Pan-Cancer Immunogenomic Perspective on the Tumor Microenvironment Based on PD-L1 and CD8 T-Cell Infiltration

Chan-Young Ock1, Bhumsuk Keam1,2, Sehui Kim3, Ju-Seog Lee4, Miso Kim1, Tae Min Kim1,2, Yoon Kyung Jeon3, Dong-Wan Kim1,2, Doo Hyun Chung5, and Dae Seog Heo1,2

Abstract

Purpose: There is currently no reliable biomarker to predict who would benefit from anti-PD-1/PD-L1 inhibitors. We comprehensively analyzed the immunogenomic properties in The Cancer Genome Atlas (TCGA) according to the classification of tumor into four groups based on PD-L1 status and tumor-infiltrating lymphocyte recruitment (TIL), a combination that has been suggested to be a theoretically reliable biomarker of anti-PD-1/PD-L1 inhibitors.

Experimental Design: The RNA expression levels of PD-L1 and CD8A in the samples in the pan-cancer database of TCGA (N = 9,677) were analyzed. Based on their median values, the samples were classified into four tumor microenvironment immune types (TMIT). The mutational profiles, PD-L1 amplification, and viral association of the samples were compared according to the four TMITs.

Results: The proportions of TMIT I, defined high PD-L1 and CD8A expression, were high in lung adenocarcinoma (67.1%) and kidney clear cell carcinoma (64.8%) among solid cancers. The number of somatic mutations and the proportion of microsatellite instable-high tumor in TMIT I were significantly higher than those in other TMITs, respectively (P < 0.001). PD-L1 amplification and oncogenic virus infection were significantly associated with TMIT I, respectively (P < 0.001). A multivariate analysis confirmed that the number of somatic mutations, PD-L1 amplification, and Epstein–Barr virus/human papillomavirus infection were independently associated with TMIT I.

Conclusions: TMIT I is associated with a high mutational burden, PD-L1 amplification, and oncogenic viral infection. This integrative analysis highlights the importance of the assessment of both PD-L1 expression and TIL recruitment to predict responders to immune checkpoint inhibitors.

See related commentary by Schalper et al., p. 2102

©2016 AACR.

Introduction

A recent strategy targeting immune checkpoints such as CTLA-4 and PD-1/PD-L1 shows promising clinical benefits and introduces a paradigm shift in cancer treatment. Immune checkpoint-blocking agents have shown a remarkable clinical efficacy with a long response duration in immunogenic tumors, such as melanoma, renal cell carcinoma, bladder cancer, non–small cell lung carcinoma, and Hodgkin’s lymphoma (1–5).

Through the expression of PD-L1 on the surface of tumor cells, as measured by immunohistochemistry, may potentially serve as a predictive factor to identify patients who would benefit from anti-PD-1/PD-L1 therapy, not all PD-L1-positive patients respond well (4–6). Therefore, PD-L1 expression on tumor cells may not be a simple predictive factor. Interestingly, the degree of tumor-infiltrating lymphocyte (TIL) infiltration and PD-L1 expression in the tumor microenvironment (TME) are also correlated with the clinical outcomes of anti-PD-1/PD-L1 therapies (6 7). Moreover, recent advances in immuno-genomics have shown that tumors with a high mutational burden, abundant neoantigen, and microsatellite (MSI)-high status are associated with a good response to anti-PD-1/PD-L1 therapy (8–14). Moreover, oncogenic viruses such as Epstein–Barr virus (EBV) or human papillomavirus (HPV) are associated with an inflamed TME, and a favorable clinical outcome in response to anti-PD-1/PD-L1 therapy would also be expected (10, 15). Because of the complex nature of tumor immunity, a comprehensive immuno-genomic analysis that include analyses of the interaction between the tumor and TIL in the TME and investigations of the underlying reasons for the promotion of tumor immunogenicity is needed.

Although TIL assessment in the TME is challenging (15), Rooney and colleagues fascinatingly showed immune cytolytic activity by measuring the mRNA expression levels of granzyme A (GZMA) and perforin 1 (PRF1) from large-scale RNA-sequencing (RNAseq) data from The Cancer Genome Atlas (TCGA) (10). These researchers showed that RNAseq data constitute an appropriate model that can be used to assess the TME because the contamination of stromal cells surrounding the tumor would proportionally influence the TME gene expression profiles in an unbiased manner. However, the cytolytic activity assessed by measuring the granzyme A and perforin 1 expression levels would be influenced by the infiltration of not only CD8 T cells.
Translational Relevance

This study classified tumors and their microenvironments into four groups based on PD-L1 expression and tumor-infiltrating lymphocyte (TIL) recruitment, as assessed by CD8A expression, in the pan-cancer RNA-sequencing database of The Cancer Genome Atlas (TCGA, N = 9,677). This immunogenic perspective of four tumor microenvironment immune types (TMIT) would be an appropriate approach for cancer immunotherapy. Across all cancer types in TCGA, TMIT I, defined by high PD-L1 and CD8A expression, was significantly associated with a high number of somatic mutations and neoantigens, PD-L1 amplification, and infection with an oncogenic virus, such as Epstein–Barr virus or human papillomavirus. This integrative analysis highlights the importance of the assessment of both PD-L1 expression and TIL recruitment to predict responders to anti-PD-1/PD-L1 therapies.

Materials and Methods

Processing of genomic data from TCGA project

We used publicly available, level 3 data of TCGA in this study. Clinical information, gene-level somatic mutation data, copy number variation (CNV) data, and mRNA expression data obtained by RNASeq of the TCGA samples were downloaded from the UCSC Cancer Browser (https://genome-cancer.ucsc.edu) on June 3, 2015. To count the number of total somatic mutations, multiple somatic mutations including nonsynonymous mutations, insertion–deletion mutations, and silent mutations were each counted and summed, and germline mutations without somatic mutations were excluded. The amplification and deletion statuses in the CNV thresholded data, which was calculated using GISTIC 2.0 (20), were documented as “2” and “−2”, respectively, and mRNA expression data, which were generated using the Illumina HiSeq V2 platform, are presented as reads per kilobase per million (RPKM) and transformed into log 2 values for analysis. The patients’ clinical data and microsatellite instability (MSI) status of the TCGA tumors were obtained from the TCGA database. The MSI status was available for 1,164 samples, which included COAD (N = 426) and STAD samples (N = 414). The detection of oncogenic viruses, such as EBV, HPV, and hepatitis B virus (HBV), in each tumor sample (N = 6,511) and the neoantigen number (N = 3,726) were also referenced in a previous report by Rooney and colleagues (10). Altogether, samples of 32 cancer types (N = 10,354) were included in the analysis (Supplementary Table S1).

Statistical analyses of genomic data

According to previous reports regarding the four types of immune evasion of cancer (15), after merging log 2-transformed values of the RPKM of PD-L1 and CD8A, we divided all of the TCGA samples into four groups as follows: type I, PD-L1 expression higher than the median and CD8A expression higher than the median; type II, PD-L1 expression lower than the median and CD8A expression higher than the median; type III, PD-L1 expression higher than the median and CD8A expression lower than the median; and type IV, PD-L1 expression lower than the median and CD8A expression lower than the median. The cytolytic activity of each sample was calculated using the log 2-transformed value of the geometric mean of GZMA and PRF1, as previously reported (10). The prognostic significance of the four TMIT was estimated using Kaplan–Meier plots (log-rank test) and Cox proportional hazards regression analysis. The statistical significance of two continuous variables, such as the total number of somatic mutations and the number of neoantigens, and the CD8A expression level, cytolytic activity, and PD-1 expression was calculated by linear regression analysis. The significance of the differences between continuous values, such as the number of mutations, interferon-gamma (IFNG) expression, and cytolytic activity, and categorical variables, such as TMIT, MSI status, and POLE mutation, was calculated by the Wilcoxon rank sum test or analysis of variance with Tukey post hoc test (for comparisons of more than three groups). Comparing the proportion of each TMIT according to categorical variables was performed by Fisher exact test. The number of somatic mutations and CNV, such as the amplification and deletion of a gene in a previously suggested set of 373 genes that are frequently altered in cancer (10, 21) with PD-L1 (CD274), were analyzed by logistic regression to determine whether these variables can significantly predict a specific TMIT. Univariate and multivariate logistic regression analyses were performed to determine whether there is a significant association between clinicopathologic characteristics and the ability to predict TMIT. P values less than 0.05 were considered statistically significant. False discovery rates of less than 5% were applied to control type I errors in the analysis to determine which gene alterations would significantly predict a specific TMIT. All statistical analyses and data presentations were performed in R language 3.1.3 (http://www.r-project.org), with the exception of the survival analysis, which was performed in STATA version 12 (StataCorp LP).

Results

Distribution and clinical implication of TMIT across cancer types

We analyzed 9,677 tumor samples from 32 cancer types included in the TCGA dataset (Supplementary Table S1). The median of the log 2-transformed RPKM values, which indicate the mRNA expression levels of PD-L1 and CD8A, was significantly different between each TMIT (Figure 1A). In this study, we classified tumors into four different types based on the presence of each gene alteration: (i) PD-L1 amplification higher than the median and CD8A expression lower than the median; type II, PD-L1 expression lower than the median and CD8A expression higher than the median; type III, PD-L1 expression higher than the median and CD8A expression lower than the median; and type IV, PD-L1 expression lower than the median and CD8A expression lower than the median. The cytolytic activity of each sample was calculated using the log 2-transformed value of the geometric mean of GZMA and PRF1, as previously reported (10). The prognostic significance of the four TMIT was estimated using Kaplan–Meier plots (log-rank test) and Cox proportional hazards regression analysis. The statistical significance of two continuous variables, such as the total number of somatic mutations and the number of neoantigens, and the CD8A expression level, cytolytic activity, and PD-1 expression was calculated by linear regression analysis. The significance of the differences between continuous values, such as the number of mutations, interferon-gamma (IFNG) expression, and cytolytic activity, and categorical variables, such as TMIT, MSI status, and POLE mutation, was calculated by the Wilcoxon rank sum test or analysis of variance with Tukey post hoc test (for comparisons of more than three groups). Comparing the proportion of each TMIT according to categorical variables was performed by Fisher exact test. The number of somatic mutations and CNV, such as the amplification and deletion of a gene in a previously suggested set of 373 genes that are frequently altered in cancer (10, 21) with PD-L1 (CD274), were analyzed by logistic regression to determine whether these variables can significantly predict a specific TMIT. Univariate and multivariate logistic regression analyses were performed to determine whether there is a significant association between clinicopathologic characteristics and the ability to predict TMIT. P values less than 0.05 were considered statistically significant. False discovery rates of less than 5% were applied to control type I errors in the analysis to determine which gene alterations would significantly predict a specific TMIT. All statistical analyses and data presentations were performed in R language 3.1.3 (http://www.r-project.org), with the exception of the survival analysis, which was performed in STATA version 12 (StataCorp LP).
expression level, of PD-L1 and CD8A were 4.72 and 6.97, respectively. The log transformation has no effect on the division at the median of the distributions, but only improves the analysis of the correlation. The expression levels of PD-L1 and CD8A were generally positively correlated even though a considerable proportion of tumor samples were found to be PD-L1-high and CD8A-low or PD-L1-low and CD8A-high tumors ($P < 0.001$, $R^2 = 0.250$; Fig. 1a and Supplementary Fig. S1). CD8A expression was significantly correlated with cytolytic activity (10) and PD-1 expression ($P < 0.001$, $R^2 = 0.718$ and 0.712, respectively; Supplementary Fig. S2).

All tumor samples were divided into four groups of TMIT according to the median values of PD-L1 and CD8A expression. Among all of the evaluated samples, 34.6% of the samples were according to the median values of PD-L1 and CD8A expression. The proportions of TMIT II (low PD-L1/low CD8A), TMIT III (high PD-L1/low CD8A), and TMIT IV (low PD-L1/high CD8A) were 34.6%, 15.4%, and 15.4%, respectively; refs. 15, 19).

The proportion of TMIT samples was analyzed according to cancer type (Fig. 1b). As expected, cancers derived from lymphoproliferative tissues, such as thymoma (THYM) and diffuse large B cell lymphoma (DLBC), had the highest proportion of TMIT I tumors among all cancer types (64.0% and 70.8%, respectively), further supporting the hypothesis that the signature reliably reflects the TMIT in cancer tissues. Among the solid cancers, lung adenocarcinoma (LUAD, 67.1%), kidney clear cell carcinoma (KIRC, 64.8%), lung squamous cell carcinoma (LUISIC, 63.5%), and head and neck squamous cell carcinoma (HNSC, 54.1%) had the highest proportion of TMIT I samples. The clinicopathological features of the TCGA patients are summarized according to TMIT in Supplementary Table S2. Older patients (>60 years) had a significantly higher proportion of immune type I tumors compared with younger patients, although the actual difference between two groups was small (37.4% vs. 32.4%, $P < 0.001$). A survival analysis according to immune type showed that the overall survival of patients with immune type I tumors was significantly more favorable compared with that of patients with immune type II (hazard ratio (HR) 1.20; 95% confidence interval (CI), 1.07–1.35; $P = 0.002$; Supplementary Fig. S3a). The poor prognostic impact of immune type III compared with that of immune type I was significant in bladder urothelial carcinoma (BLCA), in which PD-L1 expression and TIL are regarded as a predictive factor for the response to anti-PD-1/PD-L1 therapy (ref. 22; HR 1.98; 95% CI, 1.16–3.39; $P = 0.012$; Supplementary Fig. S3b). Although the overall prognosis of TMIT II and IV across all cancer types was not significant, a subgroup analysis showed that SKCM patients with immune type II and IV tumors had a significantly poorer prognosis compared with those with immune type I tumors (immune type II vs. type I: HR 2.04; 95% CI, 1.48–2.82; $P < 0.001$, immune type IV vs. type I: HR 1.58; 95% CI, 1.05–2.38; $P = 0.028$; Supplementary Fig. S4).

High mutational burden is associated with TMIT I

Because previous reports have shown that the degree of mutational burden and the presence of neoantigen are correlated with the immunogenic features of the tumor and reliably predict a good response to an anti-PD-1/PD-L1 treatment strategy (10,12–14), we compared the proportion of different TMIT groups according to the mutational burden. According to cancer type, we observed a tendency toward a correlation between the total number of somatic mutations (hereby, number of mutations) and the number of neoantigens and the proportion of TMIT I (Fig. 1c; Supplementary Fig. S5). Interestingly, TMIT I tumors had a significantly higher number of mutations and a higher number of neoantigens compared with the other TMIT groups ($P < 0.001$, Fig. 2a). Tumor samples with a higher number of mutations than the median value (46) had a significantly higher proportion of TMIT I compared with those with fewer mutations (40.3% vs. 30.4%; $P < 0.001$; Fig. 2b).

The MSI has also been reported as a type of high mutational burden with immunogenic features (14, 23). As expected, MSI-high tumors had a significantly higher number of mutations compared with MSI-low tumors and microsatellite stable (MSS) tumors (Fig. 2c). Moreover, MSI-high tumors had a higher proportion of TMIT I compared with MSI-low tumors and MSS tumors (52.2% versus 25.0% and 29.9%, respectively, $P < 0.001$; Fig. 2d). This association was still clearly observed in the subgroup of COAD and STAD (Supplementary Fig. S6). Because recent studies showed that the POLE mutation is associated with defects in DNA proofreading and an enhanced immune response (24), we also determined whether the POLE mutation is associated with TMIT I. Among samples with different mutational profiles ($N = 5,830$), the POLE mutation was detected in 112 (1.9%) samples, which had a significantly increased number of mutations compared with the number of mutations found in the samples without the POLE mutation (Supplementary Fig. S7). Interestingly, the proportion of TMIT I samples was higher in the tumor samples with the POLE mutation than in the wild-type samples (46.4% vs. 34.8%, $P = 0.021$; Supplementary Fig. S7).

Because the mutational burden is significantly associated with TMIT I, we then assessed the association of TMIT I with somatic mutations in 373 genes that are frequently mutated in many cancers (21) and the PD-L1 (CD274) gene (Table 1 and Supplementary Table S3a and Supplementary Fig. S8). Other than the POLE mutation, we found that VHL, PBRM1, CASP8, and ATM mutations are significantly associated with TMIT I. In contrast, EGFR and BRAF mutations were found to be correlated with TMIT III.

PD-L1 amplification is associated with TMIT I

The association of PD-L1 amplification and TMIT was analyzed because PD-L1 amplification has been reported to serve as a good predictive marker of the response to anti-PD-1/PD-L1 therapy in Hodgkin’s lymphoma (4, 25) and has been found to be associated with high immune cytolytic activity (10). The frequency of PD-L1 amplification in all cancers was 1.6% (149 of 9,364 samples), and HNSC (4.8%), sarcoma (4.7%), ovarian serous cystadenocarcinoma (4.3%), and DLBC (4.2%) had a relatively high frequency of PD-L1 amplification (Fig. 3a). Compared with the wild-type tumors, a high proportion of tumors with PD-L1 amplification were TMIT I (64.4% vs. 34.4%; $P < 0.001$; Fig. 3b and c).

In addition to PD-L1 amplification, the analysis of 373 oncogenes in cancer revealed that APC, NPM1, and CDX1 amplifications as well as CTNNB1 deletion are associated with TMIT I. Moreover, EGRF amplification and CDKN2A deletion were found to be associated with TMIT III (Table 1 and Supplementary Table S3b and S3c and Supplementary Fig. S9).
Oncogenic virus infection is associated with TMIT I

We then aimed to determine whether there is an association between infection with an oncogenic virus and TMIT. Considerable proportions of EBV in STAD, HPV in CESC and HNSC, and HBV in liver hepatocellular carcinoma were detected (LIHC; Supplementary Fig. S10), which is consistent with previous
A high proportion of tumors resulting from EBV infection and HPV infection were classified as TMIT I compared with those resulting from HBV infection or those with no virus (84% and 56.4% vs. 18.5% and 35.9%, respectively, \( P < 0.001 \); Fig. 4a and b). This trend was also significant in EBV-positive and EBV-negative STAD (Supplementary Fig. S11a) and HPV-positive and HPV-negative HNSC (Supplementary Fig. S11b), although HBV positivity did not influence the proportion of TMIT tumors in LIHC (Supplementary Fig. S11c).

IFNG expression and cytolytic activity were significantly higher in TMIT I tumors (Supplementary Figs. S5c and S5d), and cytolytic activity was significantly higher in EBV-infected and HPV-infected tumors compared with HBV-infected and uninfected samples (Supplementary Fig. S5e).
Table 1. Specific genomic alterations of cancer driver genes to predict tumor microenvironment immune type (TMIT)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Genomic alteration</th>
<th>Odds ratio</th>
<th>P value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>VHL</td>
<td>Mutation</td>
<td>3.37</td>
<td>2.23 × 10⁻⁹</td>
<td>4.15 × 10⁻⁹</td>
</tr>
<tr>
<td>CASP8</td>
<td>Mutation</td>
<td>3.13</td>
<td>2.56 × 10⁻⁸</td>
<td>8.74 × 10⁻⁸</td>
</tr>
<tr>
<td>PBRM1</td>
<td>Mutation</td>
<td>2.40</td>
<td>6.59 × 10⁻¹⁰</td>
<td>8.23 × 10⁻¹⁰</td>
</tr>
<tr>
<td>NCOI</td>
<td>Mutation</td>
<td>2.37</td>
<td>2.58 × 10⁻⁹</td>
<td>8.81 × 10⁻⁹</td>
</tr>
<tr>
<td>MXRAS</td>
<td>Mutation</td>
<td>2.13</td>
<td>5.93 × 10⁻¹⁰</td>
<td>4.44 × 10⁻¹⁰</td>
</tr>
<tr>
<td>FAT1</td>
<td>Mutation</td>
<td>1.99</td>
<td>1.44 × 10⁻⁸</td>
<td>6.00 × 10⁻⁸</td>
</tr>
<tr>
<td>ANK3</td>
<td>Deletion</td>
<td>1.95</td>
<td>1.20 × 10⁻⁷</td>
<td>5.64 × 10⁻⁷</td>
</tr>
<tr>
<td>MUC17</td>
<td>Deletion</td>
<td>1.94</td>
<td>1.90 × 10⁻⁷</td>
<td>1.78 × 10⁻⁷</td>
</tr>
<tr>
<td>BRAF</td>
<td>Deletion</td>
<td>1.70</td>
<td>5.50 × 10⁻⁹</td>
<td>3.44 × 10⁻⁹</td>
</tr>
<tr>
<td>FLG</td>
<td>Deletion</td>
<td>1.69</td>
<td>7.05 × 10⁻⁹</td>
<td>3.78 × 10⁻⁹</td>
</tr>
<tr>
<td>CD274</td>
<td>Amplification</td>
<td>3.46</td>
<td>2.05 × 10⁻¹⁴</td>
<td>7.43 × 10⁻¹²</td>
</tr>
<tr>
<td>APC</td>
<td>Amplification</td>
<td>3.33</td>
<td>6.10 × 10⁻⁷</td>
<td>3.68 × 10⁻⁷</td>
</tr>
<tr>
<td>DIAH8</td>
<td>Amplification</td>
<td>3.09</td>
<td>1.16 × 10⁻⁶</td>
<td>1.40 × 10⁻⁶</td>
</tr>
<tr>
<td>CDX1</td>
<td>Amplification</td>
<td>2.75</td>
<td>1.69 × 10⁻⁹</td>
<td>8.75 × 10⁻⁹</td>
</tr>
<tr>
<td>NPM1</td>
<td>Amplification</td>
<td>2.54</td>
<td>2.27 × 10⁻⁹</td>
<td>1.65 × 10⁻⁹</td>
</tr>
<tr>
<td>PIK3</td>
<td>Amplification</td>
<td>1.67</td>
<td>4.94 × 10⁻¹⁰</td>
<td>8.94 × 10⁻¹⁰</td>
</tr>
<tr>
<td>MUC4</td>
<td>Amplification</td>
<td>1.64</td>
<td>2.03 × 10⁻⁹</td>
<td>1.65 × 10⁻⁹</td>
</tr>
<tr>
<td>ACVR2B</td>
<td>Deletion</td>
<td>4.50</td>
<td>1.54 × 10⁻⁸</td>
<td>5.57 × 10⁻⁸</td>
</tr>
<tr>
<td>MYD88</td>
<td>Deletion</td>
<td>4.32</td>
<td>3.56 × 10⁻¹⁰</td>
<td>5.58 × 10⁻¹⁰</td>
</tr>
<tr>
<td>TGFBR2</td>
<td>Deletion</td>
<td>3.42</td>
<td>4.64 × 10⁻¹¹</td>
<td>5.58 × 10⁻¹⁰</td>
</tr>
<tr>
<td>RPSA</td>
<td>Deletion</td>
<td>4.26</td>
<td>1.52 × 10⁻¹²</td>
<td>1.31 × 10⁻¹¹</td>
</tr>
<tr>
<td>SLC22A14</td>
<td>Deletion</td>
<td>3.92</td>
<td>1.81 × 10⁻¹⁰</td>
<td>1.31 × 10⁻¹⁰</td>
</tr>
<tr>
<td>EAF1</td>
<td>Deletion</td>
<td>4.10</td>
<td>6.79 × 10⁻¹⁰</td>
<td>3.50 × 10⁻¹⁰</td>
</tr>
<tr>
<td>MRPS25</td>
<td>Deletion</td>
<td>4.19</td>
<td>6.45 × 10⁻¹⁰</td>
<td>3.50 × 10⁻¹⁰</td>
</tr>
<tr>
<td>CTHNB1</td>
<td>Deletion</td>
<td>3.50</td>
<td>7.86 × 10⁻¹⁰</td>
<td>3.55 × 10⁻¹⁰</td>
</tr>
<tr>
<td>ZNF620</td>
<td>Deletion</td>
<td>3.78</td>
<td>1.55 × 10⁻⁹</td>
<td>6.21 × 10⁻⁹</td>
</tr>
</tbody>
</table>

Please see Supplementary Table S3 for raw data of all significant genomic alterations of genes. Odds ratio was calculated that the proportion of TMIT I in each genomic alteration-positive divided by those in the genomic alteration-negative, by logistic regression. Filtering condition of this table is odds ratio > 1.5 and FDR < 1.00 × 10⁻⁷. FDR, false discovery rate.

Summary of clinicopathologic features correlated with TMIT I

Taken together, the multivariate analysis results show that a high number of mutations, EBV infection, HPV infection, and PD-L1 amplification are independently associated with TMIT I (Table 2).

Discussion

Using a large-scale TCGA pan-cancer dataset, we classified all types of cancer into one of four TMITs based on their PD-L1 and CD8A mRNA expression levels assessed by RNAseq. TMIT varies across cancer type, and this result is likely accompanied by the prevalence of somatic mutations across the cancer types. TMIT I, defined by high PD-L1 and high CD8A expression, is associated with a high mutual burden/neoantigen, high MSI, PD-L1 amplification, and infection with an oncogenic virus such as EBV and HPV. Our results suggest that PD-L1 positivity should be comprehensively interpreted with TME and that classification into four TMITs may be an appropriate model through which to tailor cancer immunotherapeutic modules (Supplementary Fig. S12).

The clinical effects of anti-PD-1/PD-L1 therapies have been modest except in a subset of cancer types, including malignant melanoma, lung cancer, and kidney cancer (2, 3, 27). The underlying biology of such limitations has not been clearly understood until recent studies, which showed that the mutational burden correlates with immunogenic features and favorable clinical outcomes of anti-PD-1/PD-L1 therapies (8, 12–14). Consistent with previous observations, the proportion of TMIT I tumors, which reflects the adaptive immune response (15), was higher in lung cancer and melanoma, as well as head and neck cancer, stomach cancer, and bladder cancer, which are expected to have a favorable clinical outcome in response to anti-PD-1/PD-L1 therapies (5, 28, 29).

The assessment of the tumor mutational burden has indicated that the mutational burden varies between studies, and the prediction of neoantigens by calculating the interaction between a specific mutation and HLA genotype would be theoretically appropriate (10, 12, 30). However, the number of nonsynonymous mutations is also clearly correlated with the mutational burden and the clinical outcome of anti-PD-1 therapy (8, 13). In this study, the number of predicted neoantigens, as referenced by Rooney and colleagues (10), was found to be significantly correlated with the total number of somatic mutations assessed by whole-exome sequencing. Considering that the prediction of neoantigens is only performed in silico (31, 32), calculating the number of somatic mutations would be a more convenient method for assessing the mutational burden because the total number of mutations is also significantly correlated with the number of nonsynonymous mutations (data not shown). The number of mutations is also correlated with the MSI status as well as POLE mutation, which was found to be correlated with TMIT I in this study. This finding is consistent with previous results, which showed that anti-PD-1 antibody exhibits a good clinical outcome in MSI-high tumors (14) and TIL recruitment in POLE-mutated tumors (24).

Consistent with a previous report that showed that oncogenic viral infection increases the cytolytic activity of a tumor (16), a high proportion of tumors associated with EBV and HPV are TMIT I. Viral infection generates various viral antigens inside the tumor, and this phenomenon increases the immunogenicity of the tumor by activating the IFNG pathway (33), as supported by recent findings of the association of PD-L1 with EBV-associated malignancies (34, 35). Moreover, PD-L1 amplification is also clearly correlated with TMIT I. Previous results have clearly shown that Hodgkin’s lymphoma presents variable copy number gains of chromosome 9p24.1, a genomic region that includes PD-L1, POLEDIL2 (encoding PD-L2, another ligand of PD-1), and JAK2, which activates the JAK/STAT/IFNG pathway (25, 36). This finding is consistent with our analysis, which showed a frequent co-amplification of PD-L1 and JAK2 in the TCGA pan-cancer data (data not shown). These findings are also clearly supported by a recent finding of anti-PD-1 antibody in Hodgkin’s lymphoma, which is known to be associated with both EBV and PD-L1 amplification (4).

Along with the investigation of PD-L1 expression as a predictive marker of the response to anti-PD-1/PD-L1 therapies (5, 6), the prognostic impact of PD-L1 has also been actively studied (37, 38). In addition to the fact that the optimal cutoff of PD-L1 has not been defined, the prognostic significance of PD-L1 has been reported to differ according to cancer type. The results from this study show that even in patients with high PD-L1, the overall prognosis differed according to CD8A expression (e.g., TMIT I vs. TMIT III), which would account for the previously published findings of an inconsistent prognostic significance of PD-L1. Therefore, the prognostic impact of PD-L1 expression, along with CD8+ TIL and the underlying etiology of high PD-L1 expression, such as high mutational burden and viral infection that could recruit immunogenicity, should be assessed.
In this study, the proportion of TMIT I tumors was found to be correlated with the number of mutations across cancer type and viral infection status, but considerable discrepancies also exist. KIRC is known as a highly immunogenic tumor type (39), and the results from this study show a high proportion of TMIT I. However, this tumor has a relatively low mutational burden. Interestingly, VHL and PBRM1 mutations, which are frequently shown in KIRC (40), were clearly observed to be correlated with TMIT I. Moreover, hepatocellular carcinoma has a low proportion of TMIT I tumors, although it has a relatively high mutational burden and an association with HBV. These findings are likely the result of the unique anatomical characteristics of immune privilege, such as in glioblastoma (41, 42).

In this study, we used RNAseq data from a mixture of cancer cells and surrounding tissues, including TILs, although this...
Figure 4.
Oncogenic viral infection with increased cytolytic activity is associated with tumor microenvironment immune type I (TMIT). A scatter plot of log2-transformed values of RPKM of PD-L1 and CD8A according to virus detection and PD-L1 amplification is shown (A). The proportion of TMIT according to PD-L1 amplification is compared (B). Box plots of log2-transformed values of RPKM of IFNG (C) and cytolytic activity (D) according to TMITs are shown. A box plot of cytolytic activity according to virus detection and TMIT is shown (E).

Table 2. Logistic regression analysis for predicting tumor microenvironment immune type I according to clinicopathological characteristics

<table>
<thead>
<tr>
<th></th>
<th>Univariate</th>
<th></th>
<th>Multivariate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td><em>P</em></td>
<td>OR (95% CI)</td>
<td><em>P</em></td>
</tr>
<tr>
<td>Age ≥Median</td>
<td>1.24 (1.14–1.35)</td>
<td>5.78 × 10⁻⁷</td>
<td>1.11 (0.96–1.28)</td>
<td>0.147</td>
</tr>
<tr>
<td>Gender Men (vs. women)</td>
<td>0.97 (0.89–1.06)</td>
<td>0.493</td>
<td>Not entered</td>
<td>Not entered</td>
</tr>
<tr>
<td>Number of mutations ≥Median</td>
<td>1.55 (1.40–1.72)</td>
<td>&lt;2.00 × 10⁻⁶</td>
<td>1.44 (1.25–1.67)</td>
<td>6.21 × 10⁻⁷</td>
</tr>
<tr>
<td>Presence of virus EBV</td>
<td>8.14 (3.07–28.0)</td>
<td>1.32 × 10⁻⁴</td>
<td>6.29 (2.34–21.8)</td>
<td>9.01 × 10⁻⁴</td>
</tr>
<tr>
<td>HBV</td>
<td>2.31 (1.75–3.07)</td>
<td>5.05 × 10⁻⁹</td>
<td>1.98 (1.47–2.67)</td>
<td>8.31 × 10⁻⁴</td>
</tr>
<tr>
<td>PD-L1 amplification Yes</td>
<td>0.19 (0.07–0.42)</td>
<td>1.54 × 10⁻⁴</td>
<td>0.18 (0.07–0.40)</td>
<td>1.08 × 10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>3.46 (2.50–4.87)</td>
<td>6.47 × 10⁻¹⁵</td>
<td>3.45 (1.96–6.33)</td>
<td>3.10 × 10⁻⁵</td>
</tr>
</tbody>
</table>

OR, odds ratio.

Ock et al.
contamination of surrounding stromal tissues was not intentional. Interestingly, as interpreted by Rooney and colleagues (10), this limitation would actually be advantageous to assess the infiltration of surrounding tissues, such as the TIL of CD8⁺ CTLs. In this regard, the assessment of CD8A gene expression in a mixture of cancer and stromal cells would be more practical in the clinical setting than the immunohistochemical assessment of CD8⁺ TIL recruitment or activity, which would be difficult to judge uniformly across different tumor types. Nevertheless, the clinical validation of this approach is definitely warranted.

Although this study has some limitations, including that the cutoff values of PD-L1 and CD8A need clinical validation, it is nevertheless valuable because we systematically assessed the TMIT of most cancer types by using data from TCGA project. Our findings are clearly consistent with previous findings. Future investigations and clinical validations regarding the use of this approach for the assessment of immuno-genomic features across cancer types are warranted.

In conclusion, from the pan-cancer immuno-genomic perspective, the classification of tumors into four TMITs based on PD-L1 status and CD8A⁺ TIL is an appropriate approach for cancer immunotherapy. TMIT I (PD-L1⁺/CD8A⁺) was found to be associated with high PD-L1 expression, high mutational burden/neoantigen, high MSI, PD-L1 amplification, and the presence of an oncogenic virus. These factors are likely good predictive factors for the response to anti-PD-1/PD-L1 therapy. Our findings will open up opportunities for finding new anti-PD-1/PD-L1 therapy strategies that can identify a subset of patients who may receive a greater benefit from anticancer immunotherapy.

**References**


Ock et al.


Pan-Cancer Immunogenomic Perspective on the Tumor Microenvironment Based on PD-L1 and CD8 T-Cell Infiltration

Chan-Young Ock, Bhumsuk Keam, Sehui Kim, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-15-2834

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2016/02/18/1078-0432.CCR-15-2834.DC2

Cited articles
This article cites 42 articles, 13 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/22/9/2261.full#ref-list-1

Citing articles
This article has been cited by 17 HighWire-hosted articles. Access the articles at:
http://clincancerres.aacrjournals.org/content/22/9/2261.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://clincancerres.aacrjournals.org/content/22/9/2261.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.