Association of *PDCD1* and *CTLA-4* Gene Expression with Clinicopathological Factors and Survival in Non–Small-Cell Lung Cancer

Results from a Large and Pooled Microarray Database

Lei Deng,† Balazs Gyorffy, MD, PhD, MSc,‡§‖ Feifei Na, MD,*† Baoqing Chen,*† Jie Lan, MD,* Jianxin Xue, MD, PhD,* Lin Zhou, MD,* and You Lu, MD**

Introduction: Immune checkpoint blockade is being investigated in clinical trials and showed great potential in lung cancer. The prognostic roles of and clinicopathological factors associated with immune checkpoint gene expression, *CTLA-4* and *PDCD1* remain largely undefined, which encodes cytotoxic-lymphocyte antigen 4 (CTLA-4) and programmed cell death 1 (PD-1), respectively.

Methods: We used a lung cancer database of 1715 patients measured by Affymetrix microarrays to analyze the association of gene expression with clinicopathological factors and survival. Hazard ratio (HR) and 95% confidence interval (CI) for overall survival (OS) were calculated. Cutoffs were determined by median across the entire database.

Results: In 909 patients with histology information, significantly higher *PDCD1* and *CTLA-4* expression were found in squamous carcinoma than adenocarcinoma. In 848 patients with known smoking history, current/former smokers were found to have significantly elevated gene expression compared with nonsmokers. Significant higher expression of both genes were found in TNM stage II versus I. Higher expression of *PDCD1* predicted worse OS in univariate analysis, but not in multivariate (HR: 1.22; 95% CI: 0.53–2.79). *CTLA-4* was marginally significant in univariate analysis of the entire set (HR: 1.15; 95% CI: 0.99–1.34). In patients with information for multivariate analysis, higher expression of *CTLA-4* was associated with worse OS (HR: 1.96; 95% CI: 1.18–3.31).

Conclusions: In this study with large number of patients, *PDCD1* and *CTLA-4* expression is significantly higher in squamous carcinoma and current/former smokers. Higher expression of *CTLA-4*, but not *PDCD1* predicts worse survival.

Key Words: Non–small-cell lung cancer, Immune checkpoint blockade, *PDCD1*, *CTLA-4*, Smoking, Squamous carcinoma, Adenocarcinoma, Prognosis, Microarray

(*J Thorac Oncol.* 2015;10: 1020–1026)

Non–small-cell lung cancer (NSCLC) is one of the leading cancer death causes in the world. Despite recent progress and newly approved drugs, prognosis of advanced stage patients undergoing systemic therapy are still suboptimal and new treatment option is needed. Recent clinical trials with immune checkpoint blockade have shown promising results.[1] A subgroup of patients has been reported to have achieved durable long-term response, suggesting the potential of such paradigm.

To avoid autoimmune reaction to normal tissues, human body utilizes the immune checkpoint to balance the pro- and anti-immune reactions.[2] Cancer cells have been known to hijack the immunosuppressive mechanism to evade the attack from immune system.[3] Among the checkpoint pathways, cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and programmed cell death 1 (PD-1) pathways are mostly studied. They are expressed on T cells and prevent the activation of cytotoxic T cells to function, with or without the interaction with its ligands.[4] Blocking these two pathways tip the balance from tumor immune tolerance to immune activation.

It has been reported that CTLA-4, PD-1, and PD-ligand 1 (PD-L1) were overexpressed in NSCLC, and correlated with poor survival in NSCLC patients. However, those studies mostly used immunohistochemistry (IHC) and enrolled small number of patients.[5–11] Antibodies and cutoffs varied greatly as well. Larger studies are required to define the prognostic roles.

Early phases of clinical trials have shown promising results of CTLA-4 and PD-1/PD-L1 pathway blockade, and a subgroup of patients seemed to respond better to the blockade. Pretreatment patient selection for immune checkpoint blockade is being advocated. In a randomized phase II trial, squamous NSCLC than nonsquamous are more responsive to

*Department of Thoracic Oncology, Cancer Center and State Key Laboratory of Biotherapy, †Huaxi Student Society of Oncology Research, West China Hospital, West China School of Medicine, Sichuan University, Chengdu, China; ‡MTA TTK Lendület Cancer Biomarker Research Group, §2nd Department of Pediatrics, Semmelweis University, and ‖MTA-SE Pediatrics and Nephrology Research Group, Budapest, Hungary.

Disclosure: The authors declare no conflict of interest.

B.G. was supported by the OTKA K108655 Grant. Y.L. was supported by National Natural Science Foundation of China No. 81172131.

Address for correspondence: You Lu, MD, Department of Thoracic Oncology, Cancer Center and State Key Laboratory of Biotherapy, West China Hospital, West China School of Medicine, Sichuan University, 37 Guoxue Lane, Chengdu, People's Republic of China. E-mail: radyoulu@hotmail.com

DOI: 10.1097/JTO.0000000000000550

ISSN: 1556-0864/15/1007-1020

ipilimumab, a CTLA-4 antibody.[12] Gettinger et al.[13] reported that patients with positive IHC PD-L1 staining have higher objective response rate to nivolumab (a PD-1 antibody), than those with negative staining. Both nivolumab[14] and MPDL3280A[15] (a PD-L1 antibody) have been reported to be more active in current/former smokers than in never-smokers.

To explore the clinicopathological factors associated with *CTLA-4* and *PDCD-1* gene expression, which encodes CTLA-4, and PD-1, respectively, and the prognostic roles in NSCLC, we utilized a large microarray database,[16] which integrated data from the Cancer Biomedical Informatics Grid (caBIG), the Gene Expression Omnibus (GEO), and the Cancer Genome Atlas for lung cancers (TCGA).

MATERIALS AND METHODS

Online Microarray Database

Details of how the database (www.kmplot.com/lung) was constructed were described previously.[16] In brief, data were extracted and integrated from caArray, GEO, and TCGA. Only tumor tissue was analyzed in all the original datasets. Clinical data from the three datasets including age, sex, smoking history, histology, stage, grade, and survival, etc. were inputted into the online database. The smoking history was directly obtained from the datasets and was categorized into never-smokers and former/current smokers.

Datasets included in the online database are as follows: GSE4573[17], GSE14814[18], GSE8894[19], GSE19188[20], GSE3141[21], GSE3120[22], caArray,[23] TCGA,[24] GSE29013[25], and GSE37745[26]. Detailed treatment information for each patient is not available.

Data Retrieval from the Online Database

Data retrieval was performed on August 2014. Gene expression cutoff value was chosen as median over entire dataset to ensure all analyses of each gene was based on the same cutoff value. All available follow-up data were used. Biased arrays were excluded. Biased arrays were defined as those having two or more of following parameters out of the 95% range of all arrays, percentage of present calls, raw *Q* value, presence of bioB-/C-/D-spikes, GAPDH, and ACTB 3:5 ratio. Not all clinicopathological data were available in each patient, but all available data were reported.

Statistical Analysis

Gene expression level was compared using Mann–Whitney *U* test (two cohorts comparison) or Kruskal–Wallis test (two or more cohorts comparison). The online database identified that histology, stage, sex, and smoking history are associated with overall survival (OS) by univariate analysis, but the above information was not available in all patients. Multivariate analysis of survival was carried out when gene expression is significantly associated with survival in patients with all the above-mentioned clinicopathological factors available (Uni in Multi in Table 2). All multivariate analysis was performed including histology, stage, sex, and smoking history by the online tool using Cox regression Kaplan–Meier survival curve was plotted and analyzed by logrank test.

RESULTS

Patient Characteristics and Gene Expression

Of 1715 patients in the database, 1432 patients had OS data. Median OS was 40 months. Median follow-up was 37 months. Clinicopathological data were not available for all patients. Altogether, 556 and 500 patients were adenocarcinoma (Adeno) and squamous cell carcinoma (Squa), respectively; 441, 186, 67, and four patients were stage I, II, III, and IV; 634 were woman, whereas 908 were men; 187 patients were never-smokers, in contrast to 689 current/previous smokers.

For *PDCD1*, the median expression value was 65, with a range of 2 to 760 across the entire database (Fig. 1A). For *CTLA-4*, the median value was 95, with a range of 2 to 689 (Fig. 1B). Interestingly, median *PDCD1* expression value was 36 in adenocarcinoma, compared with 78 in squamous carcinoma ($p < 0.01$; Fig. 2A; Table 1). Similarly, the median *CTLA-4* was also higher in squamous cancer patients compared with adenocarcinoma (median: 63 vs. 111; $p < 0.01$; Fig. 2A; Table 1).

Current/previous smokers had much higher *PDCD1* and *CTLA-4* expression compared with never-smokers (*PDCD1* median 142 vs. 36, $p < 0.01$; *CTLA-4* median 152 vs. 59, $p < 0.01$, Fig. 2B).

Gene expression in TNM stages I, II, and III patients was compared in 693 patients. Stage IV patients were excluded because only four stage IV patients were in the database. *PDCD1* expression was progressively higher as TNM stage rises. The median was 43, 73, and 90 for stages I, II, and III patients, respectively ($p < 0.01$; Fig.2C). Similar trend was also observed in *CTLA-4*. The median expression value was 80, 107, and 115 in stages I, II, and III, respectively ($p < 0.01$; Fig. 2C). The statistical significance was not found between stages II and III patients, likely due to insufficient sample size.

No difference of gene expression was found in different sex. The median of *PDCD1* was 95 and 99 for woman and men, respectively. For *CTLA-4*, the median was 110 and 108 (Fig. 2D).

Association of Gene Expression with Survival

Using the median as cutoff value, the number of patients with lower *PDCD1* expression was 629, and the number with higher was 803 (Table 2). Higher expression of *PDCD1* was significantly correlated with worse OS (hazard ratio [HR]: 1.29, 95% confidence interval [CI]: 1.22–1.50) in univariate analysis (Table 2; Fig. 1C). Less data were available for multivariate analysis with a patient number of 412 patients, and the HR was 1.22 (95% CI: 0.53–2.79). For 412 patients, the univariate HR was 1.29 (95% CI: 1.11–1.5; Table 2).

Using the median as cutoff value, the number of patients with lower *CTLA-4* expression was 646, and that with higher expression was 786 (Table 2; Fig. 2D). Univariate analysis of the entire patient cohort suggested a trend of correlation with OS (HR: 1.15, 95% CI: 0.99–1.34). However, in 412 patients with available data for multivariate analysis, the univariate analysis showed significant correlation with OS (HR: 3.46,

