IMPORTANCE Therapies to inhibit programmed cell death 1 and its ligand (anti–PD-1/PD-L1) provide significant survival benefits in many cancers, but the efficacy of these treatments varies considerably across different cancer types. Identifying the underlying variables associated with this cancer type-specific response remains an important open research challenge.

OBJECTIVE To evaluate systematically a multitude of neoantigen-, checkpoint-, and immune response–related variables to determine the key variables that accurately predict the response to anti–PD-1/PD-L1 therapy across different cancer types.

DESIGN, SETTING, AND PARTICIPANTS This analysis of a broad range of data used whole-exome and RNA sequencing of 7187 patients from the publicly available Cancer Genome Atlas and the objective response rate (ORR) data of 21 cancer types obtained from a collection of clinical trials. Thirty-six variables of 3 distinct classes considered were associated with (1) tumor neoantigens, (2) tumor microenvironment and inflammation, and (3) the checkpoint targets. The performance of each class of variables and their combinations in predicting the ORR to anti–PD-1/PD-L1 therapy was evaluated. Accuracy of predictions was quantified with Spearman correlation measured using a standard leave-one-out cross-validation, a statistical method of evaluating a statistical model by dividing data into 2 segments: one to train the model and the other to validate the model. Data were collected from October 19 through 31, 2018, and were analyzed from November 1 through December 14, 2018.

MAIN OUTCOMES AND MEASURES Response to anti-PD-1/PD-1 therapy.

RESULTS Among the 36 variables, estimated CD8+ T-cell abundance was the most predictive of the response to anti–PD-1/PD-L1 therapy across cancer types (Spearman $R = 0.72; P < 2.3 \times 10^{-4}$), followed by the tumor mutational burden (Spearman $R = 0.68; P < 6.2 \times 10^{-4}$), and the fraction of samples with high PD1 gene expression (Spearman $R = 0.68; P < 6.9 \times 10^{-8}$). Notably these top 3 variables cover the 3 classes considered, and their combination is highly correlated with response (Spearman $R = 0.90; P < 4.1 \times 10^{-8}$), explaining more than 80% of the ORR variance observed across different tumor types.

CONCLUSIONS AND RELEVANCE That we know of, this is the first systematic evaluation of the different variables associated with anti–PD-1/PD-L1 therapy response across different tumor types. The findings suggest that the 3 key variables can explain most of the observed cross-cancer response variability, but their relative explanatory roles may vary in specific cancer types.
Targeting programmed death 1 (PD-1) or its ligand (PD-L1) offers significant clinical benefits in many cancers. One emerging biomarker of response to anti-PD-1/PD-L1 therapy is the tumor mutational burden,\(^1,2\) which quantifies the number of somatic mutations in the tumor. The mutational burden has recently been shown by Yarchoan et al\(^3\) to strongly correlate with the response rate for anti-PD-1/ PD-L1 therapies across multiple cancer types. However, tumors containing comparably high mutational burden may exhibit variable responses,\(^4-6\) suggesting that additional factors may contribute to anti-PD-1/PD-L1 response. Herein we systematically study this hypothesis, aiming to identify the most important additional factor(s) that contribute to response prediction. To this end we analyze whole-exome sequencing and RNA sequencing data of pretreated samples from 21 different cancer types in the Cancer Genome Atlas (TCGA). We find that a trivariate model composed of tumor mutational burden, estimated CD8\(^+\) T-cell abundance (eCD8T), and fraction of high PD-1 messenger RNA expression samples (fPD1) markedly enhances the prediction of the response to anti-PD-1/PD-L1 therapy across different cancer types.

**Methods**

Data were collected from October 19 through 31, 2018, and analyzed from November 1 through December 14, 2018. We considered 36 different variables that have previously been reported to be associated with anti-PD-1/PD-L1 response within a given cancer type, of 3 distinct classes: (1) descriptors associated with tumor neoantigens, including the tumor mutational burden, intratumor heterogeneity, neoantigen burden, and hydrophobicity;\(^7\) (2) descriptors of the tumor microenvironment, including the estimated abundance of different immune cells, the cytolytic storm;\(^8\) T-cell exhaustion signatures, interferon-\(\gamma\) signature, and estimated T-cell receptor diversity; and (3) checkpoint target-related variables, including PD-L1 protein expression, the combined positive score,\(^9\) and fPD1\(^10\) (eMethods and eTable in the Supplement). All our predictions were performed using a standard leave-one-out cross-validation. This study does not require additional approval from an institutional review board because its human data is gathered from publicly available datasets (eg, TCGA) that already received such approvals.

**Statistical Analysis**

For this study, we performed a standard regression analysis with leave-one-out cross-validation in predicting objective response rate (ORR) across cancer types using the R package caret (R Foundation for Statistical Computing). The performance of the prediction was evaluated based on Spearman rank correlation (\(R\)) and unexplained variance (1-\(R^2\)). For comparing goodness of fit between different models, a log-likelihood ratio test was performed. The association between tumor mutational burden, eCD8T, and their combination vs the patient survival after anti-PD-1/PD-L1 therapy in those with melanoma was evaluated using log-rank test. Statistical significance was set at 2-sided \(P < .05\).

**Key Points**

**Question** What are the most important variables that predict the response to therapy to inhibit programmed cell death 1 and its ligand across different cancer types?

**Findings** This analysis of multiomics data from the Cancer Genome Atlas cohort and objective response rates to therapy data across 21 cancer types found that estimated CD8\(^+\) T-cell abundance is the most predictive, followed by tumor mutational burden and the fraction of samples with high programmed cell death 1 gene expression.

**Meaning** Immune, neoantigen, and checkpoint target variables are required in combination for accurately predicting the response to therapy to inhibit programmed cell death 1 and its ligand across multiple cancers.

**Results**

We studied 21 cancer types for which molecular TCGA data from 7187 patients and anti–PD-1/PD-L1 ORR from Yarchoan et al\(^3\) are available. First, we observed a high and significant correlation between mutational burden and ORR (Spearman \(R = 0.68; P < 6.2 \times 10^{-4}\) (Figure 1A). The slightly higher correlation (\(R = 0.74; P < .001\)) Yarchoan et al\(^3\) originally reported may result from the consideration of a larger number of cancer types (\(n = 27\)) available from Chalmers et al.\(^11\) Second, we found that eCD8T was the strongest positive correlate of ORR (\(R = 0.72; P < 2.3 \times 10^{-4}\)) (Figure 1B). The estimated abundance of all other immune cell types was not significantly associated with PD-1/PD-L1 response, except for M1-macrophage (\(R = 0.51; P < .02\)) and CD4\(^+\) T-cell abundance (\(R = 0.44; P < .05\)) (eFigure 1A and B in the Supplement). The associations found for other neoantigen- or immune system-related variables were weaker (Figure 1C and eFigure 1C-I in the Supplement). Notably, most of the cancer types that showed higher response rates than predicted by the mutational burden regression model had higher eCD8T levels; those showing lower than mutational burden–based predicted response rates had lower eCD8T levels (Figure 1A). Correspondingly, most of the cancer types that showed a higher response rate than predicted by an eCD8T regression model had higher mutational burden and vice versa (Figure 1B). Accordingly, the combined mutational burden and eCD8T model markedly enhanced the response prediction (\(R = 0.85; P < 1.1 \times 10^{-6}\)) (Figure 2A) with a significant log-likelihood model improvement compared with each of the univariate models (\(P < .001\)). The resulting bivariate linear regression model assigned approximately equal weights to the mutational burden and eCD8T (when \(z\) scores of both variables were considered), which were not significantly correlated each other (\(R = 0.32; P = .16\)).

As exhibited in Figure 2A, several outlier cancer types had actual response rates that still significantly deviated from the predicted values of the combined bivariate regression model. We asked whether the third class of factors, which were associated with checkpoint targets, could possibly enhance the prediction accuracy. We focused on the fPD1 (>80-percentile across
cancer types in TCGA) reported to correlate with anti–PD-1 ORR across a smaller set of cancer types (n = 17). We found that the performances of an fPD1-based model ($R = 0.70$) and the bivariate mutational burden–eCD8T model ($R = 0.71$) were comparable. In the larger data set of Yarchoan et al, however, the correlation of the fPD1 model ($R = 0.68$) was much weaker than that of the bivariate model (eFigure 1J in the Supplement). Notably, however, most cancer types that showed higher response rate than predicted by the mutational burden–eCD8T model had higher fPD1 levels (Figure 2A) and those showing lower than predicted response rates had lower fPD1 levels. Indeed, adding fPD1 to the bivariate mutational burden–eCD8T model led to a trivariate regression model with a significantly improved accuracy ($R = 0.90$; $P < 4.1 \times 10^{-8}$; log-likelihood ratio test $P < .02$) in the Yarchoan et al data set (Figure 2B). There were 2 clear outliers whose actual response rate was far from that predicted by the trivariate regression model, glioblastoma multiforme and microsatellite-stable colon cancer. The discrepancy observed for glioblastoma multiforme could be explained by its strong cytolytic score (eFigure 1F in the Supplement). Other factors in this class associated with checkpoint targets, including PD-L1 protein expression.
Some tumors are known to exhibit high response rates to anti–PD-1/PD-L1 therapy despite having low mutational burden, such as Merkel cell carcinoma, kidney renal cell carcinoma, and mesothelioma. On the other hand, some tumors such as small cell lung cancer have low response rates, despite their high mutational burden. The high response rate of kidney renal cell carcinoma and mesothelioma can be explained by the high eCD8T and fPD1 expression observed in these cancer types. Because Merkel cell carcinoma and small cell lung cancer samples were not available in TCGA, we analyzed independent cohorts of these cancer types. Although incorporating eCD8T and fPD1 levels does not fully explain the observed ORRs, its significance improved on the mutational burden-based predictions beyond a single variate mutational burden model to include the increase in predicted ORR was more prominent in polyomavirus-positive cases in which the mutational burden was low (Figure 2C) than in the polyomavirus-negative cases with a high mutational burden (eFigure 2B in the Supplement); however, better descriptors remain to be discovered in this case. As noted by Yarchoan et al, high CD8 T-cell infiltration induced by presentation of viral antigens may confer an increased response rate to anti–PD-1/PD-L1 therapy in Merkel cell carcinoma.

Finally, we evaluated the performance of our models in predicting the PD-1/PD-L1 response of individual patients with melanoma by analyzing a large data set providing whole-exome and RNA sequencing data. Because fPD1 is a cancer type-specific property that is difficult to translate to individual patients in a straightforward manner, we resorted to testing the combined association of mutational burden and eCD8T and compared its performance to those of the respective univariate models. Although both individual variables were not predictive, their combination was (eFigure 2C in the Supplement).

**Discussion**

This systematic evaluation of the different variables associated with anti–PD-1/PD-L1 therapy response across different tumor types is, to our knowledge, the first such study. Going beyond a single variate mutational burden model to include immune- and checkpoint target-related variables has enabled us to markedly reduce the unexplained variance in predicting anti–PD-1/PD-L1 response across cancer types from 0.45 (1 - 0.74²) to 0.19 (1 - 0.9²). Both analyses were based on public data of the patients who were not given anti–PD-1/PD-L1 therapy. A potential limitation is that our analysis is based on bulk tumor data, which obviously only reflects the mean expression of different cell populations in the tumor’s microenvironment. Nonetheless, it is quite remarkable and non-trivial that such accuracy can be obtained from the analysis of bulk tumor data, considering the complex interplay be-
between different cell types in the tumor microenvironment, which modulates the clinical response to checkpoint inhibitors. With the accumulation of tumor-specific single cell data and the advancement of computational deconvolution methods, we expect that our results could be further improved in a significant manner, possibly leading to a better understanding of the mechanisms underlying the variability of immune response observed across different cancer types.

Conclusions

Our findings show that variables from the 3 different classes all play an important and quite independent role in mediating anti–PD-1/PD-L1 therapy response across multiple cancers. Their relative explanatory roles may vary in specific cancer types.

REFERENCES