Prediction of immune checkpoint inhibition with immune oncology-related gene expression in gastrointestinal cancer using a machine learning classifier

Zhihao Lu,1 Huan Chen,2 Xi Jiao,1 Wei Zhou,2 Wenbo Han,2 Shuang Li,1 Chang Liu,1 Jifang Gong,1 Jian Li,1 Xiaotian Zhang,1 Xicheng Wang,1 Zhi Peng,3 Changsong Qi,1 Zhenghang Wang,1 Yanyan Li,1 Jie Li,1 Yan Li,1 Malcolm Brock,3 Henghui Zhang,4 Lin Shen1

ABSTRACT

Immune checkpoint inhibitors (ICIs) have revolutionized the therapeutic landscape of gastrointestinal cancer. However, biomarkers correlated with the efficacy of ICIs in gastrointestinal cancer are still lacking. In this study, we performed 395-plex immune oncology (IO)-related gene target sequencing in tumor samples from 96 patients with metastatic gastrointestinal cancer patients treated with ICIs, and a linear support vector machine learning strategy was applied to construct a predictive model. Results All 96 patients were randomly assigned into the discovery (n=72) and validation (n=24) cohorts. A 24-gene RNA signature (termed the IO-score) was constructed from 395 immune-related gene expression profiling using a machine learning strategy to identify patients who might benefit from ICIs. The durable clinical benefit rate was higher in patients with a high IO-score than in patients with a low IO-score (discovery cohort: 92.0% vs 43.4%, p<0.001; validation cohort: 85.7% vs 17.6%, p=0.004). The IO-score may exhibit a higher predictive value in the discovery (area under the receiver operating characteristic curve (AUC)=0.97) and validation (AUC=0.74) cohorts compared with the programmed death ligand 1 positivity (AUC=0.52), tumor mutational burden (AUC=0.69) and microsatellite instability status (AUC=0.59) in the combined cohort. Moreover, patients with a high IO-score also exhibited a prolonged overall survival compared with patients with a low IO-score (discovery cohort: HR, 0.29; 95% CI 0.15 to 0.56; p=0.003; validation cohort: HR, 0.32; 95% CI 0.10 to 1.05; p=0.04). Taken together, our results indicated the potential of IO-score as a biomarker for immunotherapy in patients with gastrointestinal cancers.

INTRODUCTION

Immune checkpoint inhibitors (ICIs) have made remarkable achievements in the treatment of cancer, including gastrointestinal (GI) cancers. However, the objective response rate of patients with GI cancers to ICIs was only 10%–20%, underscoring the need for effective biomarker-based patient selection.1 To date, the predictive biomarkers used to guide patient selection mainly fall into two categories: those reflecting tumor intrinsic properties, such as the microsatellite instability (MSI)/mismatch repair (MMR) status and a high tumor mutational burden (TMB) and those associated with the tumor microenvironment (TME), such as programmed death ligand 1 (PD-L1) expression and immune-related RNA expression.2–4 However, previous studies have reported conflicting results regarding the association of those biomarkers with ICI efficacy in GI cancers,4–7 emphasizing the urgent need to explore new optimal predictors. Here, we applied a 395-plex immune oncology (IO)-related gene profiling platform and machine learning strategy to determine a novel RNA signature that may reflect the ‘immune-responsive feature’ of both cancer cells and the immune microenvironment.

METHODS

Patients

Patients with metastatic GI cancers who failed standard therapy and had received anti-programmed cell death 1 (PD-1)/PD-L1 therapy or in combination with a cytotoxic T lymphocyte-associated protein 4 (CTLA-4) inhibitor between January 2016 and January 2018 were recruited. The tumor response was measured by imaging studies or physical examinations according to the Response Evaluation Criteria in Solid Tumors (RECIST) V1.1 and iRECIST. Durable clinical benefit (DCB) was defined
as complete response, partial response, or stable disease (SD) lasting for ≥24 weeks, and no durable benefit (NDB) was defined as progressive disease or SD that lasted for <24 weeks.8

**Gene expression profiling**

Total RNA was isolated from 5 µm thick pretreatment archival formalin-fixed, paraffin-embedded (FFPE) sections of tumors. Gene expression profiling was performed using a panel sequencing platform that determines the expression of 395 human genes belonging to the following functional categories:9 immunological response to immunotherapy, infiltrating immune cell markers, tumor-specific antigens, tumor markers and essential signaling pathways. Detailed information on the RNA profiling method and the data processing pipeline is provided in the online supplementary methods.

**Construction of a linear support vector machine classifier**

In the discovery cohort, the linear support vector machine (linear SVM) model was applied to distinguish 25 individuals with DCB from 47 patients with NDB based on the 395-gene RNA profiling data (online supplementary methods). Briefly, after data standardization and normalization, recursive feature elimination with cross-validation was performed to select features. We used the sigmoid function to scale the distance to the SVM classifier boundary into the range [0–1], which was designated as the IO-score. The prediction of the independent validation cohort was generated with the training test set procedure. The area under the receiver operating characteristic curve (AUC) was then calculated from the IO-score and the binary group labels (DCB and NDB). We used the Youden index as the cut-off point (high IO-score >0.52) to stratify patients with different prognoses and DCB rates.

**Evaluation of TMB**

We extracted genomic DNA from FFPE specimens and matched white blood cell samples using the blackPREP FFPE DNA Kit (Analytik Jena AG, Jena, Germany) and the Tiangen Whole Blood DNA Kit (Tiangen, Beijing, PRC) according to the manufacturer’s instructions. Whole exome sequencing was then performed to detect genomic alterations. The TMB was determined by analyzing the number of somatic mutations per megabase (mb). The cut-off value for TMB-high and TMB-low was defined as 5 mutations/mb as previously reported.10

**Immunohistochemical staining for PD-L1**

PD-L1 status was assessed by immunohistochemical (IHC)-stained archival FFPE sections using an anti-PD-L1 (rabbit, clone SP142, 1:100; Spring Bioscience, California, USA) antibody. PD-L1 expression was evaluated in tumor cells and tumor-infiltrating immune cells. PD-L1 positivity was defined as ≥1% of the tumor/stromal cell membrane staining.

**Assessing the MSI/MMR status**

To determine the MMR status, IHC was performed on archival FFPE sections using monoclonal anti-mutL homolog 1, anti-mutS homolog 2, anti-mutS homolog 6 and PMS1 homolog 2. Tumors lacking expression of any one of the four proteins were considered deficient MMR (dMMR); otherwise, they were considered proficient MMR (pMMR). In some cases, to determine MSI status, PCR-based molecular testing was employed, which assesses five microsatellite loci comprising BAT-25, BAT-26, D2S123, D5S346, and D17S250.11–13 MSI-high (MSI-H) tumors were defined as instability at two or more of these markers, whereas MSI-low and microsatellite stability (MSS) were defined as instability at a single locus and no instability at any locus, respectively.13

**Statistical analysis**

Statistical tests included the χ² test, Student’s t-test and the Mann-Whitney U test. A receiver operating characteristic curve was generated to evaluate the diagnostic accuracy. The AUC was calculated to measure the discriminatory ability of the potential biomarkers. Progression-free survival (PFS) and overall survival (OS) were analyzed with the Kaplan-Meier method and log-rank test.

**RESULTS**

**Patient characteristics**

A total of 96 patients with metastatic GI cancers were randomly assigned into a discovery cohort (n=72) and a validation cohort (n=24; figure 1A). Clinical characteristics of the subjects are summarized in online supplementary table S1. All patients had received anti-PD-1/anti-PD-L1 therapy or in combination with anti-CTLA4 antibodies. The clinical end points include OS and PFS, and the therapeutic response.

**Model construction and performance evaluation of the linear SVM-derived IO-score**

Based on the expression levels of 395-genes in patients with DCB and NDB in the discovery cohort, a linear SVM classification algorithm was adopted to construct a prediction model (figure 1A and online supplementary figure S1). The best-performing features and hyperparameters were determined by 10-fold cross-validation within the discovery cohort. Intriguingly, a linear SVM classifier composed of 24 IO-related genes was developed, and the IO-score of each sample was subsequently derived (figure 1B and online supplementary table S2).

The performance of the IO-score was comprehensively evaluated, and overall accuracies of 94% and 83% were achieved for discriminating DCB and NDB in the discovery (AUC=0.97; 95% CI 0.93 to 1.00) and validation (AUC=0.74; 95% CI 0.51 to 0.97) cohorts, respectively (figure 2A and online supplementary table S3). Moreover, a higher DCB rate was observed in patients with a high IO-score than in patients with a low IO-score (discovery cohort: 92.0% (23 of 25) vs 4.3% (2 of 47),...

p<0.001; validation cohort: 85.7% (6 of 7) vs 17.6% (3 of 17), p=0.004, figure 2B). Furthermore, a favorable prognosis was identified in the high IO-score subgroup when compared with that in the low IO-score subgroup (figure 2C–D).

**Predictive and prognostic values of the TMB, PD-L1 positivity and MSI/MMR status**

We then evaluated the predictive and prognostic values of other candidate biomarkers in the combined cohort (online supplementary table S4 and online supplementary figures S2-S4), with AUCs of 0.69 (95% CI 0.57 to 0.80), 0.59 (95% CI 0.45 to 0.73), and 0.52 (95% CI 0.37 to 0.67) for TMB, MSI/MMR, and PD-L1 positivity, respectively. As expected, a better prognosis was also identified in the TMB-high and MSI-H/dMMR subgroups than in the TMB-low and MSS/pMMR subgroups (online supplementary figures S2 and S3). However, the IO-score tended to have greater power for patient stratification, as revealed by the high odds ratio (OR) for DCB (figure 3 and online supplementary table)

**Figure 1** Flow diagram of the machine learning strategy and IO-score composition. (A), Flow diagram of the construction of the linear SVM classifier construction. (B), Feature importance of the linear SVM-classifier-derived 24-gene RNA signature, namely, the IO-score. DCB, durable clinical benefit; ICI, immune checkpoint inhibitor; IO, immune oncology; NDB, no durable benefit; nRPM, normalized reads per million; SVM, support vector machine.

**Figure 2** Predictive and prognostic value of the IO-score in the discovery and validation cohorts. (A) ROC curve of the IO-score in predicting the clinical benefit. (B) Comparison of the DCB rates between the IO-score-high and IO-score-low groups. (C) and (D) Kaplan-Meier curves comparing PFS (C) and OS (D) between the IO-score-high and IO-score-low subgroups. AUC, area under the receiver operating characteristic curve; DCB, durable clinical benefit; IO, immune oncology; OS, overall survival; PFS, progression-free survival; ROC, receiver operating characteristic.
S4) and the low HR of the survival analysis (figure 2C and D).

Notably, the waterfall plot of maximal tumor reduction in the two independent cohorts also sustained the superiority of the IO-score predictive model based on the IO-score (figure 4A and B). Furthermore, we analyzed the relationship between the IO-score and other potential biomarkers in the subgroup with a complete data set (n=52). Higher percentages of MSI-H and/or TMB-high patients were identified in the high IO-score subgroup than in the low IO-score subgroup (online supplementary figure S5).

**DISCUSSION**

Based on our analysis, a linear SVM-derived RNA signature, the IO-score, may be a superior prognostic and predictive biomarker for GI cancer patients treated with ICIs. In previous studies, gene expression profile (GEP) signatures, such as the interferon-γ (IFN-γ) RNA signature and T-cell inflamed GEP, predict the efficacy of an anti-PD-1 treatment in patients with various tumor types, including head and neck squamous cell carcinoma and melanoma.2 3 However, data exploring the correlation between RNA signatures and GI cancers, particularly esophageal cancer and colorectal cancers, are still limited.14 Moreover, a prospective evaluation of the IFN-γ signature was performed in patients with gastric cancer,15 and the results showed that it failed to guide patient selection, possibly due to the great tumor heterogeneity and complexity of the TME.

Intriguingly, machine learning methodologies offer a novel approach for measuring the interactions between the tumor and the TME. Here, we used a state-of-the-art classification linear SVM strategy to construct a predictive model based on the IO-score. This novel RNA signature, the IO-score, comprises genes encoding tumor antigens, tumor suppressors/oncogenes, lymphocyte markers, interferons and checkpoint signaling pathways (online supplementary table S2 and online supplementary figure S6). This model may comprehensively capture the intrinsic pathological features of both tumor cells and infiltrating immune cells, which may partially explain the superiority of the IO-score compared with other predictors.

Conventionally, for clinical classification problems with a binary response, a logistic regression (LR) analysis is one of the most widely used models. Recently,
SVM became a relatively new alternative based on the principle of statistical learning theory to solve classification and regression problems. In terms of the algorithm, SVM has some unique advantages. For instance, the introduction of kernel function greatly simplifies the complexity of computation. SVMs also usually require fewer variables to achieve the same misclassification rate than LR, and it is possible to manage classification problem when the sample size is limited.

Several limitations to the current study must be acknowledged. First, this was a retrospective study, and the predictive model may require further validation in a prospective study. Second, a larger cohort study would be helpful to optimize the cut-off value of the IO-score.

CONCLUSIONS
In summary, the IO-score, a machine learning-based RNA signature, effectively identifies the ‘immune responsive features’ of patients with GI cancers. The predictive model based on the IO-score exhibited superior predictive and prognostic value in both the discovery and validation cohorts and may help facilitate the individualized management of immunotherapy in patients with metastatic GI cancers.

REFERENCES