chen, L.; Teer, J.K.; Thompson, Z.J.; Engel, B.E.; Xie, M.; Berglund, A.E.; Creelan, B.C.; et al. Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. *Oncogene* **2016**, *35*, 3209–3216. [CrossRef] [PubMed]
18. Gene Expression Omnibus. Platform GPL96: [HG-U133A] Affymetrix Human Genome U133A Array. Available online: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL96 (accessed on 16 November 2026).
19. DerSimonian, R.; Laird, N. Meta-analysis in clinical trials. *Control. Clin. Trials* **1986**, *7*, 177–188. [CrossRef] [PubMed]
20. IntHout, J.; Ioannidis, J.P.A.; Borm, G.F. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med. Res. Methodol.* **2014**, *14*, 25. [CrossRef] [PubMed]
21. Han, Y.; Wang, Y.; Dong, X.; Sun, D.; Liu, Z.; Yue, J.; Wang, H.; Li, T.; Wang, C. TISCH2: Expanded datasets and new tools for single-cell transcriptome analyses of the tumor microenvironment. *Nucleic Acids Res.* **2023**, *51*, D1425–D1431. [CrossRef] [PubMed]
22. Kim, N.; Kim, H.K.; Lee, K.; Hong, Y.; Cho, J.H.; Choi, J.W.; Lee, J.-I.; Suh, Y.-L.; Ku, B.M.; Eum, H.H.; et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* **2020**, *11*, 2285. [CrossRef] [PubMed]
23. Wu, F.; Fan, J.; He, Y.; Xiong, A.; Yu, J.; Li, Y.; Zhang, Y.; Zhao, W.; Zhou, F.; Li, W.; et al. Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat. Commun.* **2021**, *12*, 2540. [CrossRef] [PubMed]
24. Faruki, H.; Mayhew, G.M.; Serody, J.S.; Hayes, D.N.; Perou, C.M.; Lai-Goldman, M. Lung adenocarcinoma and squamous cell carcinoma gene expression subtypes demonstrate significant differences in tumor immune landscape. *J. Thorac. Oncol.* **2017**, *12*, 943–953. [CrossRef] [PubMed]
25. Tian, L.; Choi, S.-C.; Murakami, Y.; Allen, J.; Morse, H.C., III; Qi, C.-F.; Krzewski, K.; Coligan, J.E. p85α recruitment by the CD300f phosphatidylserine receptor mediates apoptotic cell clearance required for autoimmunity suppression. *Nat. Commun.* **2014**, *5*, 3146. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.