Identification of a costimulatory molecule-based signature for predicting prognosis risk and immunotherapy response in patients with lung adenocarcinoma

CURRENT STATUS: POSTED

Chaoqi Zhang
National Cancer Center/National Clinical Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Zhihui Zhang
National Cancer Center/National Clinical Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Nan Sun
National Cancer Center/National Clinical Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Zhen Zhang
The First Affiliated Hospital of Zhengzhou University

Guochao Zhang
National Cancer Center/National Clinical Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Feng Wang
National Cancer Center/National Clinical Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Yuejun Luo
National Cancer Center/National Clinical Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Yun Che
National Cancer Center/National Clinical Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Jie He
National Cancer Center/National Clinical Center for Cancer/Cancer Hospital, Chinese Academy of
Abstract

Background Costimulatory molecules play significant roles in mounting anti-tumor immune responses, and antibodies targeting these molecules are recognized as promising adjunctive cancer immunotherapies. Here, we aim to conduct a first full-scale exploration of costimulatory molecules from the B7-CD28 and TNF families in patients with lung adenocarcinoma (LUAD) and generated a costimulatory molecule-based signature (CMS) to predict survival and response to immunotherapy.

Methods We enrolled 1549 LUAD cases across 10 different cohorts and included 502 samples from TCGA for discovery. The validation set included 970 cases from eight different Gene Expression Omnibus (GEO) datasets and 77 frozen tumor tissues with qPCR data. The underlying mechanisms and predictive immunotherapy capabilities of the CMS were also explored.

Results A five gene-based CMS (CD40LG, TNFRSF6B, TNFSF13, TNFRSF13C, and TNFRSF19) was initially constructed using the bioinformatics method from TCGA that classifies cases as high- vs. low-risk groups per OS. Multivariable Cox regression analysis confirmed that the CMS was an independent prognostic factor. As expected, CMS exhibited prognostic significance in the stratified cohorts and different validation cohorts. Additionally, the prognostic meta-analysis revealed that CMS was superior to the previous signature. Samples in high- and low-risk groups exhibited significantly different tumor-infiltrating leukocytes and inflammatory activities. Importantly, we found that signature high-risk patients were optimal candidates for immunotherapy.

Conclusion We conducted the first and most comprehensive costimulatory molecule landscape analysis of patients with LUAD and built a clinically feasible CMS for prognosis and immunotherapy response prediction, which will be helpful for further optimize immunotherapies for cancer.

Background

Over the last few years, lung cancer has become the most common malignant tumor and is a grave danger to global human health, with an annual incidence increasing at a rate of 7.5%.\(^1\)

Approximately four out of five lung cancers are classified as non-small cell lung cancer (NSCLC). As the major histological subtype of NSCLC, lung adenocarcinoma (LUAD) accounts for over 1 million worldwide deaths annually.\(^2\) Despite the amplification of traditional approaches which—in
combination with targeted therapy—have reduced mortality, the five-year OS (OS) rate of LUAD remains about 15%.\cite{3} The introduction of immunotherapy, especially immune checkpoint inhibitors (ICIs) targeting programmed cell death protein 1 (PD-1) and programmed cell death 1 ligand 1 (PD-L1), has revolutionized lung cancer treatment.\cite{4, 5} More recently, pembrolizumab monotherapy was approved to replace chemotherapy as the frontline treatment for patients with PD-L1 positive metastatic NSCLC.\cite{6} Although treatment for lung cancer has been improved with the development of ICI-based immunotherapies, only a small proportion of patients with lung cancer can benefit from this schedule. Therefore, we must be able to predict the best candidates for immunotherapy and develop other novel immune checkpoint targets.

The success of ICIs has emerged from a deep understanding of the functions of the immune system and immunosuppressive conditions that are generated in the tumor microenvironment (TME).\cite{7, 8} In the TME, T cells help distinguish cancer cells from healthy cells and initiate subsequent attacks. Before the attack, the naïve T cells need two signals to be active. The first signal is generated once a specific antigen is recognized by the T cell receptor (TCR). The second signal is a non-specific costimulatory signal.\cite{9} Based on the fact that the naïve T cells cannot be activated in without costimulatory signals,\cite{10} cancer cells prevent the recognition of these signals by changing the costimulatory molecule signals and expressions in the TME.\cite{11} Hence, ICIs prevent tumor cells from delivering incorrect messages to T cells, thereby selectively restoring a tumor-induced immune deficiency in the TME.\cite{12} In addition to the best described immune checkpoint pathways (PD-L1/PD-1, CD86/CTLA4) that belong to the B7-CD28 family,\cite{13, 14} other co-stimulation pathways mainly arise from the tumor necrosis factor (TNF) family.\cite{15}

Currently, 13 molecules are classified as B7-CD28 family members, including eight molecules (CD80, CD86, PD-L1, PD-L2, ICOSLG, B7-H3, B7x, and HHLA2) that belong to the B7 family and five molecules (CD28, CTLA4, ICOS, PD-1, and TMIGD2) that belong to the CD28 family.\cite{13} The TNF family consists of the TNF ligand superfamily (TNFSF) and the TNF receptor superfamily (TNFRSF) with 48 molecules.
Nineteen legends were defined as TNFSF, and other 29 receptors considered members of the TNFRSF (Table 1). These costimulatory molecules—consisting of members of the B7-CD28 and TNF families—constitute potential molecular targets for the development of novel ICIs and may make excellent additions to existing immunotherapeutic strategies.\textsuperscript{17, 18} However, the expression patterns and clinical significance of the majority of these members remain unknown. There is a need for full-scale investigations of these molecules in patients with LUAD.

Table 1

| Official symbol | Aliases | Family | HR   | 95%CI       | P value |
|-----------------|---------|--------|------|-------------|---------|
| CD27            | TNFRSF7 | TNFRSF | 0.8682 | 0.7858–0.9592 | 0.0055  |
| CD274           | PD-L1, B7-H1 | B7 | 1.0125 | 0.9239–1.1097 | 0.7897  |
| CD276           | B7-H3   | B7    | 1.3865 | 1.0728–1.7919 | 0.0125  |
| CD28            | Tp44    | CD28  | 0.8635 | 0.7721–0.9658 | 0.0102  |
| CD40            | TNFRSF5 | TNFRSF | 0.9089 | 0.8063–1.0246 | 0.1182  |
| CD40LG          | TNFRSF5, CD154, CD40L | TNFRSF | 0.8194 | 0.7421–0.9049 | 0.0000  |
| CD70            | TNFRSF7, CD27L | TNFRSF | 1.0688 | 0.9757–1.1708 | 0.1522  |
| CD80            | B7-1, CD28LG1 | B7 | 0.8878 | 0.7986–0.9869 | 0.0275  |
| CD86            | B7-2, CD28LG2 | B7 | 0.9099 | 0.8076–1.0252 | 0.1210  |
| CTLA4           | CD152   | CD28  | 0.8772 | 0.7924–0.9711 | 0.0116  |
| EDA             | EDA-A1, EDA-A2 | TNFRSF | 0.9561 | 0.8727–1.0474 | 0.3345  |
| EDA2R           | TNFRSF27, XEDAR | TNFRSF | 0.9107 | 0.8387–0.9988 | 0.0259  |
| EDAR            | EDA-A1R  | TNFRSF | 1.0425 | 0.9744–1.1154 | 0.2275  |
| FAS             | TNFRSF6, CD95 | TNFRSF | 0.9501 | 0.8396–1.0752 | 0.4176  |
| FASLG           | TNFRSF6, CD95-L | TNFRSF | 0.9083 | 0.8204–1.0056 | 0.0641  |
| HHLA2           | B7-H5   | B7    | 1.0016 | 0.9602–1.0448 | 0.9414  |
| ICOS            | CD278, CVID1 | CD28 | 0.8970 | 0.8092–0.9942 | 0.0384  |
| ICOSLG          | B7-H2, CD275 | B7 | 0.8427 | 0.7104–0.9996 | 0.0495  |
| LTA             | TNFRSF1 | TNFRSF | 0.8918 | 0.8002–0.9939 | 0.0385  |
| LTB             | TNFRSF3 | TNFRSF | 0.8771 | 0.7949–0.9679 | 0.0092  |
| LTBR            | TNFRSF3 | TNFRSF | 1.3077 | 1.0591–1.6147 | 0.0126  |
| NGFR            | TNFRSF16, CD271 | TNFRSF | 1.0011 | 0.9123–1.0986 | 0.9808  |
| PDCD1           | PD-1, CD279 | CD28 | 0.9625 | 0.8693–1.0658 | 0.4630  |
| PDCD1LG2        | PD-1, B7DC, CD273 | B7 | 0.9691 | 0.8733–1.0754 | 0.5544  |
| RELT            | TNFRSF19L | TNFRSF | 0.9259 | 0.7884–1.0873 | 0.3477  |
| TMIGD2          | CD28H   | CD28  | 0.9290 | 0.8279–1.0423 | 0.2096  |
| TNF             | TNFRSF2, TNFA | TNFRSF | 0.9098 | 0.8257–1.0024 | 0.0560  |
| TNFRSF10A       | TRAILR1, CD261 | TNFRSF | 1.0582 | 0.9025–1.2408 | 0.4858  |
| TNFRSF10B       | TRAILR2, CD262 | TNFRSF | 1.0336 | 0.8482–1.2596 | 0.7428  |
| TNFRSF10C       | TRAILR3, CD263 | TNFRSF | 0.8628 | 0.7727–0.9634 | 0.0087  |
| TNFRSF10D       | TRAILR4, CD264 | TNFRSF | 1.0799 | 0.9334–1.2495 | 0.3015  |
| TNFRSF11A       | RANK, CD265 | TNFRSF | 1.1316 | 0.9925–1.2902 | 0.0647  |
| TNFRSF11B       | OPG     | TNFRSF | 1.0088 | 0.9139–1.1135 | 0.8619  |
| TNFRSF12A       | FN14, TWEAKR, CD266 | TNFRSF | 1.1040 | 0.9768–1.2477 | 0.1131  |
| TNFRSF13B       | TACI, TNFRSF14B, CD267 | TNFRSF | 0.8682 | 0.7987–0.9438 | 0.0009  |
| TNFRSF13C       | BAFFR, CD268 | TNFRSF | 0.8788 | 0.7977–0.9682 | 0.0090  |
| TNFRSF14        | LIGHTR, HVEM, CD270 | TNFRSF | 0.8253 | 0.7112–0.9577 | 0.0114  |
| TNFRSF17        | BCMA, TNFRSF13A, CD269 | TNFRSF | 0.9023 | 0.8370–0.9727 | 0.0073  |
| TNFRSF18        | GITR, AITR, CD266 | TNFRSF | 0.9919 | 0.9060–1.0860 | 0.8607  |
We used LUAD gene expression data from The Cancer Genome Atlas (TCGA) to systematically explore the expression patterns and prognoses of these costimulatory molecules. Then, through a series of statistical methods, we built a costimulatory molecule-based signature (CMS) with significantly different prognoses. The CMS was well-validated in nine different cohorts from Gene Expression Omnibus (GEO) datasets and an independent cohort using clinical samples. Also, according to a prognostic meta-analysis, we determined that CMS was superior to the previous costimulatory molecule-related model. We also found that the CMS was characterized by distinct inflammatory profiles and specific immune infiltrating lymphocytes. What’s more, the CMS was able to predict the immunotherapy response in patients with LUAD. Therefore, our work describes the systemic landscape of costimulatory molecules based on B7-CD28 and TNF families and highlights the potential underlying clinical applications for the CMS, thereby supporting the development of rationales to guide prognosis management and immunotherapy in patients with LUAD.

Materials And Methods

mRNA expression datasets and clinical information

A total of nine public datasets, including 1472 cases with corresponding mRNA expression data and clinical data, were gathered in this study. The training set consisted of data from 502 patients with genetic information and matching OS data from TCGA that were downloaded from the Cancer
Genomics Browser of University of California Santa Cruz (UCSC) (https://genomecancer.ucsc.edu).[19]

Eight other public datasets with mRNA microarray data were collected from GEO datasets (http://www.ncbi.nlm.nih.gov/geo), including GSE11969 (n = 91),[20] GSE13213 (n = 117),[21] GSE19188 (n = 40),[22] GSE30219 (n = 83),[23] GSE31210 (n = 226),[24] GSE37745 (n = 106),[25] GSE41271 (n = 180),[26] and GSE50081 (n = 127).[27] All the microarray data from the GEO datasets were first log2 transformed and quantile normalized. Moreover, for the genes with one more probe, mean expression values were recognized as the expression data. The clinical characteristics of these patients from multiple institutions are summarized in Table 2.

**RNA extraction and quantitative real-time reverse transcription–PCR**

We used 77 surgically resected LUAD tissues, collected from The First Affiliated Hospital of Zhengzhou University between August 2013 and January 2015, as the independent cohort. Then total RNA was extracted from LUAD tissues using the RNAiso Plus reagent (Takara, #9109) according to the manufacturer’s instructions. The first strand of complementary DNA was synthesized from total RNA using the Prime Script™ RT reagent kit (Takara, #RR047A). Quantitative real-time PCR was performed with SYBR Premix Ex Taq II (Takara, #RR820A), and data were analyzed in the Agilent Mx3005P. With the endogenous control for normalization of GAPDH, the expression data of all the selected genes were log2 transformed before signature validation. All the primer sequences in this research are displayed in Supplementary Table 1. The samples used in the study were approved by the Institutional Review Boards of the First Affiliated Hospital of Zhengzhou University.

**Functional enrichment analysis**

After deleting the genes with low expression values, functional enrichment analysis based on CMS related genes were conducted through the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) in DAVID 6.8 (http://david.abcc.ncifcrf.gov).

**Estimated the profiling of immune cell infiltration**

The mRNA expression data of LUAD from TCGA was used for estimating the fractions of 22 immune cell types in the TME by CIBERSORT.[28] The profiling of multiple immune cell types was performed through the leucocyte gene signature matrix, termed LM22, for the CIBERSORT software.
LM22 consists of 547 genes that can distinguish 22 immune cells, including different subtypes of B cell types, T cell types, natural killer cells (NKs), plasma cells, and myeloid cell types.

**Biomarkers for predicting immunotherapy response**

The potential immunotherapy response prediction performance of CMS was estimated with the following biomarkers: tumor mutation burden (TMB), neoantigen, PD-L1 protein expression, and Tumor Immune Dysfunction and Exclusion (TIDE) score. The TMB and neoantigen data of LUAD patients in the TCGA dataset were separated from The Cancer Immunome Atlas (TCIA) (https://tcia.at/home).[29] The protein data of PD-L1 expression was realized through the reverse-phase protein array (RPPA) analysis, which was retrieved from cbioPortal (http://www.cbioportal.org). TIDE has been proven to outperform known immunotherapy biomarkers in predicting immunotherapy response in patients with melanoma and lung cancer, especially those treated with ICIs.[30] TIDE scores, T cell dysfunction scores, and T cell exclusion scores were downloaded from the TIDE web (http://tide.dfci.harvard.edu) after following the instructions on the website to upload input data.

**Signature identification and statistical analysis**

A univariate Cox proportional hazards regression analysis and stepwise Cox proportional hazards regression model were used to construct the signature. Then, the CMS was constructed using five selected genes with a linear combination of their expression values. These inputs were weighted with the regression coefficients from the stepwise Cox regression analyses. All the patients in different cohorts were divided into high- and low-risk groups based on the optimal cutoff point. The prognostic significance of the CMS between the high- and low-risk groups in different sets and subgroups were calculated with Kaplan-Meier curves and a 2-tailed log-rank test. The Mann-Whitney U-test was applied to analyze the between-group differences for immune cell fractions and immunotherapy biomarkers. Univariate and multivariate Cox regression analyses were conducted to clarify the independent prognostic factors. $P < 0.05$ was considered statistically significant for all statistical methods. STATA software (version 12.0) was used to perform the prognostic meta-analysis of CMS and B7-CD 28 signature. The two overall hazard ratio (HR) values were calculated using the random-
effects model. R software version 3.5.1 (https://www.r-project.org) was used for other data analyses.

Results
The panorama and prognostic significance of costimulatory molecule genes in LUAD

A total of 60 costimulatory molecule genes were separated from the TCGA LUAD data, which consisted of 13 well-defined B7-CD28 family costimulatory molecules,[13] and 47 TNF family costimulatory molecules (Table 1).[16] The relationships of these molecules are shown in Supplementary Fig. 1. Correlation analysis revealed that most of the costimulatory molecules were highly relevant to others. Then, 502 LUAD patients with 60 costimulatory molecule expression data and matched complement OS information from the TCGA data were used to evaluate the prognostic significance of these candidate genes. Univariate cox proportional hazards regression analysis was conducted, and the results showed that 23 genes were significantly associated with OS ($P < 0.05$, Table 2). Among the significant genes, four genes (CD276, LTBR, TNFRSF1A and TNFRSF6B) were confirmed as risky factors with HRs (HR) > 1, and 19 genes (CD27, CD28, CD40LG, CD80, CTLA4, EDA2R, ICOS, ICOSLG, LTA, LTB, TNFRSF10C, TNFRSF13B, TNFRSF13C, TNFRSF14, TNFRSF17, TNFRSF19, TNFSF12, TNFSF13 and TNFSF8) were confirmed as protective factors with HR < 1.
Table 2
Clinical characteristics of the enrolled patients.

| Characteristic | TCGA N=502 | GSE11969 N=90 | GSE19188 N=40 | GSE30 N=83 | GSE31210 N=226 | GSE37 N=106 | GSE41271 N=180 | GSE50081 N=127 | Independent N=77 |
|----------------|------------|---------------|---------------|------------|---------------|-------------|----------------|---------------|----------------|
| Age, year Mean | 65.3       | 61.0          | 60.7          | -          | 61.1          | 59.6        | 63.0           | 64.1          | 68.7          | 60.0          |
| Gender Male    | 231        | 47            | 60            | 25         | 65            | 105         | 46             | 91            | 65            | 39            |
|               | 271        | 43            | 57            | 15         | 18            | 121         | 60             | 89            | 62            | 38            |
| Smoking history Yes | 416       | 45            | 61            | -          | -             | 111         | -              | -             | 92            | 46            |
|               | 72         | 45            | 56            | -          | -             | 115         | -              | -             | 23            | 31            |
|               | NA         | 14            | 0             | -          | -             | 0           | -              | -             | 12            | 0             |
| TNM stage I and II | 388       | 65            | 92            | -          | 83            | 226         | 89             | 129           | 127           | 62            |
| III and IV    | 105        | 25            | 25            | -          | 0             | 0           | 17             | 51            | 0             | 15            |
|               | NA         | 9             | 0             | -          | 0             | 0           | 0              | 0             | 0             | 0             |
| OS state Alive | 320        | 50            | 68            | 16         | 40            | 191         | 29             | 111           | 76            | 57            |
|               | Death      | 182           | 40            | 49         | 24            | 43          | 35             | 77            | 69            | 51            |

NA, not available; OS, overall survival.

Identification of CMS for prognostication

With the tremendous success in the clinical use of ICIs targeting costimulatory factors for lung cancer, we sought to establish CMS for prognostication. A stepwise Cox proportional hazards regression model was then used to filter out the redundant candidate genes and construct an optimized model. We then developed a risk score formula for patients with LUAD based on the gene’s expression levels to predict patient survival: risk score=(-0.1075 × CD40LG) + (0.1418 × TNFRSF6B) + (-0.1603 × TNFSF13) + (-0.1069 × TNFRSF13C) + (-0.0803 × TNFRSF19). The expression panel of the five genes, the distribution of risk scores, and survival status of each patient are shown in Fig. 1A. Next, we classified all the patients in the TCGA cohort into high-risk (n = 292) and low-risk groups (n = 210) based on the optimal cutoff point. We found that patients in the high-risk group showed significantly worse OS (Fig. 1B, HR 2.0435, 95% confidence interval (CI) 1.4811–2.8195, P < 0.0001). When we further applied the signature into different clinical stages, the results indicated that the formula still worked well. Specifically, we observed significant OS time between the high- and low-risk groups both for early-stage (stage I and II) (Fig. 1C, HR 1.9961, 95% CI 1.3641–2.9210, P = 0.0003) and advanced...
stage disease (stage III and IV) (Fig. 1D, HR 2.7529, 95% CI 1.6335–4.6394, P < 0.0001).

To further explore whether the signature-based risk score was an independent factor in patients with LUAD, univariate, and multivariate Cox regression analyses in the TCGA database were conducted. The results of the multivariate Cox regression model confirmed that the risk score was a significant factor (HR = 1.7952, 95% CI 1.2254–2.6298, P = 0.0027) independent of age, sex, smoking history, clinical stage, and mutation (MUT) status (Table 3).

### Table 3

Univariable and multivariable Cox regression analysis of the costimulatory molecule-based signature and survival in TCGA dataset.

| Variable               | Univariable analysis | Multivariable analysis |
|------------------------|----------------------|------------------------|
|                        | HR               | 95%CI                  | P value | HR               | 95%CI                  | P value |
| Age ≥ 60 or < 60       | 1.1575            | 0.7957–1.6838          | 0.4445  | 1.3959            | 0.9490–2.0533          | 0.0903  |
| Gender Male or Female  | 1.1568            | 0.8401–1.5928          | 0.3722  | 1.0080            | 0.7183–1.4145          | 0.9634  |
| Smoking history Yes or No | 1.0374          | 0.6532–1.6476          | 0.8763  | 1.0613            | 0.6366–1.7692          | 0.8195  |
| T stage 1, 2, 3 or 4   | 1.5458            | 1.2602–1.8961          | <0.0001 | 1.2890            | 1.0171–1.6336          | 0.0357  |
| Lymphatic metastasis   | 2.4053            | 1.7466–3.3124          | <0.0001 | 1.6992            | 1.1150–2.5897          | 0.0137  |
| TNM stage I, II, III or IV | 1.5587        | 1.3381–1.8156          | <0.0001 | 1.1711            | 0.9242–1.4839          | 0.1912  |
| ERFR status MUT or WT  | 1.4658            | 0.9584–2.2418          | 0.0777  | 1.4682            | 0.9181–2.3480          | 0.1089  |
| KRAS status MUT or WT  | 1.2159            | 0.8598–1.7195          | 0.2689  | 1.2525            | 0.8706–1.8019          | 0.2250  |
| Risk score High or low | 2.1058            | 1.4571–3.0433          | 0.0001  | 1.7952            | 1.2254–2.6298          | 0.0027  |

Abbreviations: HR, hazard ratio; CI, confidence interval; WT, wild-type; MUT, mutation.

Evaluation of the performance of CMS in different clinical subgroups

Sex, age, smoking history, and MUT status were factors that influenced the TME, especially the expression of immune checkpoints. Consequently, patients from TCGA were then divided into different subgroups based on these parameters: sex (male or female), age [older (age ≥ 60) or younger (age < 60)], smoking (smoker or non-smoker), and MUT status [EGFR wide-type (WT), EGFR MUT, KRAS WT, KRAS MUT, or EGFR/KRAS WT]. All the patients in different subgroups were stratified into high- and low-risk groups based on the risk score with the same formula. The results showed
that all the high-risk groups had significantly different OS compared to the paired low-risk groups (Supplementary Figs. 2 and 3, P < 0.05).

Validation of the CMS in nine independent cohorts
To identify whether the CMS derived from the TCGA cohort was robust, we first evaluated its performance in eight independent public validation cohorts. These consisted of the remaining GSE11969, GSE13213, GSE19188, GSE30219, GSE31210, GSE37745, GSE41271, and GSE50081 datasets. The CMS stratified all patients from different public cohorts into the high- and low-risk groups using the same formula with the optimal cutoff points. As shown in Fig. 2, significant differences between the high- and low-risk groups were found in most of the GEO datasets, including GSE13213 (HR 2.5990, 95% CI 1.3539–4.9890, P = 0.0029), GSE19188 (HR 2.4817, 95% CI 1.0571–5.8262, P = 0.0308), GSE30219 (HR 2.2955, 95% CI 1.1495–4.5839, P = 0.0156), GSE31210 (HR 2.2037, 95% CI 1.0960–4.4308, P = 0.0229), GSE41271 (HR 2.3023, 95% CI 1.4267–3.7153, P = 0.0004) and GSE50081 (HR 2.2958, 95% CI 1.2393–4.2530, P = 0.0066). Meanwhile, in the GSE11969 and GSE37745 datasets, the signature showed a borderline difference between the high- and low-risk groups with P values of 0.1015 and 0.1192, respectively (Fig. 2A and F).

To further measure whether the signature could be used in clinical practice, we validated the signature in an independent cohort that contained 77 frozen tissue samples with qRT-PCR data. By using the same model and the optimal cutoff point, patients were classified into high- (n = 32) and low-risk groups (n = 45). As expected, a significant difference in mortality was found between these two groups (Fig. 2I, HR 2.9189, 95% CI 1.1622–7.3309, P = 0.0169).

Compare the CMS with the previous model
Prior to the creation of our signature, Shanbo Zheng et al. constructed a signature for LUAD based on the costimulatory molecules from the B7-CD28 family (B7-CD28 signature) with a risk score of 0.3313 × CD276–0.1559 × CD28.¹³¹ We then comprehensively assessed the prognostic significance of our CMS and the B7-CD28 signature by examining public datasets and conducting prognostic meta-analyses based on the nine groups (n = 1472) of the two different signature groups. As shown in Fig. 3A, our CMS performed very well in the different cohorts, producing HRs larger than 1. On the
contrary, the B7-CD28 signature was not that stable in different cohorts and some of the HRs were smaller than 1 (Fig. 3B). More importantly, the meta-analysis combined HR of our CMS was far larger than that of the B7-CD28 signature. These findings indicate that our signature was superior to the previous model.

**CMS related biological processes and pathways**

The consistent prognostic performance of the CMS was confirmed in 10 different cohorts. This prompted us to investigate the biological features of patients with different risk scores. We first filtered out low-expression genes (genes where half or more than half of the values were 0) and then extracted the genes that strongly correlated with risk score (the top 600 genes ranked by Pearson |R|) from the TCGA dataset. Collectively, 36 positively related genes and 564 negatively related genes were screened out (Fig. 4A). Then, these selected genes were chosen for GO and KEGG analyses through use of the online DAVID tool (https://david.ncifcrf.gov). The results revealed that signature-related genes were more involved in the biological process of the immune response, especially B cell and T cell-related immune response (Fig. 4B). KEGG analysis further confirmed that these genes were closely related to immune-specific pathways (Fig. 4C).

**CMS-related immune cell infiltration and inflammatory activities**

To further increase our understanding of the CMS-related immune landscape, we first explored the relationship between CMS and immune cell infiltration. The estimated fractions of different immune cells in the TME of LUAD were calculated by CIBERSORT, in combination with the LM22. The results demonstrated that the panorama of immune cells between high- and low-risk patients were dramatically different (Fig. 5A). In particular, high-risk patients showed a significantly higher proportion of activated NK cells, activated dendritic cells (DCs), neutrophils, macrophages M0, resting DCs, and regulator T cells (Tregs) (Fig. 5B and C). On the contrary, low-risk patients featured a high proportion of memory B cells, resting CD4 memory T cells, and gamma delta T cells (Fig. 5B and C).

Next, to increase our understanding of CMS-related inflammatory activities, we assessed the relationship between CMS and seven clusters of metagenes. These consisted of 104 genes and represented different inflammation and immune response.[32] The expression details of the collected
genes and risk scores were displayed in Fig. 5D. Then, to explore the correlation between CMS and the entire metagenes of every cluster, the expression of corresponding gene clusters was calculated by Gene Sets Variation Analysis (GSVA).

Finally, the correlations were portrayed according to Pearson r-values between risk scores and metagenes (Fig. 5E). The results revealed that CMS was negatively associated with HCK, LCK, MHC-I, and MHC-II. This indicated that patients with high CMS scores were characterized by an immune-suppressive state.

**Association of CMS and immunotherapy response in patients with LUAD**

Presently, immunotherapy is considered a first-line treatment for patients with LUAD. Costimulatory molecules are major candidates for immunotherapy. Therefore, we further assessed the association of CMS and immunotherapy response through analyzing the correlation of CMS and widely recognized immunotherapy biomarkers.

Totally, we enrolled eight indices, including TMB, the number of neoantigens, the number of clonal neoantigens, the number of subclonal neoantigens, the protein level of PD-L1, the TIDE score, the T cell dysfunction score, and the T cell exclusion score, to get a comprehensive evaluation. The results, as depicted in Fig. 6, illustrated that high-risk patients were distinguished by a high level of TMB, neoantigens, protein level of the PD-L1 and T cell exclusion scores, and low level of the TIDE and T cell dysfunction scores. These results indicate that CMS-based high-risk patients are the optimum candidates for immunotherapy, especially ICIs.

**Discussion**

There is plenty of evidence pointing out that the immunosuppressive TME exhausts T cells and renders them anergic. This subsequently enables tumor cells to evade host immune-mediated elimination.

Costimulatory molecules, especially the immune checkpoints, expressed on cancer cells or tumor-infiltrating lymphocytes play vital roles in regulating the anti-tumor immune response. Further, the blocking antibody targeting PD-L1/PD-1 has directly prolonged survival in patients with metastatic cancer. Presently, the costimulatory molecules mainly consist of two major families: the B7-CD28 family and the TNF family. In this study, we simultaneously detected the expression pattern and clinical significance of 60 costimulatory molecules in patients with LUAD. Based on the significant genes, we developed a novel survival prediction model (CMS) based on the
expression of five costimulatory molecular features in the TCGA dataset. The CMS score was found as an independent risk factor for patients with LUAD. Furthermore, the CMS was well validated in eight different public GEO datasets and 77 cases from frozen tissues with qRT-PCR data. Interestingly, through prognostic meta-analysis, we proved that our CMS had better prognostic value than the previous costimulatory molecule-related signature. We also explored the immune panorama—including immune cell distribution and inflammatory activities—in CMS high- and low-risk patients. More importantly, we found that CMS high-risk patients were considered the optimum candidates for immune checkpoint-based immunotherapies. To our knowledge, this is the first and most comprehensive study to date to describe the prognostic and immunotherapy response prediction value of a CMS in patients with LUAD.

To get the whole picture of costimulatory molecule expression in patients with LUAD, we collected the 13 members from the B7-CD28 family and the 47 members from the TNF family into our analysis.[13, 16] After the univariate Cox proportional hazards regression analysis and stepwise Cox proportional hazards regression model, we found that all five selected genes (CD40LG, TNFRSF6B, TNFSF13, TNFRSF13C, and TNFRSF19) belonged to the TNF family. This indicated that costimulatory signals and pathways in the TNF family had a more important prognostic value than those in the B7-CD28 family in patients with LUAD. CD40LG—also known as CD40L, TNFSF5, or CD154—is a membrane-bound protein belonging to the TNFSF family. CD40LG has been a therapy target in cancer treatment because of its ability to trigger Th1-type immune responses.[38] The expression and prognostic states of the CD40LG-CD40 axis was previously reported in lung cancer.[39] TNFRSF6B, a soluble decoy receptor, is also known as Decoy receptor 3 (DcR3), belongs to the TNFRSF family.[40] TNFRSF6B inhibits apoptosis and promotes angiogenesis through binding with FASL, LIGHT, and TL1A.[41, 42] Moreover, studies found that DcR3 is a potential immunotherapy target for cancer treatment.[43] TNFSF13, also known as APRIL and CD256, is a proliferation-inducing ligand that plays an important role in B cell development.[44] The clinical significance of TNFSF13 in several cancers was previously
revealed and included NSCLC,[45] breast cancer,[46] B-cell chronic lymphocytic leukemia,[47] and other tumor types. TNFRSF13C (BAFFR or CD268), a receptor of BAFF, is a crucial regulatory factor in B cell proliferation, development, and maturation.[48] Hong Qin et al. reported that a novel anti-BAFFR antibody may be a promising strategy for drug-resistant B-cell malignancies.[49] TNFRSF19, also known as TROY or TAJ, is a member of the TNFRSF family and demonstrates complex and pleiotropic functions in different cellular contexts.[50] Present evidence displayed that TNFRSF19 acted as a tumor suppressor in patients with lung cancer.[51] Although the expression details of these five members in various cancer types have been described, the combination and functions of these molecules still warrants further exploration.

To verify the robustness of CMS, we reproduce the model in nine different cohorts, and the significance of CMS was finally confirmed by prognosis meta-analysis. It is worth mentioning that the number of validation cohorts in our research was larger than that of any other studies in the LUAD population. This made our signature more reliable and clinically feasible. Before our study, a signature based on the expression of costimulatory molecules from the B7-CD28 family was constructed.[31] Through meta-analysis, we obtained two crucial conclusions: the CMS signature had prognostic significance across these public datasets, although some of the $P$-values were not statistically significant and our CMS model demonstrated an advantage over the reported B7-CD28 model. These conclusions are consistent with our finding that the TNF family has a more important prognostic value for patients with LUAD.

Through analysis, the most related genes of CMS, the potential mechanisms of CMS in LUAD was proved to be closely associated with the immune-related process. Hence, the details of CMS-specific immune profiles were further analyzed. We found that there were higher proportions of DCs, NKs, and Tregs in CMS high-risk patients TME. Simultaneously, inflammatory metagene analysis revealed that CMS score was negatively related to monocyte/myeloid lineage- and T cell-specific functions (HCK and LCK). What's more, CMS score was also found negatively related to the antigen-presenting process of T cells (MHC-I and MHC-II) in LUAD. Thus, CMS high-risk patients appear to exhibit a high immune cell
infiltration microenvironment while in an immune-suppressive state. Interestingly, this research highlighted the potential role of CMS in predicting the response to immunotherapy in patients with LUAD. Because the immune checkpoint targets (PD-L1 and PD-1) are costimulatory molecules, CMS may have the ability to predict the response to ICIs-based immunotherapy. Due to the lack of details regarding mRNA expression in cases with immunotherapy, we had to evaluate the relationship indirectly. We collected TMB, the number of neoantigens, the protein level of PD-L1, and the TIDE scores. TMB is one of the classic biomarkers for immunotherapy response, and neoantigen burden is always increased by TMB. This will be useful for T cell recognition. The PD-L1 expression level was another well-known biomarker for ICIs in lung cancer. The TIDE score is a newly-developed method for immunotherapy response prediction, and considered a more accurate biomarker than TMB or PD-L1 expression. Collectively, high-risk patients exhibited high TMB and PD-L1 expression. From a mechanical standpoint, this resonated with the results of the immune profile analysis. By comparing the CMS scores with these different verified biomarkers, we confirmed that CMS high-risk patients were more suitable for immunotherapy. These findings give us additional confidence that the CMS scores may act as a novel predictive biomarker for immunotherapy response.

There are some limitations to this study that warrant consideration. Firstly, although we tried our best to include as many independent datasets as possible for validation, this study was retrospective. Secondly, the CMS-specific immune landscape was realized through bioinformatic methods with RNA-seq data. This analysis may have been influenced by noise. Thirdly, because the mRNA expression data from patients with immunotherapy was not available, the prediction ability of CMS for immunotherapy response was estimated indirectly. Future prospective studies could affirm the complete prediction ability and a molecular picture of the CMS signature.

Conclusions
Here, we have performed a first costimulatory molecule landscape analysis in patients with LUAD. We built a reliable, clinically feasible prognostic signature named CMS and identified the potential underlying immune-related mechanisms of this signature. Importantly, the CMS had potential value
for immunotherapy response prediction. Thus, the CMS could be a clinically useful tool for prognostic management and predicting immunotherapy response in patients with LUAD. Future validation of the predictive capability of this formula will be helpful for patients seeking counseling and individualized treatment.

Abbreviations

| Abbreviation | Definition |
|--------------|------------|
| NSCLC        | non-small cell lung cancer |
| LUAD         | lung adenocarcinoma |
| OS           | overall survival |
| ICIs         | immune checkpoint inhibitors |
| PD-1         | programmed cell death protein 1 |
| PD-L1        | programmed cell death 1 ligand 1 |
| TME          | tumor microenvironment |
| TCR          | T cell receptor |
| TNF          | tumor necrosis factor |
| TNFSF        | TNF ligands superfamily |
| TNFRSF       | TNF receptors superfamily |
| TCGA         | The Cancer Genome Atlas |
| CMS          | costimulatory molecule-based signature |
| GEO          | Gene Expression Omnibus |
| UCSC         | University of California Santa Cruz |
| qPCR         | quantitative real-time polymerase chain reaction |
| GO           | gene ontology |
| KEGG         | Kyoto Encyclopedia of Genes and Genomes |
| NKs          | natural killer cells |
| TMB          | tumor mutation burden |
| TIDE         | Tumor Immune Dysfunction and Exclusion |
| TCIA         | The Cancer Immunome Atlas |
| RPPA         | the reverse phase protein array |
| HR           | hazard ratio |
| CI           | confidence interval |
| WT           | wild-type |
| MUT          | mutation |
| Dcs          | dendritic cells |
| Tregs        | T cells regulatory |
| GSVA         | Gene Sets Variation Analysis |
| DcR3         | Decoy Receptor 3 |

Declarations

**Ethics approval and consent to participate**

Not applicable

**Consent for publication**

Not applicable

**Availability of data and materials**

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.
**Competing interests**

The authors declare that they have no competing interests.

**Funding**

This work was supported by the CAMS Innovation Fund for Medical Sciences (2017-I2M-1-005, 2016-I2M-1-001), the National Key R&D Program of China (2016YFC1303201), the National Natural Science Foundation of China (81802299, 81502514), the Fundamental Research Funds for the Central Universities (3332018070), and the National Key Basic Research Development Plan (2018YFC1312105).

**Authors’ contributions**

All authors contributed to data analysis, drafting or revising the article, gave final approval of the version to be published. Chaoqi Zhang, Zhihui Zhang, Nan Sun were contributed equally to this work. All authors read and approved the final manuscript.

**Acknowledgments**

All authors would like to thank the specimen donors and research groups for TCGA, GSE11969, GSE13213, GSE19188, GSE30219, GSE31210, GSE37745, GSE41271, and GSE50081 that provided data for this collection.

**References**

1. Jemal A, Bray F, Center MM, et al. Global cancer statistics. Cancer J Clin. 2011;61:69-90.

2. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin. 2018;68:394–424.

3. Chen Z, Fillmore CM, Hammerman PS, et al. Non-small-cell lung cancers: a heterogeneous set of diseases. Nature reviews Cancer. 2014;14:535–46.
4. Herbst RS, Baas P, Kim DW, Felip E, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet. 2016;387:1540–50.

5. Borghaei H, Paz-Ares L, Horn L, et al Vokes EE, Felip E, Holgado E, Barlesi F, Kohlhauff M, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. N Engl J Med. 2015;373:1627–39.

6. Reck M, Rodriguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. N Engl J Med. 2016;375:1823–33.

7. Nurieva R, Thomas S, Nguyen T, et al. T-cell tolerance or function is determined by combinatorial costimulatory signals. EMBO J. 2006;25:2623–33.

8. Singh S, Hassan D, Aldawsari HM, et al. Immune checkpoint inhibitors: a promising anticancer therapy. Drug discovery today. 2020;25:223–9.

9. Bluestone JA. New perspectives of CD28-B7-mediated T cell costimulation. Immunity. 1995;2:555–9.

10. Goronzy JJ, Weyand CM. T-cell co-stimulatory pathways in autoimmunity. Arthritis research therapy. 2008;10(Suppl 1):3.

11. Sanmamed MF, Chen L. A Paradigm Shift in Cancer Immunotherapy: From Enhancement to Normalization. Cell. 2019;176:677.

12. Fife BT, Bluestone JA. Control of peripheral T-cell tolerance and autoimmunity via the CTLA-4 and PD-1 pathways. Immunological reviews. 2008;224:166–82.

13. Janakiram M, Chinai JM, Zhao A, et al. HHLA2 and TMIGD2: new immunotherapeutic targets of the B7 and CD28 families. Oncoimmunology. 2015;4:e1026534.

14. Zhang C, Zhang Z, Li F, et al. Large-scale analysis reveals the specific clinical and immune features of B7-H3 in glioma. Oncoimmunology. 2018;7:e1461304.

15. Ward-Kavanagh LK, Lin WW, Sedy JR, et al. The TNF Receptor Superfamily in Co-
stimulating and Co-inhibitory Responses. Immunity. 2016;44:1005–19.

16. Dostert C, Grusdat M, Letellier E, et al. The TNF Family of Ligands and Receptors: Communication Modules in the Immune System and Beyond. Physiological reviews. 2019;99:115–60.

17. Croft M, Benedict CA, Ware CF. Clinical targeting of the TNF and TNFR superfamilies. Nature reviews Drug discovery 2013;12: 147 – 68.

18. Schildberg FA, Klein SR, Freeman GJ, et al. Coinhibitory Pathways in the B7-CD28 Ligand-Receptor Family. Immunity. 2016;44:955–72.

19. Liu J, Lichtenberg T, Hoadley KA, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell 2018;173: 400 – 16.e11.

20. Takeuchi T, Tomida S, Yatabe Y, et al. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2006;24:1679–88.

21. Tomida S, Takeuchi T, Shimada Y, et al. Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2009;27:2793–9.

22. Hou J, Aerts J, den Hamer B, et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. PloS one. 2010;5:e10312.

23. Rousseaux S, Debernardi A, Jacquiau B, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. Science translational medicine. 2013;5:186ra66.

24. Okayama H, Kohno T, Ishii Y, et al. Identification of genes upregulated in ALK-positive
and EGFR/KRAS/ALK-negative lung adenocarcinomas. Cancer research. 2012;72:100-11.

25. Botling J, Edlund K, Lohr M, et al. Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. Clinical cancer research: an official journal of the American Association for Cancer Research. 2013;19:194-204.

26. Sato M, Larsen JE, Lee W, et al. Human lung epithelial cells progressed to malignancy through specific oncogenic manipulations. Molecular cancer research: MCR. 2013;11:638-50.

27. Der SD, Sykes J, Pintilie M, et al. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer. 2014;9:59-64.

28. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nature methods. 2015;12:453-7.

29. Charoentong P, Finotello F, Angelova M, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell reports. 2017;18:248-62.

30. Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nature medicine. 2018;24:1550-8.

31. Zheng S, Luo X, Dong C, et al. A B7-CD28 family based signature demonstrates significantly different prognoses and tumor immune landscapes in lung adenocarcinoma. International journal of cancer. 2018;143:2592-601.

32. Rody A, Holtrich U, Pusztai L, et al. T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers. Breast cancer
33. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinform. 2013;14:7.

34. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–74.

35. Turley SJ, Cremasco V, Astarita JL. Immunological hallmarks of stromal cells in the tumour microenvironment. Nature reviews Immunology. 2015;15:669–82.

36. Topalian SL, Taube JM, Pardoll DM. Neoadjuvant checkpoint blockade for cancer immunotherapy. New York: Science; 2020. p. 367.

37. Tang J, Jiang W, Liu D, et al. The comprehensive molecular landscape of the immunologic co-stimulator B7 and TNFR ligand receptor families in colorectal cancer: immunotherapeutic implications with microsatellite instability. Oncoimmunology. 2018;7:e1488566.

38. Xu W, Li Y, Yuan WW, et al. Membrane-Bound CD40L Promotes Senescence and Initiates Senescence-Associated Secretory Phenotype via NF-kappaB Activation in Lung Adenocarcinoma. Cellular physiology and biochemistry: international journal of experimental cellular physiology, biochemistry, and pharmacology 2018;48: 1793 – 803.

39. Mu CY, Qin PX, Qu QX, et al. Soluble CD40 in plasma and malignant pleural effusion with non-small cell lung cancer: A potential marker of prognosis. Chronic diseases translational medicine. 2015;1:36-41.

40. Lin S, Wu B, Lin Y, et al. Expression and Clinical Significance of Decoy Receptor 3 in Acute-on-Chronic Liver Failure. BioMed research international 2019;2019: 9145736.

41. Lin WW, Hsieh SL. Decoy receptor 3: a pleiotropic immunomodulator and biomarker for inflammatory diseases, autoimmune diseases and cancer. Biochem Pharmacol.
42. Yu KY, Kwon B, Ni J, et al. A newly identified member of tumor necrosis factor receptor superfamily (TR6) suppresses LIGHT-mediated apoptosis. J Biol Chem. 1999;274:13733–6.

43. Hsieh SL, Lin WW. Decoy receptor 3: an endogenous immunomodulator in cancer growth and inflammatory reactions. Journal of biomedical science. 2017;24:39.

44. Bossen C, Schneider P. BAFF, APRIL and their receptors: structure, function and signaling. Seminars in immunology 2006;18: 263 – 75.

45. Sun B, Wang H, Wang X, et al. A proliferation-inducing ligand: a new biomarker for non-small cell lung cancer. Experimental lung research. 2009;35:486-500.

46. Pelekanou V, Kampa M, Kafousi M, et al. Expression of TNF-superfamily members BAFF and APRIL in breast cancer: immunohistochemical study in 52 invasive ductal breast carcinomas. BMC Cancer. 2008;8:76.

47. Bojarska-Junak A, Hus I, Chocholska S, et al. BAFF and APRIL expression in B-cell chronic lymphocytic leukemia: correlation with biological and clinical features. Leukemia research. 2009;33:1319-27.

48. Gross JA, Dillon SR, Mudri S, et al. TACI-Ig neutralizes molecules critical for B cell development and autoimmune disease. impaired B cell maturation in mice lacking BLyS. Immunity. 2001;15:289-302.

49. Qin H, Wei G, Sakamaki I, et al. Novel BAFF-Receptor Antibody to Natively Folded Recombinant Protein Eliminates Drug-Resistant Human B-cell Malignancies In Vivo. Clinical cancer research: an official journal of the American Association for Cancer Research. 2018;24:1114-23.

50. Eby MT, Jasmin A, Kumar A, et al. TAJ, a novel member of the tumor necrosis factor receptor family, activates the c-Jun N-terminal kinase pathway and mediates
caspase-independent cell death. J Biol Chem. 2000;275:15336–42.

51. Shao L, Zuo X, Yang Y, et al. The inherited variations of a p53-responsive enhancer in 13q12.12 confer lung cancer risk by attenuating TNFRSF19 expression. Genome biology. 2019;20:103.

52. Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. 348. New York: Science; 2015. pp. 124–8.

53. Gong Z, Jia Q, Chen J, et al. Impaired Cytolytic Activity and Loss of Clonal Neoantigens in Elderly Patients With Lung Adenocarcinoma. Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer. 2019;14:857–66.

54. Herbst RS, Soria JC, Kowanetz M, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. Nature. 2014;515:563–7.

Supplemental Figure Legends

**Supplementary Figure 1. The relationship between the costimulatory molecules.** Red represents a positive correlation, and blue represents a negative correlation.

**Supplementary Figure 2. Validation of the prognostic performance of CMS in different clinical subgroups.** Kaplan-Meier survival analysis showed that high-risk patients conferred a worse prognosis in male (A), female (B), older (C), younger (D), smokers (E), and non-smokers (F) patients based on risk score in LUAD population.

**Supplementary Figure 3. Validation of the prognostic performance of CMS with different mutation statuses.** Kaplan-Meier survival analysis for patients carrying EGFR-WT (A), EGFR-MUT (B), KRAS-WT (C), KRAS-MUT (D), and EGFR/KRAS-WT (E) based on risk score in LUAD.
Supplementary Figure 4. Validation of the prognostic performance of CMS in different molecular subtypes. The CMS score allowed the segmentation of patients into high- and low-risk groups in bronchiod (A), magnoid (B), and squamiod (C) subtypes.

Figures

Identification of the CMS in the TCGA dataset. A, the distribution of risk score, survival status, and the five-gene expression panel. Kaplan-Meier curves were conducted to estimate overall survival for the high- and low-risk groups based on the risk score; B, total patients with LUAD C; patients with early-stage (stage I and II) LUAD. D; patients with advanced-stage (stage III and IV) LUAD.
Figure 2

Identification of the CMS in the TCGA dataset. A, the distribution of risk score, survival status, and the five-gene expression panel. Kaplan-Meier curves were conducted to estimate overall survival for the high- and low-risk groups based on the risk score; B, total patients with LUAD C; patients with early-stage (stage I and II) LUAD. D; patients with advanced-stage (stage III and IV) LUAD.
Identification of the CMS in the TCGA dataset. A, the distribution of risk score, survival status, and the five-gene expression panel. Kaplan-Meier curves were conducted to estimate overall survival for the high- and low-risk groups based on the risk score; B, total patients with LUAD C; patients with early-stage (stage I and II) LUAD. D; patients with advanced-stage (stage III and IV) LUAD.
Figure 4

The association between CMS and overall survival in nine different validation cohorts.

Kaplan-Meier curves were created to estimate overall survival for in high- and low-risk groups based on the risk score. A, GSE11969; B, GSE13213; C, GSE19188; D, GSE30219; E, GSE31210; F, GSE37745; G, GSE41271; H, GSE50081; I, an independent cohort with qPCR data.
Figure 5

The association between CMS and overall survival in nine different validation cohorts. Kaplan-Meier curves were created to estimate overall survival for in high- and low-risk groups based on the risk score. A, GSE11969; B, GSE13213; C, GSE19188; D, GSE30219; E, GSE31210; F, GSE37745; G, GSE41271; H, GSE50081; I, an independent cohort with qPCR data.
The association between CMS and overall survival in nine different validation cohorts.

Kaplan-Meier curves were created to estimate overall survival for in high- and low-risk groups based on the risk score. A, GSE11969; B, GSE13213; C, GSE19188; D, GSE30219; E, GSE31210; F, GSE37745; G, GSE41271; H, GSE50081; I, an independent cohort with qPCR data.
Compare CMS with previous costimulatory molecules signature. A, a meta-analysis was performed using the prognostic results of CMS in nine public datasets. B, a meta-analysis was performed using the prognostic results of the B7-CD28 signature in nine public datasets.

**Figure 7**

| Groups   | CMS         | Hazard Ratio (95% CI) | P value |
|----------|-------------|-----------------------|---------|
| TCGA     | 2.04 (1.48–2.82) | 0                     |
| GSE11969 | 1.68 (0.9–3.15)  | 0.101                 |
| GSE13213 | 2.6 (1.35–4.99)  | 0.003                 |
| GSE19188 | 2.48 (1.06–5.83) | 0.031                 |
| GSE30219 | 2.3 (1.15–4.58)  | 0.016                 |
| GSE31210 | 2.2 (1.1–4.43)   | 0.023                 |
| GSE37745 | 2.3 (0.88–2.92)  | 0.119                 |
| GSE41271 | 2.3 (1.43–3.72)  | 0.007                 |
| GSE50081 | 2.3 (1.24–4.25)  | 0.007                 |
| Overall  | 2.11 (1.75–2.53) | 0                     |

**Figure 8**

Compare CMS with previous costimulatory molecules signature. A, a meta-analysis was performed using the prognostic results of CMS in nine public datasets. B, a meta-analysis was performed using the prognostic results of the B7-CD28 signature in nine public datasets.

| Groups   | B7-CD28 signature | Hazard Ratio (95% CI) | P value |
|----------|-------------------|-----------------------|---------|
| TCGA     | 1.86 (1.38–2.51)  | 0                     |
| GSE11969 | 0.38 (0.12–1.23)  | 0.092                 |
| GSE13213 | 1.68 (0.95–2.99)  | 0.073                 |
| GSE19188 | 0.35 (0.14–0.90)  | 0.023                 |
| GSE30219 | 2.33 (1.25–4.32)  | 0.006                 |
| GSE31210 | 3.12 (1.60–6.11)  | 0                     |
| GSE37745 | 1.54 (0.96–2.46)  | 0.071                 |
| GSE41271 | 1.57 (0.96–2.56)  | 0.069                 |
| GSE50081 | 2.30 (1.30–4.07)  | 0.003                 |
| Overall  | 1.58 (1.14–2.18)  | 0.005                 |
Figure 9

Compare CMS with previous costimulatory molecules signature. A, a meta-analysis was performed using the prognostic results of CMS in nine public datasets. B, a meta-analysis was performed using the prognostic results of the B7-CD28 signature in nine public datasets.
Figure 10

CMS-related biological pathways. A, the most related genes of TNF family-based signature in patients with LUAD (Top 600 genes ranked by Pearson |R|). B and C, GO and KEGG analyses of the related genes.
Figure 11

CMS-related biological pathways. A, the most related genes of TNF family-based signature in patients with LUAD (Top 600 genes ranked by Pearson |R|). B and C, GO and KEGG analyses of the related genes.
Figure 12

CMS-related biological pathways. A, the most related genes of TNF family-based signature in patients with LUAD (Top 600 genes ranked by Pearson |R|). B and C, GO and KEGG analyses of the related genes.
Figure 13

CMS-related immune cell infiltration and inflammatory activities. A, the relative proportion of immune cell expression in high- and low-risk patients. B and C, differentially expression immune cells in high- and low-risk patients. D, the details of seven inflammatory metagenes and risk score. E, correlogram of risk score, and inflammatory metagenes.
Figure 14

CMS-related immune cell infiltration and inflammatory activities. A, the relative proportion of immune cell expression in high- and low-risk patients. B and C, differentially expression immune cells in high- and low-risk patients. D, the details of seven inflammatory metagenes and risk score. E, correlogram of risk score, and inflammatory metagenes.
Figure 15

CMS-related immune cell infiltration and inflammatory activities. A, the relative proportion of immune cell expression in high- and low-risk patients. B and C, differentially expression immune cells in high- and low-risk patients. D, the details of seven inflammatory metagenes and risk score. E, correlogram of risk score, and inflammatory metagenes.
The expression pattern of immunotherapy response makers in high- and low-risk groups. The distribution of TMB (A), number of neoantigens (B), number of clonal neoantigens (C), number of subclonal neoantigens (D), protein level of PD-L1 (E), TIDE score (F), T cell dysfunction score (G) and T cell exclusion score (H) in high- and low-risk groups.
Figure 17

The expression pattern of immunotherapy response makers in high- and low-risk groups.

The distribution of TMB (A), number of neoantigens (B), number of clonal neoantigens (C), number of subclonal neoantigens (D), protein level of PD-L1 (E), TIDE score (F), T cell dysfunction score (G) and T cell exclusion score (H) in high- and low-risk groups.
The expression pattern of immunotherapy response makers in high- and low-risk groups.

The distribution of TMB (A), number of neoantigens (B), number of clonal neoantigens (C), number of subclonal neoantigens (D), protein level of PD-L1 (E), TIDE score (F), T cell dysfunction score (G) and T cell exclusion score (H) in high- and low-risk groups.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

fig s4.tif
fig s3.tif
Supplementary Table 2 5 genes in prognostic model.pdf
fig s2.tif
fig s1.tif
Supplementary Table 1 Primer Sequences for qRT-PCR.pdf
fig s3.tif
fig s3.tif
fig s4.tif
fig s1.tif
fig s4.tif
Supplementary Table 2 5 genes in prognostic model.pdf
fig s2.tif
fig s1.tif
fig s2.tif
Supplementary Table 1 Primer Sequences for qRT-PCR.pdf
