Original Article

Tumor mutational and indel burden: a systematic pan-cancer evaluation as prognostic biomarkers

Hao-Xiang Wu[1#], Zi-Xian Wang[1#], Qi Zhao[2#], Dong-Liang Chen[1#], Ming-Ming He[1], Lu-Ping Yang[1], Ying-Nan Wang[2], Ying Jin[1], Chao Ren[1], Hui-Yan Luo[1], Zhi-Qiang Wang[1], Feng Wang[1]

[1]Department of Medical Oncology, [2]Department of Experimental Research, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, China

Contributions: (I) Conception and design: F Wang, HX Wu, ZX Wang; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: HX Wu, ZX Wang, Q Zhao, DL Chen; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Prof. Feng Wang, MD, PhD. Department of Medical Oncology, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, China. Email: wangfeng@sysucc.org.cn.

Background: Tumor mutational burden (TMB) has been widely studied as a predictive biomarker of response to immune checkpoint inhibitors (ICIs). Besides, evidence suggests frameshift indels are a highly immunogenic mutational class and thus a potentially superior biomarker. However, the general prognostic impact of TMB and indel burden in patients with solid tumors has not been systematically investigated.

Methods: We analyzed 20 primary solid cancer types from The Cancer Genome Atlas (TCGA) database. Clinicopathologic factors, TMB and indel burden were collected or calculated. For each cancer type, the impact of TMB or indel burden on overall survival (OS) was evaluated using the Kaplan-Meier method and Cox regression with the method of inverse probability of treatment weighting.

Results: Twenty cancer types from 6,035 patients were analyzed. Survival analysis showed that TMB had a significant impact on OS in 14 out of these 20 cancer types. According to the general survival impact of TMB, they could be classified into three groups, namely the TMB-Worse (eight cancer types), TMB-Better (six cancer types) and TMB-Similar (six cancer types) group, in which higher TMB was associated with inferior, superior, or similar OS, respectively. The survival impacts of TMB in the TMB-Worse and TMB-Better groups were generally consistent when limited to genes from two FDA-approved panels. Notably, in two out of the six cancer types in the TMB-Similar group, the indel burden significantly affected OS.

Conclusions: TMB, as well as indel burden, has divergent prognostic impact in different cancer types, thus could be incorporated in prognostication and risk stratification. More importantly, the general prognostic impact should be taken into account when establishing the predictive function of TMB to ICI treatment.

Keywords: Indel burden; pan-cancer analysis; prognosis; The Cancer Genome Atlas (TCGA); tumor mutational burden (TMB)

Submitted Jul 31, 2019. Accepted for publication Oct 10, 2019.
doi: 10.21037/atm.2019.10.116
View this article at: http://dx.doi.org/10.21037/atm.2019.10.116

Introduction

Cancer originates from gene mutations (1). Particular gene mutations can be both predictive to treatment response and prognostic for survival (2,3). The total number of mutations is named tumor mutational burden (TMB), which was defined as the number of non-synonymous somatic, coding, base substitution, and indel mutations per megabase (Mb) of genome examined (4). Nowadays, whole-exome sequencing is considered the gold standard to determine the TMB, and

estimating TMB by targeted next-generation sequencing (NGS) panels has been promoted as a simpler and cheaper option (5).

The adaptive immune system detects and identifies tumors by non-self neoantigens that arise as the result of somatic mutations. Thus, it is reasonable to hypothesize that the number of non-synonymous somatic mutations in a tumor, i.e. TMB, may affect the odds of generating immunogenic peptides and thereby influence immune checkpoint inhibitors (ICIs) response in patients (6). Therefore, in the booming era of immuno-oncology, TMB has been widely studied as the predictor of response to ICIs (4,7,8). Its predictive function has been tested retrospectively or prospectively in melanoma, non-small-cell lung cancer, small cell lung cancer and bladder cancer (8-11). What's more, Turajlic and colleagues found that frameshift indels had a more significant association with ICIs response than TMB across three separate melanoma cohorts, indicating that indel burden might be a better biomarker to ICIs (12). However, for clinical application, TMB has yet to address several pitfalls in order to become a reliable predictive biomarker to ICIs. Before establishing the ability of TMB to predict ICI response and survival benefit, its general prognostic impact on overall survival (OS) should be clarified (13), which has never been investigated systematically due to the lack of matched and high-quality data.

The recently completed The Cancer Genome Atlas (TCGA) project provides matched molecular and clinical data (14), which makes it possible to systematically analyze the survival impact of TMB and indel burden. Therefore, we conducted this pan-cancer analysis to evaluate the general prognostic impact of TMB and indel burden in patients newly diagnosed with cancer.

Methods

Patients

For a decade, the TCGA program collected clinicopathologic annotation data together with multi-platform molecular profiles of more than 11,000 human tumors across 33 different cancer types (15). A standardized dataset named the TCGA Pan-Cancer Clinical Data Resource was developed by Liu and colleagues (15) (https://doi.org/10.1016/j.cell.2018.02.052), from which the curated and filtered clinical and survival outcome data with high quality were obtained. Curated and filtered somatic mutation data, which could be readily used for pan-cancer analysis, were obtained from the Multi-Center Mutation Calling in Multiple Cancers (MC3) project (https://gdc.cancer.gov/about-data/publications/mc3-2017), which generated a comprehensive encyclopedia of somatic mutation calls from the TCGA data by Ellrott and colleagues (16). We retrieved all records from 11,160 patients. The exclusion criteria were as follows: (I) patients with incomplete survival data and follow-up information; (II) patients without somatic mutation information.

Within these 33 cancer types, 10 cancer types (diffuse large B-cell lymphoma, pheochromocytoma and paraganglioma, testicular germ cell tumors, thymoma, breast invasive carcinoma, kidney chromophobe, lower grade glioma, prostate adenocarcinoma, rectum adenocarcinoma and thyroid carcinoma) were excluded because the number of death events was too small for OS analysis, as determined by Liu *et al.* (15). For glioblastoma multiforme, acute myeloid leukemia and sarcoma, no stage information or other crucial confounding factors were available in the TCGA Pan-Cancer Clinical Data Resource, which might substantially compromise the reliability of the prognostic analysis of TMB; thus these three cancer types were also not included in this pan-cancer analysis. Eventually, adrenocortical carcinoma (ACC), bladder urothelial carcinoma (BLCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), mesothelioma (MESO), ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), uterine corpus endometrial carcinoma (UCEC), uterine carcinosarcoma (UCS) and uveal melanoma (UVM) were all included in the current analysis.

Variables

Examined clinicopathological variables included age, sex, race, stage, and year of initial pathologic diagnosis. Race was divided into white, black, and other ethnicity. Age and year of initial pathologic diagnosis were both retained as continuous variables. The AJCC staging system was adopted for most cancer types, except for CESC, OV, UCEC, and

UCS, for which clinical stages were adopted.

TMB and indel burden

We used the uniform somatic called variants determined by TCGA MC3 project, which were comprehensively curated from detection using seven methods (MuTect, MuSE, VarScan2, Radia, Pindel, Somatic Sniper, Indelocator). TMB was defined as the number of non-synonymous somatic, coding, base substitution, and indel mutations per megabase (Mb) of genome examined (4), while indel burden consisted of frameshift small insertions and deletions, which created a novel open reading frame and could produce a large quantity of neoantigenic peptides highly distinct from self (17). We used 38 Mb as the estimate of the exome size (18). For the two FDA-authorized or approved panels, the MSK-IMPACT panel and the FoundationOne CDx panel, the coding region captured covers 1.22 Mb and 1.1 Mb, respectively (18,19). The indel burden was calculated as the absolute frameshift indel count per case.

Statistical analysis

The primary outcome of this study was OS. OS was defined as the time from diagnosis to the date of death. Patients who were still alive at the follow-up cut-off date were treated as censored observations. Although retrospective analyses have showed the predictive function of high TMB for a better response to ICIs, the optimal cutpoint to define high TMB varied among studies (20). Therefore, for each cancer type, we varied the threshold of TMB from the 50th to 90th percentiles, and selected the one that yielded the highest statistical significance level to define the TMB-high (TMB > threshold) and TMB-low (TMB ≤ threshold) subsets (21). Associations between TMB and patient demographic characteristics were assessed using Pearson χ^2 or Fisher's exact test for categorical data and the Wilcoxon rank sum test for ordinal and continuous data. Inverse probability of treatment weight (IPTW)-based analysis was adopted to evaluate the impact of TMB on OS. Propensity score of TMB was generated from a multivariable logistic regression model for the association between baseline covariates and TMB. The logistic model was constructed via stepwise variable selection, with a threshold of P<0.20 required for initial inclusion and P<0.10 required to remain in the model. On the basis of the propensity score, IPTW was calculated for each patient (22). The impact of TMB on OS was then evaluated using univariate analysis based on the Kaplan-Meier estimator and Cox regression weighted by IPTW (22). Similar analyses were performed to evaluate the impact of indel burden on OS. The nominal level of significance was set at 0.05. Statistical analyses were performed using R v. 3.5.1 (http://www.r-project.org).

Results

Figure 1 depicted the design of this pan-cancer analysis. We obtained curated and filtered somatic mutation data from the Multi-Center Mutation Calling in Multiple Cancers (MC3) project (16), and clinical and survival outcome data from the TCGA Pan-Cancer Clinical Data Resource (15). In total, 6,035 eligible patients from 20 different cancer types were enrolled in this study. The TMB across 20 cancer types are presented in *Figure 2A*. The median TMB ranged from 0.34 mutations per megabase (mut/Mb) to 13.09 mut/Mb, with SKCM having the highest median TMB while UVM had the lowest. The median number of indels ranged from 0 to 9, UVM still had the lowest median number of indels, but LUSC other than SKCM had the highest (*Figure 2B*). The baseline patient characteristics for each cancer type are shown in *Table 1*.

For each cancer type, we varied the threshold of TMB from the 50th to 90th percentiles, and the one that yielded the highest statistical significance level was selected (21) (*Tables S1-S10*). As shown in *Tables S1-S10*, clinicopathologic factors were not equally distributed between TMB-high and TMB-low patients across different cancer types. And TMB-high patients tended to be older, with statistically significant differences observed in eight out of 20 cancer types.

To mitigate against potential bias caused by unbalanced variables between the two subsets, we employed the IPTW method for bias adjustment. After being weighted by IPTW, we found that the impact of TMB on OS was different across cancers. These 20 cancer types could be classified into three distinct groups, namely the TMB-Worse, TMB-Better and TMB-Similar groups. Eight cancer types, ACC, CHOL, COAD, ESCA, KIRC, LIHC, MESO and PAAD were included in the TMB-Worse group, in which patients with high TMB had a poorer prognosis compared with those with low TMB (*Figure 3A,B,C,D,E,F,G,H*, all P<0.05). However, in the TMB-Better group, which included BLCA, KIRP, STAD, CESC, OV and UCEC, high TMB was a statistically significant prognostic indicator of decreased mortality (*Figure 3I,J,K,L,M,N*, all P<0.05). In the TMB-Similar group, TMB did not have a significant impact on

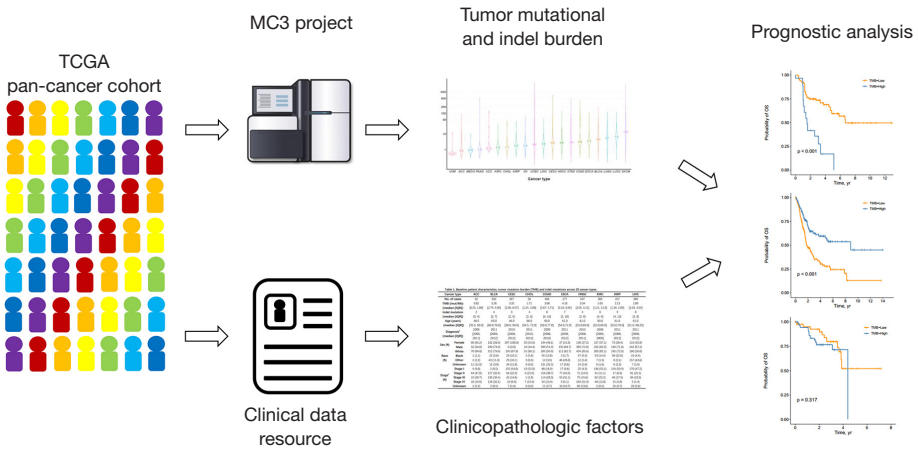

Figure 1 Study design.

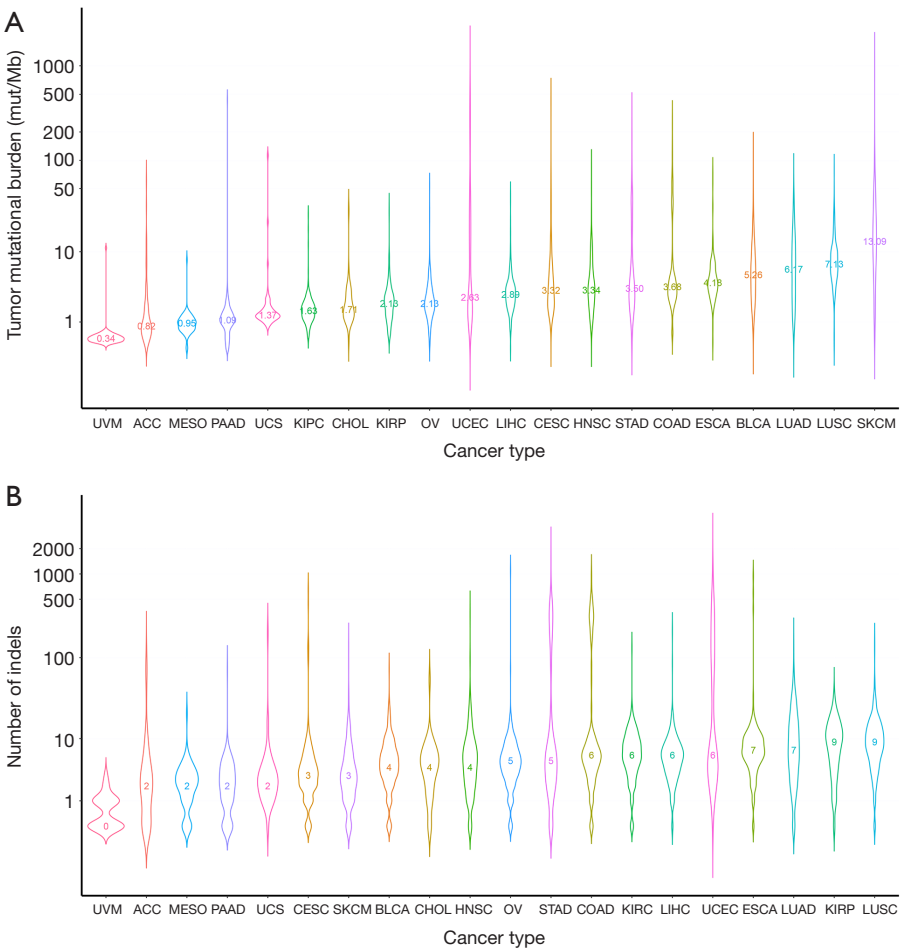

Figure 2 Tumor mutational burden (TMB) and indel burden across 20 cancer types. Number in each violin plot denotes the median. (A) The median TMB ranges from 0.34 to 13.09 mut/Mb, with skin cutaneous melanoma having the highest median TMB while uveal melanoma had the lowest; (B) the median number of indels ranges from 0 to 9, with lung squamous cell carcinoma having the highest median number of indels while uveal melanoma had the lowest.

Table 1 Baseline patient characteristics, tumor mutational burden (TMB) and indel burden across 20 cancer types

Cancer type	ACC	BLCA	CESC	CHOL	COAD	ESCA	HNSC	KIRC	KIRP	LIHC	LUAD	LUSC	MESO	OV	PAAD	SKCM	STAD	UCEC	UCS	UVM
No. of cases	92	392	287	36	403	177	507	369	257	360	496	458	81	408	176	448	429	522	57	80
TMB (mut/Mb), median [IQR]	0.82 [0.55–1.86]	5.26 [2.79–9.36]	3.32 [2.00–6.07]	1.71 [1.25–2.80]	3.68 [2.67–5.71]	4.18 [3.16–6.05]	3.34 [2.05–5.51]	1.63 [1.21–2.13]	2.13 [1.34–2.89]	2.89 [2.03–3.95]	6.17 [2.47–12.66]	7.13 [5.01–10.47]	0.95 [0.66–1.16]	2.13 [1.50–3.16]	1.09 [0.76–1.47]	13.09 [5.91–28.09]	3.5 [2.16–7.45]	2.63 [1.58–19.47]	1.37 [1.21–1.79]	0.34 [0.26–0.45]
Indel mutation, median [IQR]	2 [0–4]	4 [3–7]	3 [2–5]	4 [2–6]	6 [4–10]	7 [5–10]	4 [2–8]	6 [4–9]	9 [4–13]	6 [3–8]	7 [3–16]	9 [5–13]	2 [1–3]	5 [3–7]	2 [1–4]	3 [1–5]	5 [2–12]	6 [3–97.75]	2 [1–4]	0 [0–1]
Age (years), median [IQR]	48.5 [35.5–60.0]	69.0 [60.0–76.0]	46.0 [38.0–56.0]	66.5 [56.5–72.0]	68.0 [58.0–77.0]	61.0 [54.0–72.0]	61.0 [53.0–69.0]	60.0 [52.0–69.0]	61.0 [53.0–70.0]	61.0 [51.0–68.25]	66.0 [59.0–73.0]	68.0 [62.0–73.0]	64.0 [57.0–69.0]	59.0 [51.0–68.0]	65.0 [57.0–73.0]	59.0 [48.0–71.0]	67.0 [58.0–73.0]	64.0 [57.0–71.0]	68.0 [62.0–76.0]	61.5 [51.0–74.3]
Diagnosis†, median [IQR]	2008 [2005–2011]	2011 [2009–2012]	2010 [2006–2012]	2011 [2010–2012]	2009 [2006–2010]	2011 [2008–2012]	2010 [2006–2011]	2006 [2005–2009]	2011 [2008–2012]	2011 [2008–2012]	2010 [2007–2011]	2009 [2005–2011]	2010 [2008–2012]	2005 [2001–2008]	2012 [2010–2012]	2008 [2003–2012]	2011 [2010–2012]	2009 [2007–2010]	2009 [2007–2011]	2012 [2011–2013]
Sex (%)																				
Female	60 (65.2)	102 (26.0)	287 (100.0)	20 (55.6)	194 (48.1)	27 (15.3)	138 (27.2)	137 (37.1)	73 (28.4)	118 (32.8)	266 (53.6)	122 (26.6)	15 (18.5)	408 (100.0)	79 (44.9)	167 (37.3)	155 (36.1)	522 (100.0)	57 (100.0)	35 (43.8)
Male	32 (34.8)	290 (74.0)	0 (0.0)	16 (44.4)	209 (51.9)	150 (84.7)	369 (72.8)	232 (62.9)	184 (71.6)	242 (67.2)	230 (46.4)	336 (73.4)	66 (81.5)	0 (0.0)	97 (55.1)	281 (62.7)	274 (63.9)	0 (0.0)	0 (0.0)	45 (56.2)
Race (%)																				
White	78 (84.8)	312 (79.6)	195 (67.9)	31 (86.1)	205 (50.9)	111 (62.7)	434 (85.6)	303 (82.1)	185 (72.0)	180 (50.0)	388 (78.2)	339 (74.0)	79 (97.5)	347 (85.0)	154 (87.5)	429 (95.8)	273 (63.6)	360 (69.0)	44 (77.2)	55 (68.8)
Black	1 (1.1)	22 (5.6)	29 (10.1)	2 (5.6)	55 (13.6)	3 (1.7)	47 (9.3)	53 (14.4)	58 (22.6)	16 (4.4)	52 (10.5)	28 (6.1)	1 (1.2)	26 (6.4)	7 (4.0)	0 (0.0)	13 (3.0)	105 (20.1)	9 (15.8)	25 (31.2)
Other	2 (2.2)	43 (11.0)	29 (10.1)	3 (8.3)	12 (3.0)	46 (26.0)	12 (2.4)	7 (1.9)	8 (3.1)	157 (43.6)	9 (1.8)	9 (2.0)	1 (1.2)	17 (4.2)	11 (6.2)	12 (2.7)	89 (20.7)	33 (6.3)	3 (5.3)	0 (0.0)
Unknown	11 (12.0)	15 (3.8)	34 (11.8)	0 (0.0)	131 (32.5)	17 (9.6)	14 (2.8)	6 (1.6)	6 (2.3)	7 (1.9)	47 (9.5)	82 (17.9)	0 (0.0)	18 (4.4)	4 (2.3)	7 (1.6)	54 (12.6)	24 (4.6)	1 (1.8)	0 (0.0)
Stage‡ (%)																				
Stage I	9 (9.8)	2 (0.5)	155 (54.0)	19 (52.8)	68 (16.9)	17 (9.6)	25 (4.9)	196 (53.1)	154 (59.9)	170 (47.2)	266 (53.6)	222 (48.5)	9 (11.1)	15 (3.7)	19 (10.8)	77 (17.2)	54 (12.6)	324 (62.1)	22 (38.6)	0 (0.0)
Stage II	44 (47.8)	127 (32.4)	64 (22.3)	9 (25.0)	156 (38.7)	77 (43.5)	71 (14.0)	41 (11.1)	17 (6.6)	81 (22.5)	117 (23.6)	143 (31.2)	16 (19.8)	22 (5.4)	145 (82.4)	138 (30.8)	128 (29.8)	52 (10.0)	5 (8.8)	39 (48.8)
Stage III	19 (20.7)	135 (34.4)	42 (14.6)	1 (2.8)	114 (28.3)	55 (31.1)	79 (15.6)	82 (22.2)	46 (17.9)	84 (23.3)	80 (16.1)	82 (17.9)	40 (49.4)	307 (75.2)	4 (2.3)	167 (37.3)	179 (41.7)	119 (22.8)	20 (35.1)	36 (45.0)
Stage IV	18 (19.6)	126 (32.1)	19 (6.6)	7 (19.4)	54 (13.4)	9 (5.1)	263 (51.9)	48 (13.0)	15 (5.8)	5 (1.4)	26 (5.2)	7 (1.5)	16 (19.8)	61 (15.0)	5 (2.8)	22 (4.9)	42 (9.8)	27 (5.2)	10 (17.5)	4 (5.0)
Unknown	2 (2.2)	2 (0.5)	7 (2.4)	0 (0.0)	11 (2.7)	19 (10.7)	69 (13.6)	2 (0.5)	25 (9.7)	20 (5.6)	7 (1.4)	4 (0.9)	0 (0.0)	3 (0.7)	3 (1.7)	44 (9.8)	26 (6.1)	0 (0.0)	0 (0.0)	1 (1.2)

†, denotes 'year of initial pathologic diagnosis'; ‡, included AJCC stage for most cancer types except for CESC, OV, UCEC, and UCS, for which clinical stage was applied. ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; ESCA, esophageal carcinoma; HNSC, head and neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; UVM, uveal melanoma; Mut/Mb, mutations per megabase; IQR, interquartile range.

OS (*Figure 3O,P,Q,R,S,T*, all P>0.05).

Hazard ratios (HRs) based on IPTW-weighted Cox proportional hazards models were shown in *Figure 4A*. In the TMB-Worse group, CHOL had the highest HR =6.10 [95% confidence interval (CI), 1.91–19.46, P=0.002], followed by ACC, MESO, KIRC, COAD, ESCA, PAAD and LIHC. In the TMB-Better group, KIRP had the lowest HR =0.34 (95% CI, 0.16–0.70, P=0.004), followed by STAD, BLCA, CESC, OV and UCEC. Trends toward better outcome of TMB-low patients were observed in some cancer types of the TMB-Similar group such as LUAD, SKCM and UVM.

When the analysis was limited to data from only those genes included in the FDA-authorized MSK-IMPACT panel (*Figure 4B*) or the FDA-approved FoundationOne CDx panel (*Figure 4C*) used in routine clinical practice, TMB remained significantly associated with OS in the majority of cancer types in the TMB-Worse and TMB-Better groups, with only numerical survival differences captured in several cancer types.

Indel burden also had divergent prognostic impact in different cancer types, but differences from that of TMB existed, with six out of 20 cancer types in which the survival impact was significant (*Figure 4D*). Notably, indel burden significantly affected OS in two out of six cancer types in the TMB-Similar group. In UCS, patients with high indel burden had significantly worse prognosis than those with low indel burden, while in SKCM, high indel burden were prognostic of better survival (*Figure S1*).

Discussion

In this pan-cancer analysis, we found that both TMB and indel burden had divergent survival impacts in different cancer types. In ACC, CHOL, COAD, ESCA, KIRC, LIHC, MESO and PAAD, high TMB was strongly associated with inferior survival; whereas in BLCA, KIRP, STAD, CESC, OV and UCEC, patients with high TMB had superior prognosis compared with those with low TMB. In the other six cancer types, HNSC, LUAD, LUSC, SKCM, UVM and UCS, TMB did not significantly impact the survival of patients newly diagnosed with cancer. Interestingly, in two out of these six cancer types, indel burden significantly impacted OS, although it did not impact OS in all the other 14 cancer types as TMB did.

From palliative chemotherapy to neoadjuvant chemotherapy, the extensive indications of ICIs have created a new era of immuno-oncology (23). Consequently, the predictive value of TMB and indel burden is considerable to help to identify patients who will derive the greatest therapeutic benefit (24). However, there appears to be groupthink to rush TMB for approval by the US FDA and widespread use in practice despite of several unsettled pitfalls, such as its general prognostic impact (13). Samstein and colleagues showed that "Tumor mutational load predicts survival after immunotherapy across multiple cancer types" (25). Meanwhile, as they evaluated the predictive function of TMB using single-arm data, they additionally investigated the general prognostic impact of TMB in several tumor entities, but failed to reveal significant survival differences between TMB-high and TMB-low patients in the majority of entities examined. However, the number of patients of each tumor entities in the non-ICI-treated cohort they studied was relatively small, and some entities were merged together, which might lack power to discover significant survival differences. What's more, the recent announced results of Checkmate-227 investigating nivolumab plus ipilimumab versus chemotherapy failed to establish the predictive function of TMB (26). Therefore, we believe that it is of great importance to thoroughly investigate the general prognostic impact of TMB in patients newly diagnosed with cancer using the high-quality and matched data from TCGA. The results of this study showed that TMB has divergent survival impact in different cancer types, which should be seriously taken into account when establishing the predictive function of TMB to ICIs. For instance, for cancer types in the TMB-Better group, we should be cautious when we observed better survival in the TMB-high group than that in the TMB-low group in ICI-treated cohort, which might be only due to the general prognostic impact of TMB and unrelated to ICIs. Whereas for cancer types in the TMB-Worse and TMB-Similar group, if significant survival benefit was found in the TMB-high group versus TMB-low group, we could confidently claim the predictive function of TMB to ICIs. Our findings also suggest that an optimized genomic classifier incorporating TMB information is likely to improve prognostication, risk stratification, and management in the majority (14/20) of the examined cancer types.

Previous studies of the survival impact of TMB have focused on several specific cancer types. Li and colleagues have found that the MUC16 mutation was associated with high TMB and subsequently better prognosis in STAD, which is consistent with our finding (27). While in microsatellite stable metastatic colorectal cancer patients

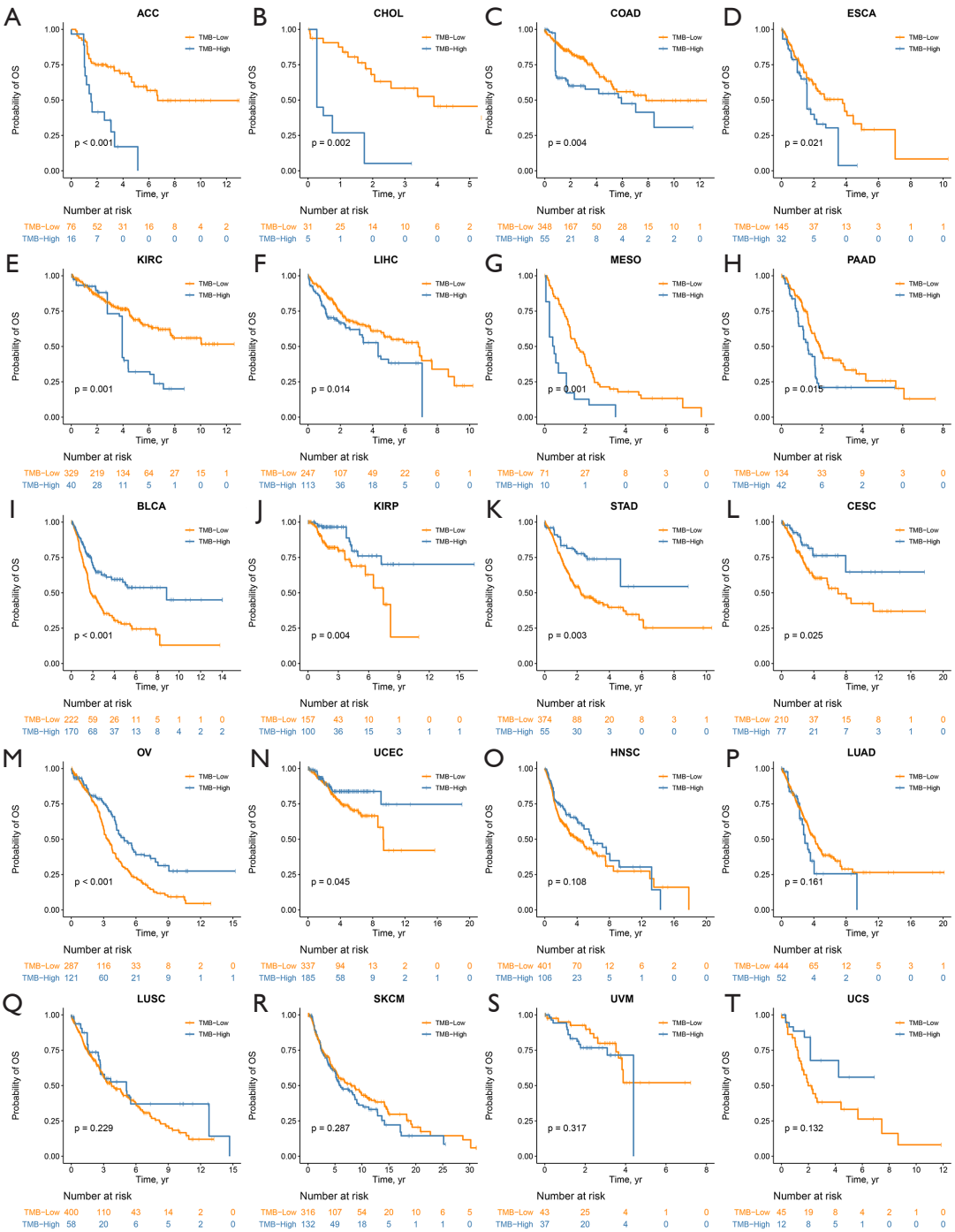

Figure 3 The inverse probability of treatment weighted Kaplan-Meier survival curves of the impact of tumor mutational burden (TMB) on overall survival. High TMB was associated with worse survival in adrenocortical carcinoma (A), cholangiocarcinoma (B), colon adenocarcinoma (C), esophageal carcinoma (D), kidney renal clear cell carcinoma (E), liver hepatocellular carcinoma (F), mesothelioma (G) and pancreatic adenocarcinoma (H). While high TMB predicted better prognosis in bladder urothelial carcinoma (I), kidney renal papillary cell carcinoma (J), stomach adenocarcinoma (K), cervical squamous cell carcinoma and endocervical adenocarcinoma (L), ovarian serous cystadenocarcinoma (M), and uterine corpus endometrial carcinoma (N). In head and neck squamous cell carcinoma (O), lung adenocarcinoma (P), lung squamous cell carcinoma (Q), skin cutaneous melanoma (R), uveal melanoma (S) and uterine carcinosarcoma (T), TMB did not have a significant impact on overall survival.

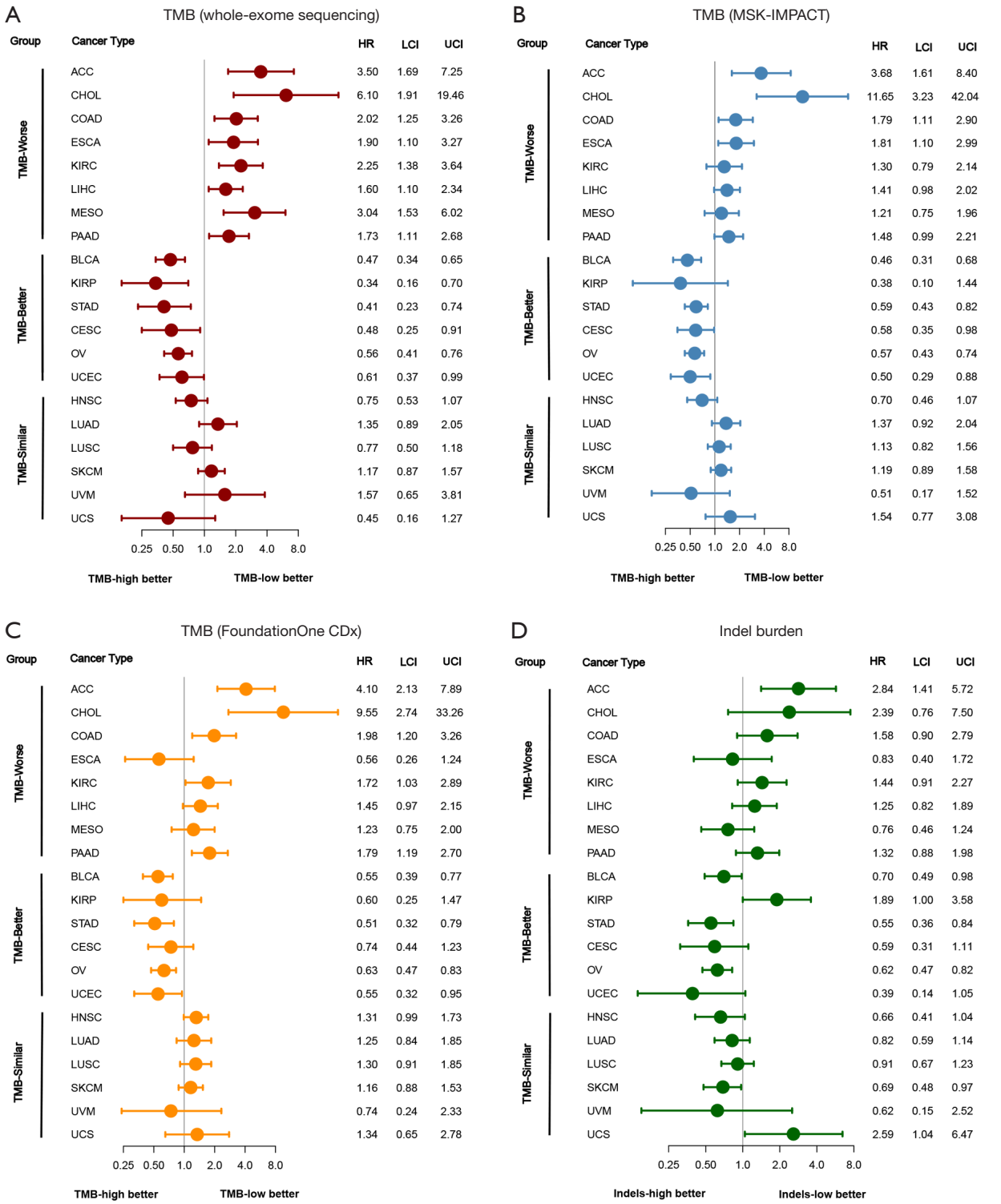

Figure 4 Forest plots of hazard ratios based on inverse probability of treatment weighted Cox proportional hazards models. Forest plots showing hazard ratios of the impact of tumor mutational burden on overall survival in 20 cancer types (A) and when analysis was limited to data from only those genes included in the MSK-IMPACT panel (B) or FoundationOne CDx panel (C), as well as hazard ratios of the impact of indel burden on overall survival in 20 cancer types (D). HR, hazard ratio; LCI, lower 95% confidence interval; UCI, upper 95% confidence interval.

receiving first-line chemotherapy plus targeted therapy, TMB-high was found to be associated with favorable OS (28), which seemed to be in conflict with our results. We assumed that microsatellite stable metastatic colorectal cancer patients with high TMB might be more sensitive to chemotherapy plus targeted therapy, thus TMB-high should probably be a predictive biomarker to chemotherapy plus targeted therapy rather than a general prognostic factor, because Samstein *et al.* also found that TMB-high was associated with worse OS in patients with colorectal cancer (25). But still, further data are required to confirm this issue. On the other hand, in non-small-cell lung cancer, Owada-Ozaki and colleagues found that high TMB was associated with poor prognosis in patients with completely resected non-small cell lung cancer (29). In contrast, results from pooled analysis of the LB2 study suggested that high TMB was associated with better prognosis in patients with resected non-small-cell lung cancer (30). The population characteristics are similar in the aforementioned two studies except for race and sample size, which could not account for the conflicting conclusions. Our analysis showed that TMB was associated with prognosis neither in patients with LUSC nor in patients with LUAD, indicating that further investigations considering specific confounders such as histology and driver gene mutations are needed. In addition, Hwang and colleagues found that high TMB resulted in inferior OS in patients with neuroblastoma, a cancer type that has not been included in the TCGA (31). Except for the aforementioned cancer types, there is no study on the survival impact of TMB in patients with other cancer types. For the first time, our pan-cancer analysis provides key information and data to address these issues.

In addition to TMB, indel burden is also an emerging biomarker for ICIs treatment. Based on the hypothesis that frameshift indels might be an ideal source of tumor-derived neoantigens and so induce multiple neoantigen reactive T cells, Turajlic and colleagues further found that frameshift indels had more significant association with ICIs response than TMB (12). However, scarce data were available regarding the general prognostic impact of indel burden. Our results showed that indel burden also had divergent prognostic impact in different cancer types. Unlike TMB, indel burden only significantly impacted OS in six out of 20 cancer types. As demonstrated by Turajlic *et al.*, the median proportion of indels (calculated as number of indels divided by total counts of TMB) in most of the cancer types were relatively low (<10%) (12). We also found that the range of the absolute indel counts were quite narrow in the majority of these 20 cancer types (*Table 1*). We speculated that the discriminative power of indel burden was limited consequently. However, indel burden significantly impacted OS in two out of six cancer types in the TMB-Similar group, suggesting a supplementary role of indel burden to TMB.

A high TMB representative of a complex genetic profile may be a hallmark of more aggressive and treatment-refractory disease (31), which may lead to worse prognosis for patients with high TMB in the TMB-Worse group. Increased TMB may also be associated with increased prevalence of resistance pathways (32). However, a high TMB was also proposed to reflect the presence of mutation-associated neoantigens, with consequent increased lymphocyte infiltration in the tumor microenvironment and better prognosis (33), which may explain why high TMB was associated with superior OS in the TMB-Better group. To sum up, high TMB is more like a double-edged sword, and which survival path it would lead to mainly depends on the interaction between tumor and microenvironment.

Routine determination of TMB through whole-exome sequencing is currently still clinically impractical (34). Determination of TMB using NGS panels is now more feasible as the commercialization of targeted NGS panels scales up (35,36). TMB determined by the FDA-authorized MSK-IMPACT panel or the FDA-approved FoundationOne CDx panel, was proven to be the predictor of response to ICIs in multiple cancer types (7,19,37). Hence, we limited the analysis to these two panels (18,38), and found that the prognostic value of TMB remained significant in the majority of the 14 cancer types in the TMB-Worse and TMB-Better groups. This suggests that TMB based on targeted NGS panel is of great clinical value. However, only numerical survival differences were captured in several cancer types, suggesting that gene panels tailored specifically for these cancer types may be necessary.

The present study has several limitations. Although TCGA provides us with high-quality data, the included confounding factors were limited as this is a pan-cancer analysis. The impact of unaccounted confounders, e.g., treatment information, on our results cannot be excluded and further validation from prospective studies is still needed. Due to the lack of inadequate survival information, we did not study the survival impact of TMB or indel burden in patients with several common cancer types, such as breast invasive carcinoma and rectal adenocarcinoma, which are pending investigation (15). Besides, in the TCGA cohort, advanced cancers are relatively underrepresented,

which might lead to potential bias.

In conclusion, our pan-cancer analysis provides key information pertaining to the general prognostic impact of TMB and indel burden across 20 primary solid tumors. We found that TMB has divergent survival impacts in different cancer types, thus could be incorporated in prognostication and risk stratification. More importantly, the prognostic impact should be taken into account when establishing the predictive function of TMB to ICI treatment.

Acknowledgments

We would like to thank the staff members of the TCGA Research Network for their involvement in the TCGA Pan-Cancer Clinical Data Resource project and the TCGA MC3 project. We would also like to thank Mr. Christopher Lavender, the native language editor from Sun Yat-sen University Cancer Center, for his assistance in editing this manuscript.

Funding: This work was supported by grants from the National Key Research and Development Program of China (2017YFC1308900); the National Natural Science Foundation of China (81872011); the Guangzhou Health and Medical Collaborative Innovation Project (201704020220); the Sun Yat-sen University Clinical Research 5010 Program (2018014); and the Young Teacher Training Program of Sun Yat-sen University (17ykpy78).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethics Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Ethical approval was waived because we used only publicly available data and materials in this study.

References

1. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science 2015;349:1483-9.
2. Sanz-Garcia E, Argiles G, Elez E, et al. BRAF mutant colorectal cancer: prognosis, treatment, and new perspectives. Ann Oncol 2017;28:2648-57.
3. Allegra CJ, Rumble RB, Hamilton SR, et al. Extended RAS Gene Mutation Testing in Metastatic Colorectal Carcinoma to Predict Response to Anti-Epidermal Growth Factor Receptor Monoclonal Antibody Therapy: American Society of Clinical Oncology Provisional Clinical Opinion Update 2015. J Clin Oncol 2016;34:179-85.
4. Yarchoan M, Hopkins A, Jaffee EM. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. N Engl J Med 2017;377:2500-1.
5. Wu HX, Wang ZX, Zhao Q, et al. Designing gene panels for tumor mutational burden estimation: the need to shift from 'correlation' to 'accuracy'. J Immunother Cancer 2019;7:206.
6. Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. Nat Rev Cancer 2019;19:133-50.
7. Gandara DR, Paul SM, Kowanetz M, et al. Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. Nat Med 2018;24:1441-8.
8. Hellmann MD, Ciuleanu TE, Pluzanski A, et al. Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden. N Engl J Med 2018;378:2093-104.
9. Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 2015;348:124-8.
10. Hugo W, Zaretsky JM, Sun L, et al. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. Cell 2016;165:35-44.
11. Rosenberg JE, Hoffman-Censits J, Powles T, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. Lancet 2016;387:1909-20.
12. Turajlic S, Litchfield K, Xu H, et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. Lancet Oncol 2017;18:1009-21.
13. Addeo A, Banna GL, Weiss GJ. Tumor Mutation Burden-From Hopes to Doubts. JAMA Oncol 2019;5:934-5.
14. The TCGA Legacy. Cell 2018;173:281-2.
15. 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.
16. Ellrott K, Bailey MH, Saksena G, et al. Scalable Open Science Approach for Mutation Calling of Tumor Exomes

Using Multiple Genomic Pipelines. Cell Syst 2018;6:271-81.e7.
17. Linnebacher M, Gebert J, Rudy W, et al. Frameshift peptide-derived T-cell epitopes: a source of novel tumor-specific antigens. Int J Cancer 2001;93:6-11.
18. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med 2017;9:34.
19. Rizvi H, Sanchez-Vega F, La K, et al. Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non-Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. J Clin Oncol 2018;36:633-41.
20. Endris V, Buchhalter I, Allgauer M, et al. Measurement of tumor mutational burden (TMB) in routine molecular diagnostics: in silico and real-life analysis of three larger gene panels. Int J Cancer 2019;144:2303-12.
21. Thomas A, Routh ED, Pullikuth A, et al. Tumor mutational burden is a determinant of immune-mediated survival in breast cancer. Oncoimmunology 2018;7:e1490854.
22. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol 2008;168:656-64.
23. Constantinidou A, Alifieris C, Trafalis DT. Targeting Programmed Cell Death -1 (PD-1) and Ligand (PD-L1): A new era in cancer active immunotherapy. Pharmacol Ther 2019;194:84-106.
24. Conway JR, Kofman E, Mo SS, et al. Genomics of response to immune checkpoint therapies for cancer: implications for precision medicine. Genome Med 2018;10:93.
25. Samstein RM, Lee CH, Shoushtari AN. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat Genet 2019;51:202-6.
26. Hellmann MD, Paz-Ares L, Bernabe Caro R, et al. Nivolumab plus Ipilimumab in Advanced Non-Small-Cell Lung Cancer. N Engl J Med 2019;381:2020-31.
27. Li X, Pasche B, Zhang W, et al. Association of MUC16 Mutation With Tumor Mutation Load and Outcomes in Patients With Gastric Cancer. JAMA Oncol 2018;4:1691-8.
28. Innocenti F, Ou FS, Qu X, et al. Mutational Analysis of Patients With Colorectal Cancer in CALGB/SWOG 80405 Identifies New Roles of Microsatellite Instability and Tumor Mutational Burden for Patient Outcome. J Clin Oncol 2019;37:1217-27.
29. Owada-Ozaki Y, Muto S, Takagi H, et al. Prognostic Impact of Tumor Mutation Burden in Patients With Completely Resected Non-Small Cell Lung Cancer: Brief Report. J Thorac Oncol 2018;13:1217-21.
30. Devarakonda S, Rotolo F, Tsao MS, et al. Tumor Mutation Burden as a Biomarker in Resected Non-Small-Cell Lung Cancer. J Clin Oncol 2018;36:2995-3006.
31. Hwang WL, Wolfson RL, Niemierko A, et al. Clinical Impact of Tumor Mutational Burden in Neuroblastoma. J Natl Cancer Inst 2018. [Epub ahead of print].
32. Offin M, Rizvi H, Tenet M, et al. Tumor Mutation Burden and Efficacy of EGFR-Tyrosine Kinase Inhibitors in Patients with EGFR-Mutant Lung Cancers. Clin Cancer Res 2019;25:1063-9.
33. Giannakis M, Mu XJ, Shukla SA, et al. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. Cell Rep 2016;15:857-65.
34. Chan TA, Yarchoan M, Jaffee E, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. Ann Oncol 2019;30:44-56.
35. Hyman DM, Solit DB, Arcila ME, et al. Precision medicine at Memorial Sloan Kettering Cancer Center: clinical next-generation sequencing enabling next-generation targeted therapy trials. Drug Discov Today 2015;20:1422-8.
36. Szustakowski JD, Green G, Geese WJ, et al. Evaluation of tumor mutation burden as a biomarker for immune checkpoint inhibitor efficacy: A calibration study of whole exome sequencing with FoundationOne®. Cancer Res 2018;78:abstr 5528.
37. Goodman AM, Kato S, Bazhenova L, et al. Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. Mol Cancer Ther 2017;16:2598-608.
38. Zehir A, Benayed R, Shah RH, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med 2017;23:703-13.

Cite this article as: Wu HX, Wang ZX, Zhao Q, Chen DL, He MM, Yang LP, Wang YN, Jin Y, Ren C, Luo HY, Wang ZQ, Wang F. Tumor mutational and indel burden: a systematic pan-cancer evaluation as prognostic biomarkers. Ann Transl Med 2019;7(22):640. doi: 10.21037/atm.2019.10.116

Supplementary

Table S1 Tumor mutational burden (TMB) threshold and baseline patient characteristics in TMB-high and TMB-low subsets across 20 cancer types (ACC, BLCA)

Variable	ACC			BLCA		
	TMB-low	TMB-high	P	TMB-low	TMB-high	P
Threshold, mut/Mb (percentile)	2.24 (83rd)			5.95 (57th)		
N	76	16		222	170	
Age (years), median [IQR]	45.00 [32.00–57.25]	59.00 [52.75–63.25]	0.012[&]	68.00 [60.00–76.00]	69.00 [61.00–76.00]	0.383
Diagnosis*, median [IQR]	2008.50 [2006.00–2011.00]	2005.50 [2001.75–2011.00]	0.386	2011.00 [2009.00–2012.00]	2011.00 [2008.25–2012.00]	0.296
Sex (%)			1.000			0.246
Female	49 (64.5)	11 (68.8)		63 (28.4)	39 (22.9)	
Male	27 (35.5)	5 (31.2)		159 (71.6)	131 (77.1)	
Race (%)			0.821			0.008[&]
White	63 (82.9)	15 (93.8)		165 (74.3)	147 (86.5)	
Black	1 (1.3)	0 (0.0)		13 (5.9)	9 (5.3)	
Other	2 (2.6)	0 (0.0)		34 (15.3)	9 (5.3)	
Unknown	10 (13.2)	1 (6.2)		10 (4.5)	5 (2.9)	
Stage[#] (%)			0.002[&]			0.581
Stage I	9 (11.8)	0 (0.0)		2 (0.9)	0 (0.0)	
Stage II	41 (53.9)	3 (18.8)		74 (33.3)	53 (31.2)	
Stage III	15 (19.7)	4 (25.0)		72 (32.4)	63 (37.1)	
Stage IV	10 (13.2)	8 (50.0)		72 (32.4)	54 (31.8)	
Unknown	1 (1.3)	1 (6.2)		2 (0.9)	0 (0.0)	

*, denotes 'year of initial pathologic diagnosis'; [#], included AJCC stage for most cancer types except for CESC, OV, UCEC, and UCS, for which clinical stage was applied; [&], P value denotes statistically significant. ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; OV, ovarian serous cystadenocarcinoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma.

Table S2 Tumor mutational burden (TMB) threshold and baseline patient characteristics in TMB-high and TMB-low subsets across 20 cancer types (CESC, CHOL)

Variable	CESC			CHOL		
	TMB-low	TMB-high	P	TMB-low	TMB-high	P
Threshold, mut/Mb (percentile)	5.84 (74th)			3.42 (86th)		
N	210	77		31	5	
Age (years), median [IQR]	45.00 [38.00–55.00]	51.00 [43.00–63.00]	0.011[&]	64.00 [56.00–71.00]	78.00 [67.00–81.00]	0.074
Diagnosis*, median [IQR]	2010.00 [2006.00–2012.00]	2010.00 [2007.00–2012.00]	0.705	2011.00 [2009.50–2012.00]	2010.00 [2010.00–2010.00]	0.469
Sex (%)			NA			0.149
Female	210 (100.0)	77 (100.0)		19 (61.3)	1 (20.0)	
Male	0 (0.0)	0 (0.0)		12 (38.7)	4 (80.0)	
Race (%)			0.546			1.000
White	145 (69.0)	50 (64.9)		26 (83.9)	5 (100.0)	
Black	19 (9.0)	10 (13.0)		2 (6.5)	0 (0.0)	
Other	23 (11.0)	6 (7.8)		3 (9.7)	0 (0.0)	
Unknown	23 (11.0)	11 (14.3)		0 (0.0)	0 (0.0)	
Stage[#] (%)			0.167			1.000
Stage I	122 (58.1)	33 (42.9)		16 (51.6)	3 (60.0)	
Stage II	43 (20.5)	21 (27.3)		8 (25.8)	1 (20.0)	
Stage III	29 (13.8)	13 (16.9)		1 (3.2)	0 (0.0)	
Stage IV	12 (5.7)	7 (9.1)		6 (19.4)	1 (20.0)	
Unknown	4 (1.9)	3 (3.9)		0 (0.0)	0 (0.0)	

*, denotes 'year of initial pathologic diagnosis'; [#], included AJCC stage for most cancer types except for CESC, OV, UCEC, and UCS, for which clinical stage was applied; [&], P value denotes statistically significant. CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangiocarcinoma; OV, ovarian serous cystadenocarcinoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma.

Table S3 Tumor mutational burden (TMB) threshold and baseline patient characteristics in TMB-high and TMB-low subsets across 20 cancer types (COAD, ESCA)

Variable	COAD			ESCA		
	TMB-low	TMB-high	P	TMB-low	TMB-high	P
Threshold, mut/Mb (percentile)	34.66 (87th)			6.74 (82nd)		
N	348	55		145	32	
Age (years), median [IQR]	68.00 [57.00–77.00]	73.00 [62.00–80.50]	0.009[&]	59.00 [53.00–69.00]	73.00 [65.75–77.00]	<0.001[&]
Diagnosis*, median [IQR]	2009.00 [2007.00–2010.00]	2009.00 [2006.00–2010.00]	0.766	2011.00 [2009.00–2012.00]	2009.50 [2001.00–2012.00]	0.016[&]
Sex (%)			0.472			0.588
Female	165 (47.4)	29 (52.7)		21 (14.5)	6 (18.8)	
Male	183 (52.6)	26 (47.3)		124 (85.5)	26 (81.2)	
Race (%)			0.445			0.069
White	175 (50.3)	30 (54.5)		89 (61.4)	22 (68.8)	
Black	50 (14.4)	5 (9.1)		3 (2.1)	0 (0.0)	
Other	9 (2.6)	3 (5.5)		42 (29.0)	4 (12.5)	
Unknown	114 (32.8)	17 (30.9)		11 (7.6)	6 (18.8)	
Stage[#] (%)			<0.001[&]			0.602
Stage I	62 (17.8)	6 (10.9)		14 (9.7)	3 (9.4)	
Stage II	118 (33.9)	38 (69.1)		62 (42.8)	15 (46.9)	
Stage III	106 (30.5)	8 (14.5)		43 (29.7)	12 (37.5)	
Stage IV	53 (15.2)	1 (1.8)		8 (5.5)	1 (3.1)	
Unknown	9 (2.6)	2 (3.6)		18 (12.4)	1 (3.1)	

*, denotes 'year of initial pathologic diagnosis'; [#], included AJCC stage for most cancer types except for CESC, OV, UCEC, and UCS, for which clinical stage was applied; [&], P value denotes statistically significant. COAD, colon adenocarcinoma; ESCA, esophageal carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; OV, ovarian serous cystadenocarcinoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma.

Table S4 Tumor mutational burden (TMB) threshold and baseline patient characteristics in TMB-high and TMB-low subsets across 20 cancer types (HNSC, KIRC)

Variable	HNSC			KIRC		
	TMB-low	TMB-high	P	TMB-low	TMB-high	P
Threshold, mut/Mb (percentile)	6.39 (80th)			2.58 (90th)		
N	401	106		329	40	
Age (years), median [IQR]	60.00 [53.00–69.00]	62.50 [56.25–68.75]	0.059	59.00 [51.00–67.00]	71.50 [60.50–76.00]	<0.001[&]
Diagnosis*, median [IQR]	2010.00 [2007.00–2012.00]	2008.00 [1999.25–2010.00]	<0.001[&]	2006.00 [2005.00–2008.00]	2007.00 [2004.00–2010.00]	0.179
Sex (%)			0.624			1.000
Female	107 (26.7)	31 (29.2)		122 (37.1)	15 (37.5)	
Male	294 (73.3)	75 (70.8)		207 (62.9)	25 (62.5)	
Race (%)			0.800			0.008[&]
White	345 (86.0)	89 (84.0)		277 (84.2)	26 (65.0)	
Black	37 (9.2)	10 (9.4)		40 (12.2)	13 (32.5)	
Other	9 (2.2)	3 (2.8)		7 (2.1)	0 (0.0)	
Unknown	10 (2.5)	4 (3.8)		5 (1.5)	1 (2.5)	
Stage[#] (%)			0.468			0.235
Stage I	23 (5.7)	2 (1.9)		169 (51.4)	27 (67.5)	
Stage II	53 (13.2)	18 (17.0)		40 (12.2)	1 (2.5)	
Stage III	64 (16.0)	15 (14.2)		75 (22.8)	7 (17.5)	
Stage IV	207 (51.6)	56 (52.8)		43 (13.1)	5 (12.5)	
Unknown	54 (13.5)	15 (14.2)		2 (0.6)	0 (0.0)	

*, denotes 'year of initial pathologic diagnosis'; [#], included AJCC stage for most cancer types except for CESC, OV, UCEC, and UCS, for which clinical stage was applied; [&], P value denotes statistically significant. HNSC, head and neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; OV, ovarian serous cystadenocarcinoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma.

Table S5 Tumor mutational burden (TMB) threshold and baseline patient characteristics in TMB-high and TMB-low subsets across 20 cancer types (KIRP, LIHC)

Variable	KIRP			LIHC		
	TMB-low	TMB-high	P	TMB-low	TMB-high	P
Threshold, mut/Mb (percentile)	2.42 (61st)			3.58 (69th)		
N	157	100		247	113	
Age (years), median [IQR]	59.00 [51.00–67.00]	66.00 [57.75–74.00]	<0.001[&]	59.00 [50.00–68.00]	64.00 [57.00–70.00]	<0.001[&]
Diagnosis*, median [IQR]	2011.00 [2008.00–2012.00]	2011.00 [2008.00–2012.00]	0.486	2011.00 [2008.00–2012.50]	2011.00 [2009.00–2012.00]	0.731
Sex (%)			0.023[&]			0.054
Female	53 (33.8)	20 (20.0)		89 (36.0)	29 (25.7)	
Male	104 (66.2)	80 (80.0)		158 (64.0)	84 (74.3)	
Race (%)			0.075			0.735
White	105 (66.9)	80 (80.0)		128 (51.8)	52 (46.0)	
Black	43 (27.4)	15 (15.0)		11 (4.5)	5 (4.4)	
Other	6 (3.8)	2 (2.0)		103 (41.7)	54 (47.8)	
Unknown	3 (1.9)	3 (3.0)		5 (2.0)	2 (1.8)	
Stage[#] (%)			0.936			0.361
Stage I	93 (59.2)	61 (61.0)		123 (49.8)	47 (41.6)	
Stage II	10 (6.4)	7 (7.0)		49 (19.8)	32 (28.3)	
Stage III	27 (17.2)	19 (19.0)		57 (23.1)	27 (23.9)	
Stage IV	10 (6.4)	5 (5.0)		3 (1.2)	2 (1.8)	
Unknown	17 (10.8)	8 (8.0)		15 (6.1)	5 (4.4)	

*, denotes 'year of initial pathologic diagnosis'; [#], included AJCC stage for most cancer types except for CESC, OV, UCEC, and UCS, for which clinical stage was applied; [&], P value denotes statistically significant. KIRP, kidney renal papillary cell carcinoma; LIHC, liver hepatocellular carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; OV, ovarian serous cystadenocarcinoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma.

Table S6 Tumor mutational burden (TMB) threshold and baseline patient characteristics in TMB-high and TMB-low subsets across 20 cancer types (LUAD, LUSC)

Variable	LUAD			LUSC		
	TMB-low	TMB-high	P	TMB-low	TMB-high	P
Threshold, mut/Mb (percentile)	20.58 (90th)			14.08 (88th)		
N	444	52		400	58	
Age (years), median [IQR]	67.00 [59.00–73.00]	61.50 [54.75–70.00]	0.016[&]	68.00 [61.00–74.00]	66.50 [62.00–73.00]	0.610
Diagnosis*, median [IQR]	2010.00 [2007.00–2011.00]	2010.00 [2007.00–2011.00]	0.843	2009.00 [2005.00–2011.00]	2008.50 [2004.00–2011.00]	0.250
Sex (%)			0.883			0.429
Female	239 (53.8)	27 (51.9)		104 (26.0)	18 (31.0)	
Male	205 (46.2)	25 (48.1)		296 (74.0)	40 (69.0)	
Race (%)			0.250			0.337
White	351 (79.1)	37 (71.2)		290 (72.5)	49 (84.5)	
Black	43 (9.7)	9 (17.3)		26 (6.5)	2 (3.4)	
Other	9 (2.0)	0 (0.0)		9 (2.2)	0 (0.0)	
Unknown	41 (9.2)	6 (11.5)		75 (18.8)	7 (12.1)	
Stage[#] (%)			0.085			0.123
Stage I	235 (52.9)	31 (59.6)		189 (47.2)	33 (56.9)	
Stage II	100 (22.5)	17 (32.7)		129 (32.2)	14 (24.1)	
Stage III	77 (17.3)	3 (5.8)		73 (18.2)	9 (15.5)	
Stage IV	25 (5.6)	1 (1.9)		7 (1.8)	0 (0.0)	
Unknown	7 (1.6)	0 (0.0)		2 (0.5)	2 (3.4)	

*, denotes 'year of initial pathologic diagnosis'; [#], included AJCC stage for most cancer types except for CESC, OV, UCEC, and UCS, for which clinical stage was applied; [&], P value denotes statistically significant. LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; OV, ovarian serous cystadenocarcinoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma.

Table S7 Tumor mutational burden (TMB) threshold and baseline patient characteristics in TMB-high and TMB-low subsets across 20 cancer types (MESO, OV)

Variable	MESO			OV		
	TMB-low	TMB-high	P	TMB-low	TMB-high	P
Threshold, mut/Mb (percentile)	1.34 (88th)			2.84 (71st)		
N	71	10		287	121	
Age (years), median [IQR]	64.00 [57.00–68.50]	65.50 [60.25–68.50]	0.937	58.00 [50.00–67.50]	61.00 [53.00–70.00]	0.125
Diagnosis*, median [IQR]	2011.00 [2008.00–2012.00]	2009.00 [2005.00–2010.00]	0.090	2005.00 [2002.00–2008.00]	2004.00 [2000.00–2008.00]	0.156
Sex (%)			0.383			NA
Female	12 (16.9)	3 (30.0)		287 (100.0)	121 (100.0)	
Male	59 (83.1)	7 (70.0)		0 (0.0)	0 (0.0)	
Race (%)			1.000			0.925
White	69 (97.2)	10 (100.0)		243 (84.7)	104 (86.0)	
Black	1 (1.4)	0 (0.0)		19 (6.6)	7 (5.8)	
Other	1 (1.4)	0 (0.0)		13 (4.5)	4 (3.3)	
Unknown	0 (0.0)	0 (0.0)		12 (4.2)	6 (5.0)	
Stage[#] (%)			0.955			0.064
Stage I	8 (11.3)	1 (10.0)		6 (2.1)	9 (7.4)	
Stage II	14 (19.7)	2 (20.0)		13 (4.5)	9 (7.4)	
Stage III	34 (47.9)	6 (60.0)		222 (77.4)	85 (70.2)	
Stage IV	15 (21.1)	1 (10.0)		44 (15.3)	17 (14.0)	
Unknown	0 (0.0)	0 (0.0)		2 (0.7)	1 (0.8)	

*, denotes 'year of initial pathologic diagnosis'; [#], included AJCC stage for most cancer types except for CESC, OV, UCEC, and UCS, for which clinical stage was applied. MESO, mesothelioma; OV, ovarian serous cystadenocarcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; OV, ovarian serous cystadenocarcinoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma.

Table S8 Tumor mutational burden (TMB) threshold and baseline patient characteristics in TMB-high and TMB-low subsets across 20 cancer types (PAAD, SKCM)

Variable	PAAD			SKCM		
	TMB-low	TMB-high	P	TMB-low	TMB-high	P
Threshold, mut/Mb (percentile)	1.47 (77th)			22.84 (71st)		
N	134	42		316	132	
Age (years), median [IQR]	65.00 [57.25–73.00]	66.00 [54.50–75.00]	0.934	58.00 [48.00–70.25]	60.00 [48.00–72.00]	0.386
Diagnosis*, median [IQR]	2012.00 [2010.25–2012.00]	2011.50 [2010.00–2012.00]	0.645	2009.00 [2003.75–2012.00]	2006.00 [2001.00–2009.00]	<0.001[&]
Sex (%)			0.050			0.001[&]
Female	66 (49.3)	13 (31.0)		134 (42.4)	33 (25.0)	
Male	68 (50.7)	29 (69.0)		182 (57.6)	99 (75.0)	
Race (%)			0.295			0.041[&]
White	120 (89.6)	34 (81.0)		299 (94.6)	130 (98.5)	
Black	5 (3.7)	2 (4.8)		0 (0.0)	0 (0.0)	
Other	7 (5.2)	4 (9.5)		12 (3.8)	0 (0.0)	
Unknown	2 (1.5)	2 (4.8)		5 (1.6)	2 (1.5)	
Stage[#] (%)			0.165			0.296
Stage I	17 (12.7)	2 (4.8)		52 (16.5)	25 (18.9)	
Stage II	109 (81.3)	36 (85.7)		98 (31.0)	40 (30.3)	
Stage III	3 (2.2)	1 (2.4)		120 (38.0)	47 (35.6)	
Stage IV	2 (1.5)	3 (7.1)		19 (6.0)	3 (2.3)	
Unknown	3 (2.2)	0 (0.0)		27 (8.5)	17 (12.9)	

*, denotes 'year of initial pathologic diagnosis'; [#], included AJCC stage for most cancer types except for CESC, OV, UCEC, and UCS, for which clinical stage was applied; [&], P value denotes statistically significant. PAAD, pancreatic adenocarcinoma; SKCM, skin cutaneous melanoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; OV, ovarian serous cystadenocarcinoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma.

Table S9 Tumor mutational burden (TMB) threshold and baseline patient characteristics in TMB-high and TMB-low subsets across 20 cancer types (STAD, UCEC)

Variable	STAD			UCEC		
	TMB-low	TMB-high	P	TMB-low	TMB-high	P
Threshold, mut/Mb (percentile)	26.68 (88th)			9.74 (65th)		
N	374	55		337	185	
Age (years), median [IQR]	66.00 [57.25–72.00]	71.00 [66.00–77.50]	<0.001[&]	65.00 [59.00–72.00]	60.00 [55.00–69.00]	0.001[&]
Diagnosis*, median [IQR]	2011.00 [2010.00–2012.00]	2011.00 [2008.00–2011.00]	0.040[&]	2009.00 [2008.00–2010.00]	2009.00 [2007.00–2010.00]	0.173
Sex (%)			0.017[&]			NA
Female	127 (34.0)	28 (50.9)		337 (100.0)	185 (100.0)	
Male	247 (66.0)	27 (49.1)		0 (0.0)	0 (0.0)	
Race (%)			0.250			0.287
White	243 (65.0)	30 (54.5)		233 (69.1)	127 (68.6)	
Black	11 (2.9)	2 (3.6)		73 (21.7)	32 (17.3)	
Other	77 (20.6)	12 (21.8)		18 (5.3)	15 (8.1)	
Unknown	43 (11.5)	11 (20.0)		13 (3.9)	11 (5.9)	
Stage[#] (%)			0.088			0.047[&]
Stage I	43 (11.5)	11 (20.0)		195 (57.9)	129 (69.7)	
Stage II	112 (29.9)	16 (29.1)		35 (10.4)	17 (9.2)	
Stage III	163 (43.6)	16 (29.1)		86 (25.5)	33 (17.8)	
Stage IV	36 (9.6)	6 (10.9)		21 (6.2)	6 (3.2)	
Unknown	20 (5.3)	6 (10.9)		0 (0.0)	0 (0.0)	

*, denotes 'year of initial pathologic diagnosis'; [#], included AJCC stage for most cancer types except for CESC, OV, UCEC, and UCS, for which clinical stage was applied; [&], P value denotes statistically significant. STAD, stomach adenocarcinoma; UCEC, uterine corpus endometrial carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; OV, ovarian serous cystadenocarcinoma; UCS, uterine carcinosarcoma.

Table S10 Tumor mutational burden (TMB) threshold and baseline patient characteristics in TMB-high and TMB-low subsets across 20 cancer types (UCS, UVM)

Variable	UCS			UVM		
	TMB-low	TMB-high	P	TMB-low	TMB-high	P
Threshold, mut/Mb (percentile)	1.79 (79th)			0.34 (54th)		
N	45	12		43	37	
Age (years), median [IQR]	68.00 [62.00–76.00]	69.00 [66.50–76.50]	0.590	57.00 [50.50–67.00]	68.00 [54.00–75.00]	0.106
Diagnosis*, median [IQR]	2009.00 [2007.00–2011.00]	2011.00 [2007.75–2012.00]	0.188	2012.00 [2011.00–2012.00]	2012.00 [2012.00–2013.00]	0.013[&]
Sex (%)			NA			0.655
Female	45 (100.0)	12 (100.0)		20 (46.5)	15 (40.5)	
Male	0 (0.0)	0 (0.0)		23 (53.5)	22 (59.5)	
Race (%)			0.489			0.629
White	36 (80.0)	8 (66.7)		31 (72.1)	24 (64.9)	
Black	6 (13.3)	3 (25.0)		0 (0.0)	0 (0.0)	
Other	2 (4.4)	1 (8.3)		0 (0.0)	0 (0.0)	
Unknown	1 (2.2)	0 (0.0)		12 (27.9)	13 (35.1)	
Stage[#] (%)			0.314			0.065
Stage I	19 (42.2)	3 (25.0)		0 (0.0)	0 (0.0)	
Stage II	4 (8.9)	1 (8.3)		17 (39.5)	22 (59.5)	
Stage III	13 (28.9)	7 (58.3)		24 (55.8)	12 (32.4)	
Stage IV	9 (20.0)	1 (8.3)		1 (2.3)	3 (8.1)	
Unknown	0 (0.0)	0 (0.0)		1 (2.3)	0 (0.0)	

*, denotes 'year of initial pathologic diagnosis'; [#], included AJCC stage for most cancer types except for CESC, OV, UCEC, and UCS, for which clinical stage was applied; [&], P value denotes statistically significant. UCS, uterine carcinosarcoma; UVM, uveal melanoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; OV, ovarian serous cystadenocarcinoma; UCEC, uterine corpus endometrial carcinoma.

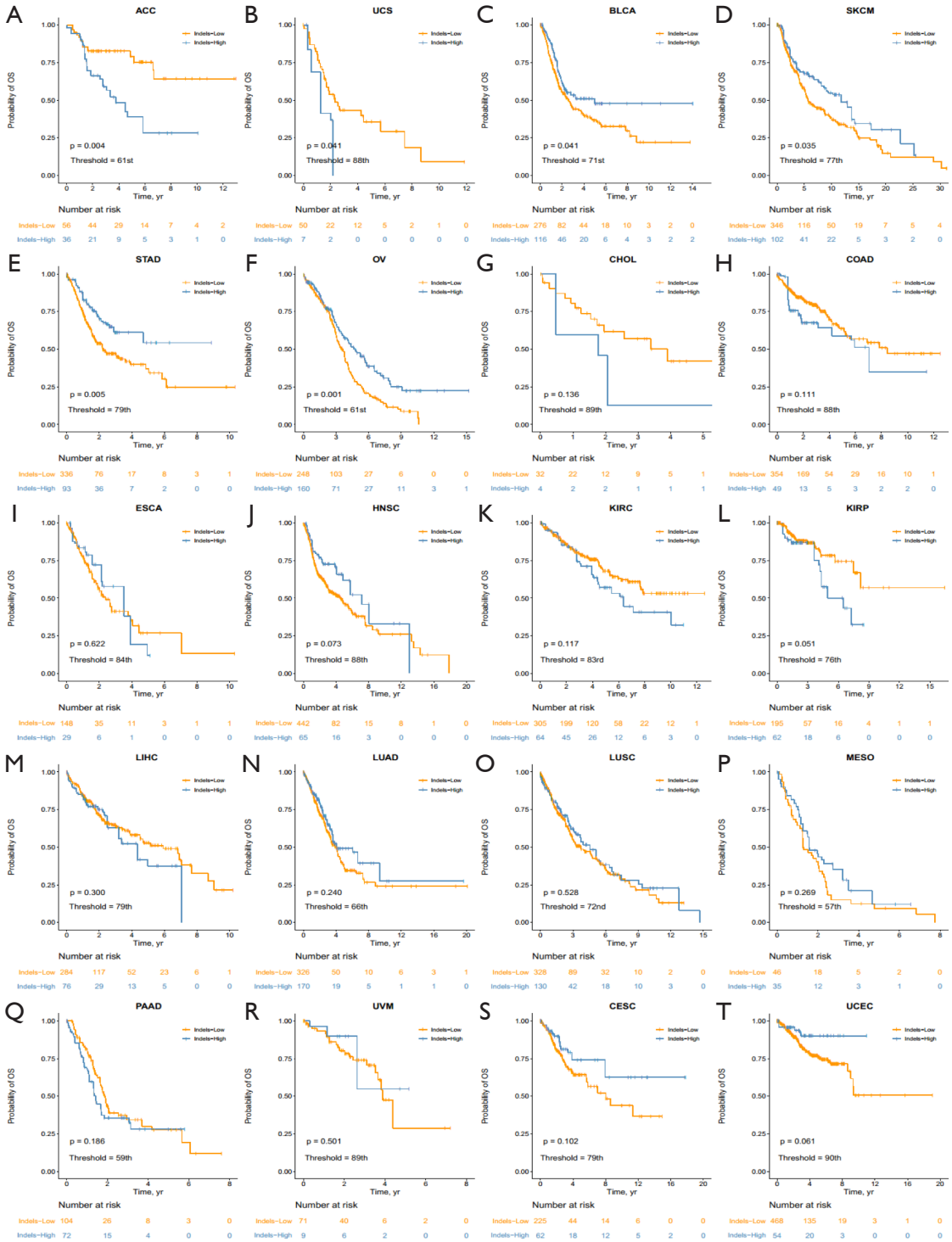

Figure S1 The inverse probability of treatment weighted Kaplan-Meier survival curves of the impact of indel burden on overall survival. High indel burden was associated with worse survival in adrenocortical carcinoma (A) and uterine carcinosarcoma (B). While high indel burden predicted better prognosis in bladder urothelial carcinoma (C), skin cutaneous melanoma (D), stomach adenocarcinoma (E) and ovarian serous cystadenocarcinoma (F). In cholangiocarcinoma (G), colon adenocarcinoma (H), esophageal carcinoma (I), head and neck squamous cell carcinoma (J), kidney renal clear cell carcinoma (K), kidney renal papillary cell carcinoma (L), liver hepatocellular carcinoma (M), lung adenocarcinoma (N), lung squamous cell carcinoma (O), mesothelioma (P), pancreatic adenocarcinoma (Q), uveal melanoma (R), cervical squamous cell carcinoma and endocervical adenocarcinoma (S) and uterine corpus endometrial carcinoma (T), indel burden did not have a significant impact on overall survival.