

Systematic assessment and optimizing algorithm of tumor mutational burden calculation and their implications in clinical decision-making

Daqiang Sun (✉ sdqmd@163.com)

Tianjin Chest Hospital, Affiliated Chest Hospital of Tianjin University

Meilin Xu

Tianjin Chest Hospital, Affiliated Chest Hospital of Tianjin University

Chaohu Pan

Jinan University

Hongzhen Tang

YuceBio Technology Co., Ltd

Dongfang Wu

YuceBio Technology Co., Ltd

Haitao Luo

YuceBio Technology Co., Ltd

Research Article

Keywords: immunotherapy, tumor mutation burden, East Asian populations, method optimizing, clinical decision-making

Posted Date: June 16th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1745663/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Tumor mutation burden (TMB) has been validated as a biomarker to predict the response of immune checkpoint inhibitors (ICIs) treatment in various cancers. However, the effects of different sequencing platforms, cancer types, and calculation algorithms on TMB as well as its cut-off value for predicting immunotherapy efficacy in the East Asian population still need to be further investigated. In this study, the data of 4126 samples generated by targeted panel sequencing or whole-exome sequencing (WES) in different platforms and public sequencing data from 3680 samples that contained targeted panel sequencing, WES and whole-genome sequencing (WGS) were obtained. The impact of different sequencing platforms and methods on TMB calculation was assessed. No significant bias was found in TMB calculated by different platforms. However, TMB calculated from WGS was significantly lower than those calculated from targeted panel sequencing and WES. The distribution of TMB at different sequencing depths and tumor purity were analyzed. There was no significant difference in the distribution of TMB when the sequencing depth was greater than 500, the tumor purity estimated by hematoxylin-eosin (HE) staining was between 0.1 to 0.8 or estimated by next-generation sequencing (NGS) was greater than 0.4. In addition, the somatic-germline-zygosity (SGZ) algorithm was optimized to calculate TMB from tumor-only sequencing samples in the East Asian population. The correlation coefficient of TMB calculated with the optimized SGZ algorithm and paired normal-tumor sequencing is 0.95. Furthermore, the optimal cut-off value of TMB in East Asian lung cancer patients treated with ICIs was determined to be 7.19 mut/Mb instead of 10 mut/Mb through the ROC curve and Log-rank analysis in the training cohort and validated in the test cohort. Patients with TMB \geq 7.19 mut/Mb had better outcomes than patients with TMB $<$ 7.19 mut/Mb. In conclusion, this study systematically analyzed the factors that influenced the TMB calculation and optimized the SGZ algorithm to calculate TMB from tumor-only sequencing samples in the East Asian population. More importantly, the cut-off value of TMB for predicting immunotherapy efficacy was determined to be 7.19 mut/Mb instead of 10 mut/Mb in East Asian lung cancer patients, which can help in clinical decision-making.

Introduction

The advent of immune checkpoint inhibitors (ICIs), including anti-programmed cell death protein 1 (PD-1)/programmed death-ligand 1 (PD-L1) and anti-cytotoxic T lymphocyte-associated antigen-4 (CTLA4), have revolutionized cancer therapy [1-5]. Several ICIs have been approved by Food and Drug Administration (FDA) in multiple tumor types [6]. However, only a subset of patients achieved durable clinical responses, and some may even suffer from unique immune-related toxicities or even hyperprogression [7-11]. Therefore, predictive biomarkers were urgently required to optimize the treatment of ICIs.

The expression of PD-L1 in tumor and/or tumor-infiltrating lymphocytes assessed by immunohistochemistry (IHC) is an established biomarker to predict efficacy in the treatment of ICIs across many cancer types including melanoma, non-small cell lung cancer (NSCLC) and colorectal cancer [10, 12]. However, on the one hand, PD-L1 alone as a biomarker is insufficient to distinguish

responders [13]; on the other hand, the detection methods and thresholds for PD-L1 expression are variable [14, 15]. Therefore, new biomarkers are required to improve the treatment decision-making and identify potential responders from ICIs therapy.

Tumor mutation burden (TMB), which was defined as the number of all non-synonymous somatic mutations per megabase based on the genome examined, has been reported to predict the efficacy of ICIs therapy in multiple tumor types [16-18]. The more mutations, the more neoantigens are produced, which ultimately activate the stronger antitumor immune response [19]. Based on the results from the phase 2 KEYNOTE-158 trial, pembrolizumab was approved by the FDA for patients with TMB ≥ 10 mut/Mb. Patients with a TMB ≥ 10 mut/Mb were defined as TMB-high and associated with better response rates [20]. However, there were racial differences in TMB across multiple cancer types [21]. Compared with European and American populations, the TMB is lower in East Asian populations. The cut-off of 10 mut/Mb may lead to fewer East Asian populations meeting eligibility for the treatment of ICIs. Therefore, the association of TMB cut-off in East Asian populations with ICIs treatment outcomes needs to be further investigated.

Currently, multiple platforms and sequencing methods have been used for next-generation sequencing (NGS) [22]. However, the effect of different sequencing platforms and methods on TMB calculation has not been systematically evaluated. Compared with sequencing both tumor and matched normal specimens, tumor-only sequencing could reduce time and cost. In addition, many clinical tumor samples lack matching normal tissue, which requires the development of algorithms to calculate TMB for these samples. However, the current algorithms, including somatic-germline-zygosity (SGZ), were designed based on European and American populations. For the East Asian population, the corresponding algorithm is lacking [23-25]. Therefore, an algorithm for TMB calculation from tumor-only sequencing samples in the East Asian population was urgently needed.

In this study, the effect of different sequencing platforms and methods on the calculation of TMB has been systematically evaluated. To calculate TMB for Asian patients with tumor-only sequencing, we optimized the SGZ algorithm and demonstrated the reliability of calculating TMB from tumor-only sequencing samples. Furthermore, the cut-off of TMB in Asian patients was determined and its efficacy in the treatment of ICIs has been investigated.

Materials And Methods

Samples and datasets

Data for 4126 samples sequenced with different sequencing platforms and methods were obtained from a CAP-accredited laboratory (YuceBio Technology Co., Ltd, China) (Table S1). To investigate the racial differences in TMB value, 3680 genomic data of European and American populations were collected from cBioPortal [26-29].

Sixty-two samples with matched control were retrospectively obtained to analyze the correlation coefficient of TMB calculated from methods of tumor-only sequencing and paired normal-tumor sequencing (Table S2 to S4). To determine the cut-off of TMB for predicting immunotherapy efficacy in the East Asian population, tumor samples of sixty-six lung patients treated with ICIs between July 2019 to September 2020 were retrospectively collected as a training cohort and sequenced without normal control (Table S5). Furthermore, genomic and clinical data of Sixty-nine East Asian NSCLC patients subjected to ICIs treatments were obtained as a test cohort to validate the cut-off of TMB [30]. Durable clinical benefit (DCB) was defined as complete response, partial response, or stable disease (SD) that lasted for ≥ 24 weeks, and non-durable benefit (NDB) was defined as SD that lasted for < 24 weeks or progressive disease.

Next-generation sequencing (NGS) and mutation analysis

Genomic profiling was implemented on tumor tissues and matched peripheral blood samples. The GeneReadDNA FFPE kit (Qiagen) and Qiagen DNA blood mini kit (Qiagen) were used to extract DNA from tumor specimens and blood, respectively. For tumor-only sequencing, DNA from the tumor sample was extracted with a GeneReadDNA FFPE kit (Qiagen). DNA quantification was performed with the dsDNA HS Assay Kit (ThermoFisher Scientific, USA). For the platform of Illumina, sequencing libraries were built by SureSelect XT Human All Exon V6 (Agilent) for WES or a customized next-generation sequencing panel targeting exons of 1267 genes for panel sequencing, respectively. Sequencing procedures were utilized by the NextSeq 550AR platform with 150-bp paired-end reads. For the platform of MGI, sequencing libraries were built by Exome Plus Panel V1.0 (IDT, USA) for WES or a customized next-generation sequencing panel targeting exons of 1267 genes for panel sequencing, respectively. Sequencing procedures were utilized by the MGISEQ-T7 platform with 100-bp paired-end reads.

Sequencing reads with $> 10\%$ N rate and/or $> 10\%$ bases with a quality score of < 20 were filtered using SOAPnuke (Version 1.5.6) [31]. Somatic single nucleotide variants and insertions and deletions (indels) were detected using VarScan (Version 2.4) [32]. Next, an in-house method was applied to filter possible false-positive mutations. Finally, SnpEff (Version 4.3) was used to functionally annotate the mutations detected in the tumor samples [33].

TMB was determined as the number of all nonsynonymous mutations and indels per megabase of the genome examined.

Tumor purity estimation

To estimate the tumor purity by hematoxylin-eosin (HE) staining, the sample was fixed in the 10% formalin solution, embedded in paraffin. Then the 5 μm slide was stained with HE. The tumor purity is the value of tumor cells divided by all cells. To estimate the tumor purity by NGS, the sequencing reads were quality controlled using SOAPnuke (Version 1.5.6) [31], then aligned to the reference genome using BWA (v0.7.12). The tumor purity was estimated by Ascatngs (v3.1.0) [34].

SGZ optimization and mutation analysis

Mutations from tumor-only sequencing samples were identified by the somatic-germline-zygosity (SGZ) algorithm. For each sample, MPS variant analysis was executed to create a genome-wide copy number profile, which is segmented and modeled to estimate the ploidy (C) and overall tumor purity (p), as well as per segment copy number (C) and minor allele count (M). The log-ratio of variants was defined by the following formula:

$$\text{Logratio} = \log_2\left(\frac{\text{Purity} * \text{CopyNumber} + 2 * (1 - \text{Purity})}{\text{Purity} * \text{Ploidy} + 2 * (1 - \text{Purity})}\right)$$

For each variant, the error log ratio was obtained by calculating the absolute value of the difference in log ratio between variant and segments. Finally, the germline variant or somatic variant was identified mainly by frequency, purity and error log ratio. However, the cut-off value of the above parameter was fit to European and American populations, which resulted in a high false-positive rate in the East Asian population [24]. In order to calculate mutation from tumor-only sequencing in the East Asian population, the SGZ algorithm was optimized as followed: (1) generating a mutation background library based on the East Asian population, (2) analyzing mutations with SGZ, (3) filtering out variants that appear more than 5 times in background library, while variants with a frequency higher than 0.9 were retained.

Statistical methods

All statistical analyses were implemented in Python (3.10.1). An independent *t*-test was used to compare TMB values between different groups. Correlation analysis was performed using the Pearson correlation analysis. Roc-curve and Log-rank test analyses were conducted to determine the cutoff of TMB. Categorical variables were evaluated with the Fisher-exact test. Kaplan-Meier curve, Log-rank test, and Cox regression were used to determine the significance of TMB on overall survival (OS) and Progression-Free-Survival (PFS). Statistical significance was set at *p*-value < 0.05.

Results

Effects of different sequencing platforms, sequencing methods and races on TMB values

To study the effect of different sequencing platforms and methods on TMB calculation, data of 4126 tumor samples sequenced with different platforms and methods were obtained. As shown in Figure 1A, no significant difference in TMB calculation from different sequencing platforms, including Illumina and MGI, was found. TMB from panel sequencing was higher than whole-exome sequencing (WES), however, there was no significant difference. To further verify the effect of different sequencing methods on TMB calculation, public sequencing data from 3680 samples performed with different methods were analyzed

(Figure 1B). TMB values calculated from whole-genome sequencing (WGS) were significantly lower than those calculated from WES and panel sequencing, which suggested that the TMB values may be affected by sequencing depth. Furthermore, TMB in different races was analyzed. As shown in Figure 1C, TMB values of the East Asian populations were significantly lower than that of European and American populations in both WES sequencing and panel sequencing.

TMB calculation was affected by the sequencing depth and tumor purity

To investigate the effect of sequencing depth on TMB calculation, TMB calculated from lung cancer at different panel sequencing depths were analyzed. As shown in Figure 2A, the TMB calculated from sequencing depths ≥ 500 was significantly higher than that calculated from sequencing depths < 500 . To determine the effect of tumor purities on TMB calculation, the distribution of TMB values with different tumor purities was analyzed. As shown in Figures 2B and 2C, the TMB values were more stable with tumor purity estimated by HE staining between 0.2-0.8 or tumor purity estimated by NGS ≥ 0.4 .

High correlation of TMB calculated with the optimized SGZ algorithm and paired normal-tumor sequencing in East Asian populations

TMB calculation requires paired normal samples to remove germline mutations, which increases the cost of sequencing. In addition, in terms of clinical accessibility, paired normal samples are sometimes unavailable, which limits the clinical application of TMB. The SGZ algorithm for TMB calculation with tumor-only sequencing samples was designed based on European and American populations. For East Asian populations, the TMB calculated by the SGZ algorithm had a low correlation with the TMB calculated by the method of paired normal-tumor sequencing (Figure 3A). In order to calculate TMB from tumor-only sequencing samples in Asian populations, the SGZ algorithm was optimized. As shown in Figure 3B, we added the mutation filtering step, and constructed a background library with normal samples from East Asian patients to filter germline mutations, which ultimately reduced the false positives of TMB. To verify the accuracy of TMB calculation with an optimized algorithm, TMB calculated from the methods of the optimized algorithm and paired normal-tumor sequencing in sixty-two samples were compared. As shown in Figures 3C and 3D, their correlation coefficient is 0.95 and 82.7% of the mutations identified from the method of paired normal-tumor sequencing could be identified with the optimized SZG algorithm. These results demonstrate the accuracy of TMB calculation from the tumor-only sequencing with the optimized SGZ algorithm in East Asian populations.

Identification of the TMB cut-off for predicting immunotherapy efficacy in the training cohort

The cut-off of TMB for predicting the efficacy of immunotherapy in European and American populations is 10 mut/Mb [20]. However, since the TMB of East Asian populations is lower than that of European and American populations, a TMB cut-off of 10 mut/Mb may not be suitable for East Asian populations. To determine the cut-off of TMB for predicting immunotherapy efficacy in the East Asian population,

tumor samples from sixty-six lung patients treated with ICIs were retrospectively collected as a training cohort. TMB was calculated with the optimized SGZ algorithm. The performance of TMB for predicting patient durable clinical benefit was analyzed with a ROC curve and Log-rank test. As shown in Figure 4A, the optimal cut-off of TMB was 7.19 mut/Mb with AUC = 0.742, and validated with Log-rank analysis (Figure 4B). Furthermore, the response rate and survival period were higher in patients with TMB \geq 7.19 mut/Mb than in those with TMB < 7.19 mut/Mb, and the TMB cut-off of 7.19 mut/Mb is better than the TMB cut-off of 10 mut/Mb in East Asian populations (Figure 4C-4F). To further investigate the role of TMB in predicting the efficacy of immunotherapy, the effects of TMB, medication type, tumor type and age group on patient survival were analyzed through multi-factor cox-regression. It was found that TMB was a favorable factor for patient survival, while other factors had no significant effect (Figure 4G).

Validation of TMB cut-off in the test cohort

To further validate the TMB cut-off of 7.19 mut/Mb, genomic and clinical data of sixty-nine East Asian NSCLC patients treated with ICIs were collected [30]. Consistent with the above results, the survival period of patients with TMB \geq 7.19 mut/Mb was longer than those with TMB < 7.19 mut/Mb (Figure 5A). Furthermore, the predicting efficacy with the TMB cut-off of 7.19 mut/Mb is better than the TMB cut-off of 10 mut/Mb (Figure 5B).

Discussion

Effects of different sequencing platforms, sequencing methods and races on TMB values

To study the effect of different sequencing platforms and methods on TMB calculation, data of 4126 tumor samples sequenced with different platforms and methods were obtained. As shown in Figure 1A, no significant difference in TMB calculation from different sequencing platforms, including Illumina and MGI, was found. TMB from panel sequencing was higher than whole-exome sequencing (WES), however, there was no significant difference. To further verify the effect of different sequencing methods on TMB calculation, public sequencing data from 3680 samples performed with different methods were analyzed (Figure 1B). TMB values calculated from whole-genome sequencing (WGS) were significantly lower than those calculated from WES and panel sequencing, which suggested that the TMB values may be affected by sequencing depth. Furthermore, TMB in different races was analyzed. As shown in Figure 1C, TMB values of the East Asian populations were significantly lower than that of European and American populations in both WES sequencing and panel sequencing.

TMB calculation was affected by the sequencing depth and tumor purity

To investigate the effect of sequencing depth on TMB calculation, TMB calculated from lung cancer at different panel sequencing depths were analyzed. As shown in Figure 2A, the TMB calculated from sequencing depths \geq 500 was significantly higher than that calculated from sequencing depths < 500. To determine the effect of tumor purities on TMB calculation, the distribution of TMB values with different

tumor purities was analyzed. As shown in Figures 2B and 2C, the TMB values were more stable with tumor purity estimated by HE staining between 0.2-0.8 or tumor purity estimated by NGS ≥ 0.4 .

High correlation of TMB calculated with the optimized SGZ algorithm and paired normal-tumor sequencing in East Asian populations

TMB calculation requires paired normal samples to remove germline mutations, which increases the cost of sequencing. In addition, in terms of clinical accessibility, paired normal samples are sometimes unavailable, which limits the clinical application of TMB. The SGZ algorithm for TMB calculation with tumor-only sequencing samples was designed based on European and American populations. For East Asian populations, the TMB calculated by the SGZ algorithm had a low correlation with the TMB calculated by the method of paired normal-tumor sequencing (Figure 3A). In order to calculate TMB from tumor-only sequencing samples in Asian populations, the SGZ algorithm was optimized. As shown in Figure 3B, we added the mutation filtering step, and constructed a background library with normal samples from East Asian patients to filter germline mutations, which ultimately reduced the false positives of TMB. To verify the accuracy of TMB calculation with an optimized algorithm, TMB calculated from the methods of the optimized algorithm and paired normal-tumor sequencing in sixty-two samples were compared. As shown in Figures 3C and 3D, their correlation coefficient is 0.95 and 82.7% of the mutations identified from the method of paired normal-tumor sequencing could be identified with the optimized SZG algorithm. These results demonstrate the accuracy of TMB calculation from the tumor-only sequencing with the optimized SGZ algorithm in East Asian populations.

Identification of the TMB cut-off for predicting immunotherapy efficacy in the training cohort

The cut-off of TMB for predicting the efficacy of immunotherapy in European and American populations is 10 mut/Mb [20]. However, since the TMB of East Asian populations is lower than that of European and American populations, a TMB cut-off of 10 mut/Mb may not be suitable for East Asian populations. To determine the cut-off of TMB for predicting immunotherapy efficacy in the East Asian population,

tumor samples from sixty-six lung patients treated with ICIs were retrospectively collected as a training cohort. TMB was calculated with the optimized SGZ algorithm. The performance of TMB for predicting patient durable clinical benefit was analyzed with a ROC curve and Log-rank test. As shown in Figure 4A, the optimal cut-off of TMB was 7.19 mut/Mb with AUC = 0.742, and validated with Log-rank analysis (Figure 4B). Furthermore, the response rate and survival period were higher in patients with TMB ≥ 7.19 mut/Mb than in those with TMB < 7.19 mut/Mb, and the TMB cut-off of 7.19 mut/Mb is better than the TMB cut-off of 10 mut/Mb in East Asian populations (Figure 4C-4F). To further investigate the role of TMB in predicting the efficacy of immunotherapy, the effects of TMB, medication type, tumor type and age group on patient survival were analyzed through multi-factor cox-regression. It was found that TMB was a favorable factor for patient survival, while other factors had no significant effect (Figure 4G).

Validation of TMB cut-off in the test cohort

To further validate the TMB cut-off of 7.19 mut/Mb, genomic and clinical data of sixty-nine East Asian NSCLC patients treated with ICIs were collected [30]. Consistent with the above results, the survival period of patients with TMB \geq 7.19 mut/Mb was longer than those with TMB $<$ 7.19 mut/Mb (Figure 5A). Furthermore, the predicting efficacy with the TMB cut-off of 7.19 mut/Mb is better than the TMB cut-off of 10 mut/Mb (Figure 5B).

Declarations

Conflict of Interest

Author Chao Pan, Hongzhen Tang, Dongfang Wu and Haitao Luo were employed by the company YuceBio Technology Co. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Conception and design: DQS and HTL. Acquisition of Data: DQS, MLX and CHP. Analysis and interpretation of data: MLX, CHP, HZT and DFW. Wrote the manuscript: MLX, CHP and HZT. Revised the manuscript: DQS and HTL. All authors have read and agreed to the published version of the manuscript.

Funding

Not applicable.

Acknowledgments

We would like to thank all the subjects and family members who gave their consent to present the data in this study, as well as the investigators and researchers at all relevant hospitals and study sites.

Ethics Statement

Written informed consent was obtained from each participant or the participant's family member at the time of sample collection, according to protocols approved by each hospital's ethics committee.

Data Availability Statement

All the analysis data can be found in supplementary table. The raw data generated in this study are available from the corresponding author on reasonable request.

References

1. Sacher AG, Gandhi L: **Biomarkers for the Clinical Use of PD-1/PD-L1 Inhibitors in Non-Small-Cell Lung Cancer: A Review.** *JAMA Oncol* 2016, **2**(9):1217-1222.

2. Perets R, Bar J, Rasco DW, Ahn MJ, Yoh K, Kim DW, Nagrial A, Satouchi M, Lee DH, Spigel DR *et al*: **Safety and efficacy of quavonlimab, a novel anti-CTLA-4 antibody (MK-1308), in combination with pembrolizumab in first-line advanced non-small-cell lung cancer.** *Ann Oncol* 2021, **32**(3):395-403.
3. Brahmer J, Reckamp KL, Baas P, Crino L, Eberhardt WE, Poddubskaya E, Antonia S, Pluzanski A, Vokes EE, Holgado E *et al*: **Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer.** *N Engl J Med* 2015, **373**(2):123-135.
4. Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, von Pawel J, Gadgeel SM, Hida T, Kowalski DM, Dols MC *et al*: **Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial.** *Lancet* 2017, **389**(10066):255-265.
5. !!! INVALID CITATION !!! .
6. Wang Y, Tong Z, Zhang W, Zhang W, Buzdin A, Mu X, Yan Q, Zhao X, Chang HH, Duhon M *et al*: **FDA-Approved and Emerging Next Generation Predictive Biomarkers for Immune Checkpoint Inhibitors in Cancer Patients.** *Front Oncol* 2021, **11**:683419.
7. Khoja L, Day D, Wei-Wu Chen T, Siu LL, Hansen AR: **Tumour- and class-specific patterns of immune-related adverse events of immune checkpoint inhibitors: a systematic review.** *Ann Oncol* 2017, **28**(10):2377-2385.
8. Weber JS, Kahler KC, Hauschild A: **Management of immune-related adverse events and kinetics of response with ipilimumab.** *J Clin Oncol* 2012, **30**(21):2691-2697.
9. Heinzerling L, Ott PA, Hodi FS, Husain AN, Tajmir-Riahi A, Tawbi H, Pauschinger M, Gajewski TF, Lipson EJ, Luke JJ: **Cardiotoxicity associated with CTLA4 and PD1 blocking immunotherapy.** *J Immunother Cancer* 2016, **4**:50.
10. Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, Powderly JD, Carvajal RD, Sosman JA, Atkins MB *et al*: **Safety, activity, and immune correlates of anti-PD-1 antibody in cancer.** *N Engl J Med* 2012, **366**(26):2443-2454.
11. Denis M, Duruisseaux M, Brevet M, Dumontet C: **How Can Immune Checkpoint Inhibitors Cause Hyperprogression in Solid Tumors?** *Front Immunol* 2020, **11**:492.
12. Mok TSK, Wu YL, Kudaba I, Kowalski DM, Cho BC, Turna HZ, Castro G, Jr., Srimuninnimit V, Laktionov KK, Bondarenko I *et al*: **Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial.** *Lancet* 2019, **393**(10183):1819-1830.
13. Gibney GT, Weiner LM, Atkins MB: **Predictive biomarkers for checkpoint inhibitor-based immunotherapy.** *Lancet Oncol* 2016, **17**(12):e542-e551.
14. Topalian SL, Taube JM, Anders RA, Pardoll DM: **Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy.** *Nat Rev Cancer* 2016, **16**(5):275-287.
15. Hirsch FR, McElhinny A, Stanforth D, Ranger-Moore J, Jansson M, Kulangara K, Richardson W, Towne P, Hanks D, Vennapusa B *et al*: **PD-L1 Immunohistochemistry Assays for Lung Cancer: Results from Phase 1 of the Blueprint PD-L1 IHC Assay Comparison Project.** *J Thorac Oncol* 2017, **12**(2):208-222.

16. Forde PM, Chaft JE, Smith KN, Anagnostou V, Cottrell TR, Hellmann MD, Zahurak M, Yang SC, Jones DR, Broderick S *et al*: **Neoadjuvant PD-1 Blockade in Resectable Lung Cancer.** *N Engl J Med* 2018, **378**(21):1976-1986.
17. Anagnostou V, Bruhm DC, Niknafs N, White JR, Shao XM, Sidhom JW, Stein J, Tsai HL, Wang H, Belcaid Z *et al*: **Integrative Tumor and Immune Cell Multi-omic Analyses Predict Response to Immune Checkpoint Blockade in Melanoma.** *Cell Rep Med* 2020, **1**(8):100139.
18. Marabelle A, Fakih M, Lopez J, Shah M, Shapira-Frommer R, Nakagawa K, Chung HC, Kindler HL, Lopez-Martin JA, Miller WH, Jr. *et al*: **Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study.** *Lancet Oncol* 2020, **21**(10):1353-1365.
19. Chen DS, Mellman I: **Oncology meets immunology: the cancer-immunity cycle.** *Immunity* 2013, **39**(1):1-10.
20. Marabelle A, Le DT, Ascierto PA, Di Giacomo AM, De Jesus-Acosta A, Delord JP, Geva R, Gottfried M, Penel N, Hansen AR *et al*: **Efficacy of Pembrolizumab in Patients With Noncolorectal High Microsatellite Instability/Mismatch Repair-Deficient Cancer: Results From the Phase II KEYNOTE-158 Study.** *J Clin Oncol* 2020, **38**(1):1-10.
21. Hsiehchen D, Espinoza M, Valero C, Ahn C, Morris LGT: **Impact of tumor mutational burden on checkpoint inhibitor drug eligibility and outcomes across racial groups.** *J Immunother Cancer* 2021, **9**(11).
22. Merino DM, McShane LM, Fabrizio D, Funari V, Chen SJ, White JR, Wenz P, Baden J, Barrett JC, Chaudhary R *et al*: **Establishing guidelines to harmonize tumor mutational burden (TMB): in silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of Cancer Research TMB Harmonization Project.** *J Immunother Cancer* 2020, **8**(1).
23. Oh S, Geistlinger L, Ramos M, Morgan M, Waldron L, Riestler M: **Reliable Analysis of Clinical Tumor-Only Whole-Exome Sequencing Data.** *JCO Clin Cancer Inform* 2020, **4**:321-335.
24. Sun JX, He Y, Sanford E, Montesion M, Frampton GM, Vignot S, Soria JC, Ross JS, Miller VA, Stephens PJ *et al*: **A computational approach to distinguish somatic vs. germline origin of genomic alterations from deep sequencing of cancer specimens without a matched normal.** *PLoS Comput Biol* 2018, **14**(2):e1005965.
25. Little P, Jo H, Hoyle A, Mazul A, Zhao X, Salazar AH, Farquhar D, Sheth S, Masood M, Hayward MC *et al*: **UNMASC: tumor-only variant calling with unmatched normal controls.** *NAR Cancer* 2021, **3**(4):zcab040.
26. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, Srinivasan P, Gao J, Chakravarty D, Devlin SM *et al*: **Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients.** *Nat Med* 2017, **23**(6):703-713.
27. Robinson DR, Wu YM, Lonigro RJ, Vats P, Cobain E, Everett J, Cao X, Rabban E, Kumar-Sinha C, Raymond V *et al*: **Integrative clinical genomics of metastatic cancer.** *Nature* 2017, **548**(7667):297-303.

28. Miao D, Margolis CA, Vokes NI, Liu D, Taylor-Weiner A, Wankowicz SM, Adeegbe D, Keliher D, Schilling B, Tracy A *et al*: **Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors.** *Nat Genet* 2018, **50**(9):1271-1281.
29. Consortium ITP-CAoWG: **Pan-cancer analysis of whole genomes.** *Nature* 2020, **578**(7793):82-93.
30. Fang W, Jin H, Zhou H, Hong S, Ma Y, Zhang Y, Su X, Chen L, Yang Y, Xu S *et al*: **Intratumoral heterogeneity as a predictive biomarker in anti-PD-(L)1 therapies for non-small cell lung cancer.** *Mol Cancer* 2021, **20**(1):37.
31. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z *et al*: **SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data.** *Gigascience* 2018, **7**(1):1-6.
32. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK: **VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing.** *Genome Res* 2012, **22**(3):568-576.
33. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM: **A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3.** *Fly (Austin)* 2012, **6**(2):80-92.
34. Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, Weigman VJ, Marynen P, Zetterberg A, Naume B *et al*: **Allele-specific copy number analysis of tumors.** *Proc Natl Acad Sci U S A* 2010, **107**(39):16910-16915.
35. Park S, Lee C, Ku BM, Kim M, Park WY, Kim NKD, Ahn MJ: **Paired analysis of tumor mutation burden calculated by targeted deep sequencing panel and whole exome sequencing in non-small cell lung cancer.** *BMB Rep* 2021, **54**(7):386-391.

Figures

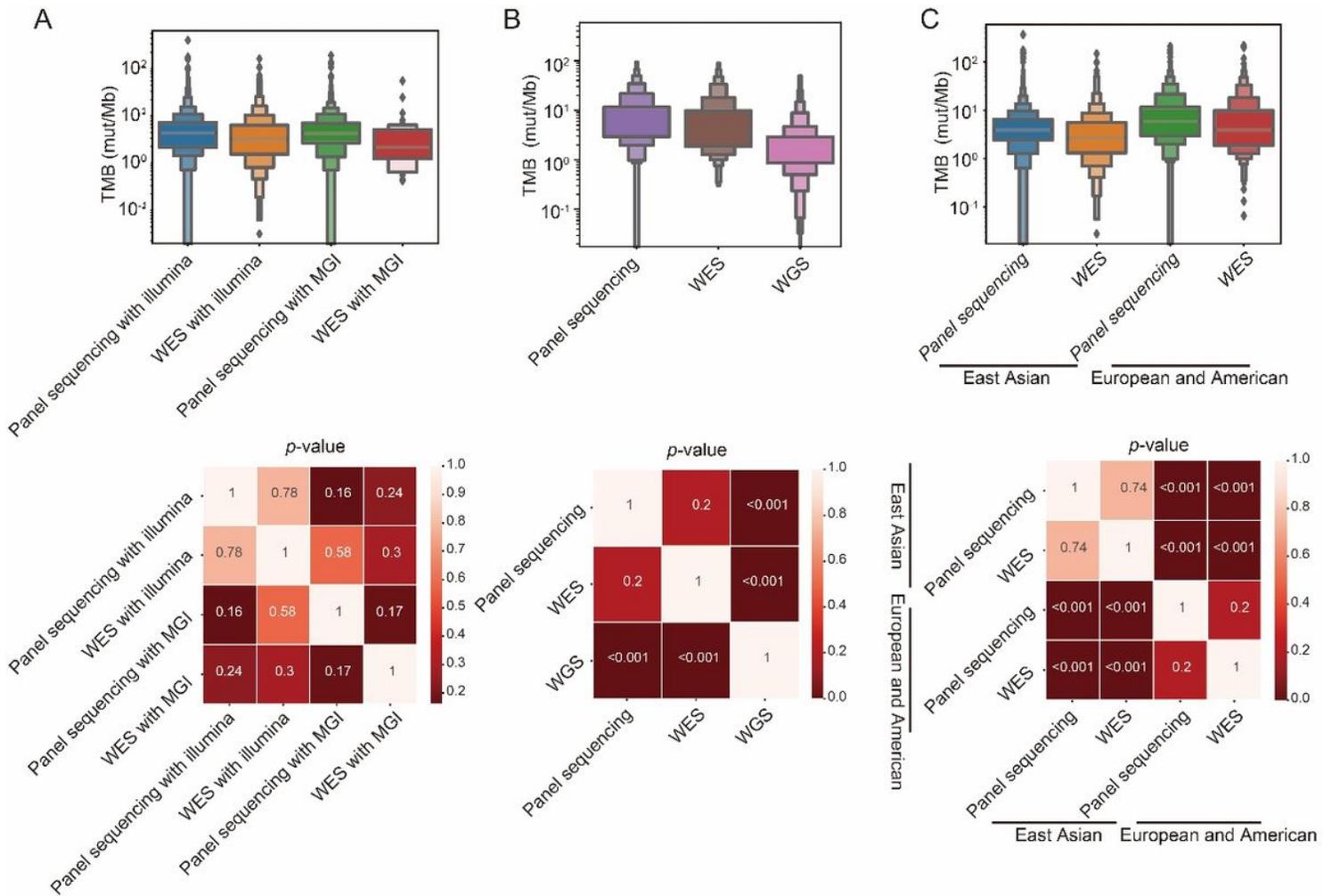
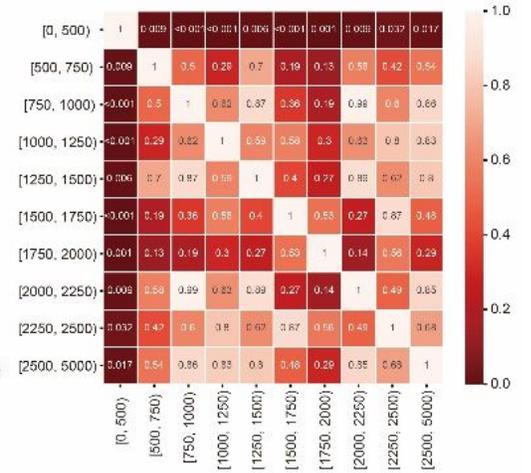
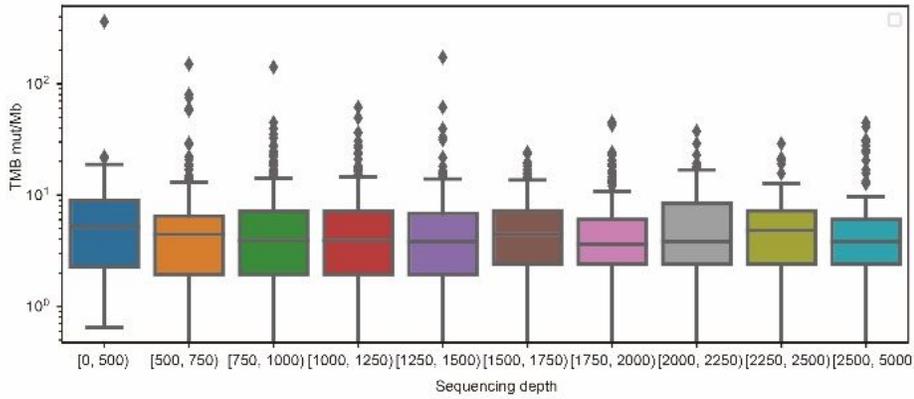


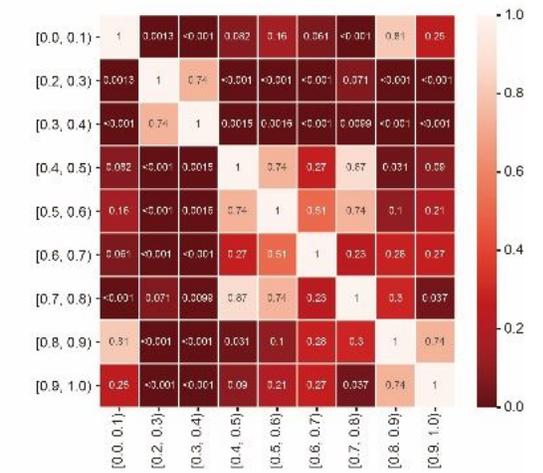
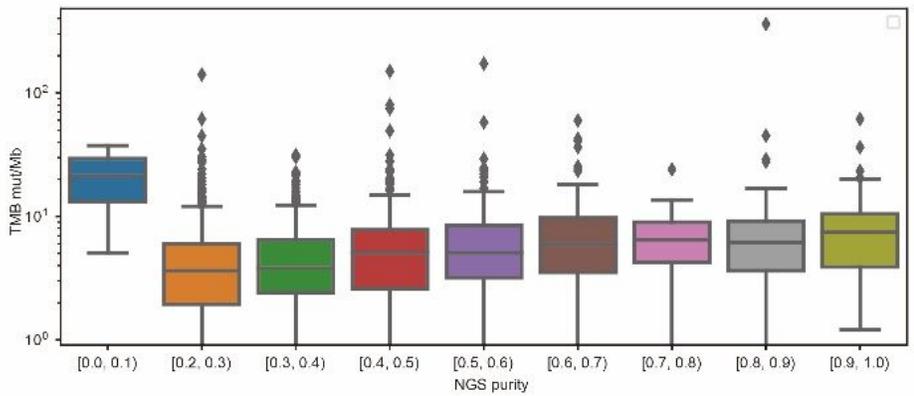
Figure 1

Effects of different sequencing platforms, sequencing methods and races on TMB. (A) Comparison of TMB between different sequencing platforms and methods in East Asian populations. (B) Comparison of TMB between different sequencing methods in European and American populations. (C) Comparison of TMB between different racial groups.

A



B



C

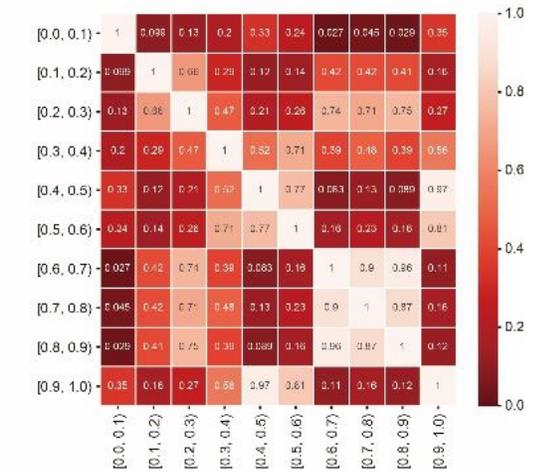
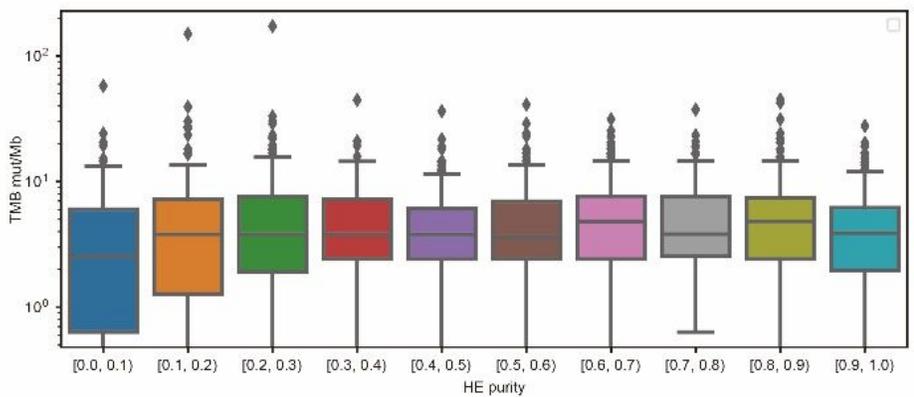


Figure 2

TMB calculation was affected by the sequencing depth and tumor purity. (A) Comparison of TMB between different sequencing depths. (B) Comparison of TMB between different purities estimated by HE staining. (C) Comparison of TMB between different purities estimated by NGS.

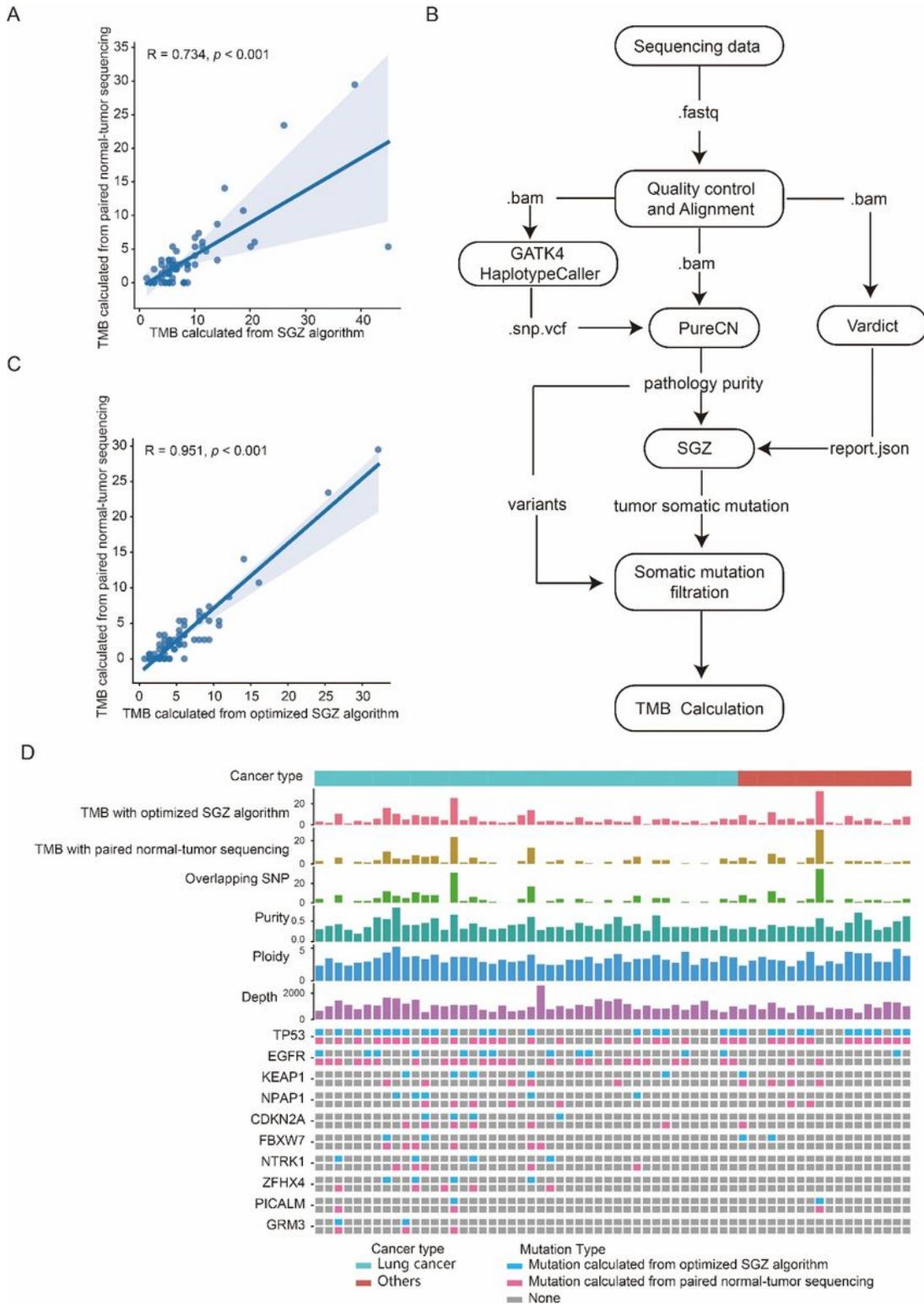
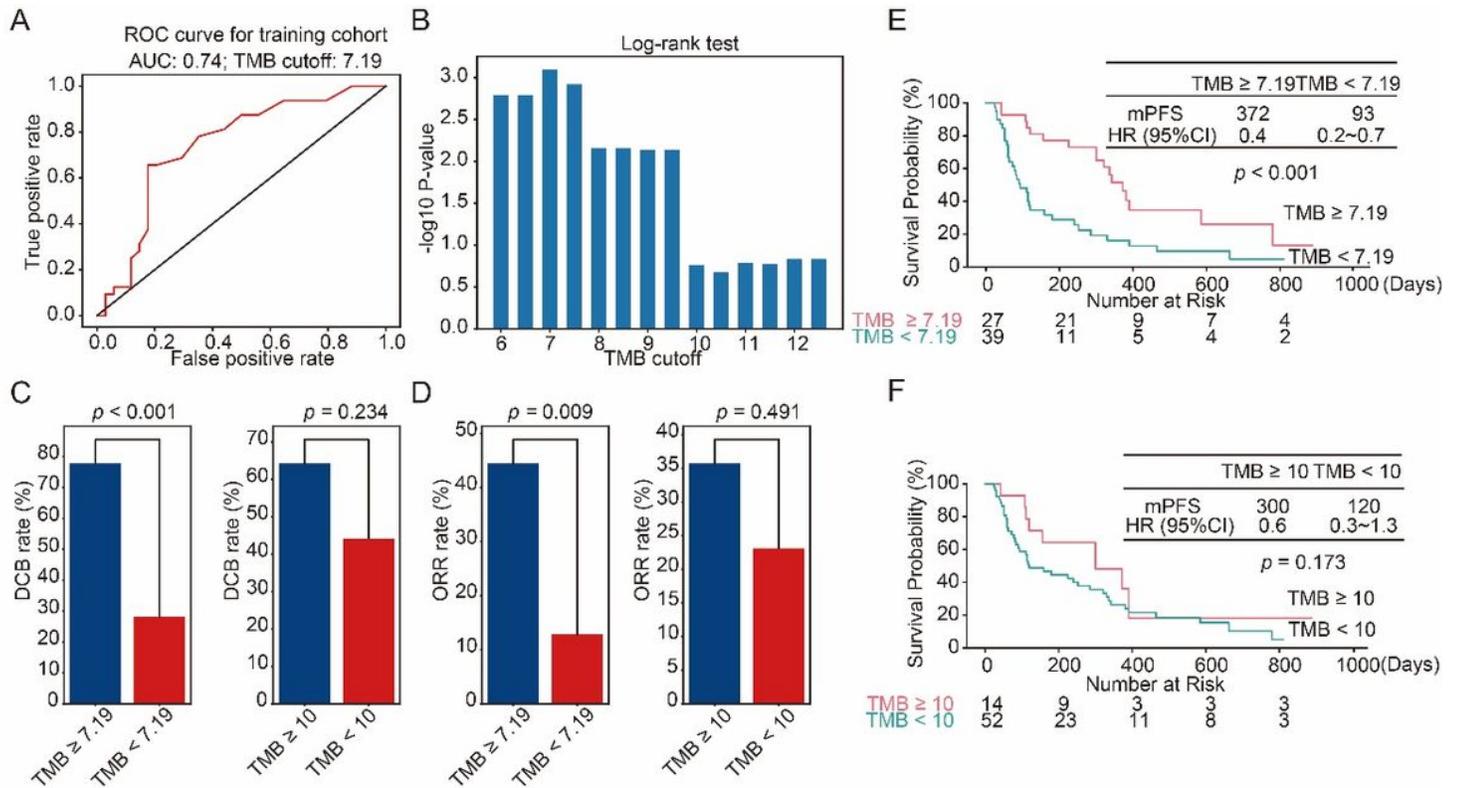


Figure 3

High correlation of TMB calculated with optimized SGZ algorithm and paired normal-tumor sequencing in East Asian populations. (A) Correlation of TMB calculated with the SGZ algorithm and the method of paired normal-tumor sequencing. (B) The process of TMB calculation from tumor-only sequencing samples with the optimized SGZ algorithm. (C) Correlation of TMB calculated with optimized SGZ algorithm and the method of paired normal-tumor sequencing. (D) The mutation landscape of sixty-two

patients calculated with methods of optimized SGZ algorithm and paired normal-tumor sequencing. The top three histograms are the values of TMB calculated with the optimized SGZ algorithm, TMB calculated with the method of paired normal-tumor sequencing and overlapping SNPs calculated by both methods. Center three histograms are purity, ploidy and sequencing depth of samples. The mutation spectrum of each patient is shown under the value of sequencing depth. The upper row is the mutation detected by the method of paired normal-tumor sequencing, and the lower row is the mutation detected by the method of the optimized SGZ algorithm.



G

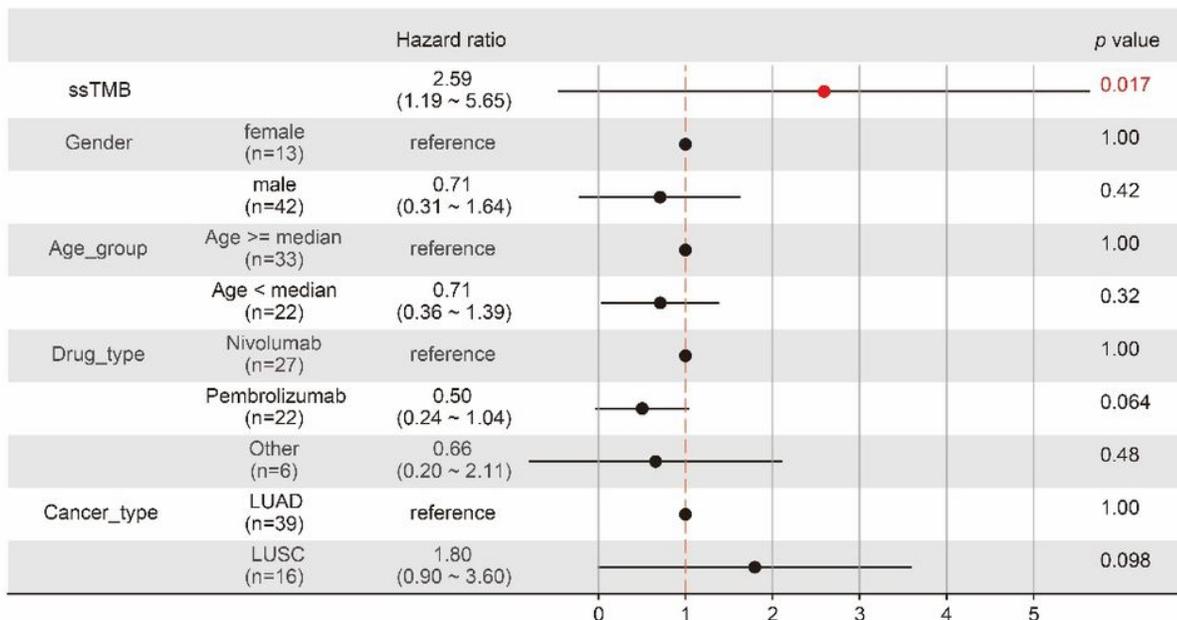


Figure 4

Identification of the cut-off of TMB for predicting immunotherapy efficacy in the training cohort. (A) and (B) ROC curves (A) and Log-rank test (B) for the identification of the TMB cut-off. (C) and (D) Barplots of DCB rate (C) and ORR rate (D) between different groups of TMB cut off 7.19 and 10. (E) Kaplan–Meier curves of OS comparing TMB ≥ 7.19 group and TMB < 7.19 group. (F) Kaplan–Meier curves of OS comparing TMB ≥ 10 group and TMB < 10 group. (G) The multivariate Cox regression analyses of the TMB, gender, age, drug, and cancer type.

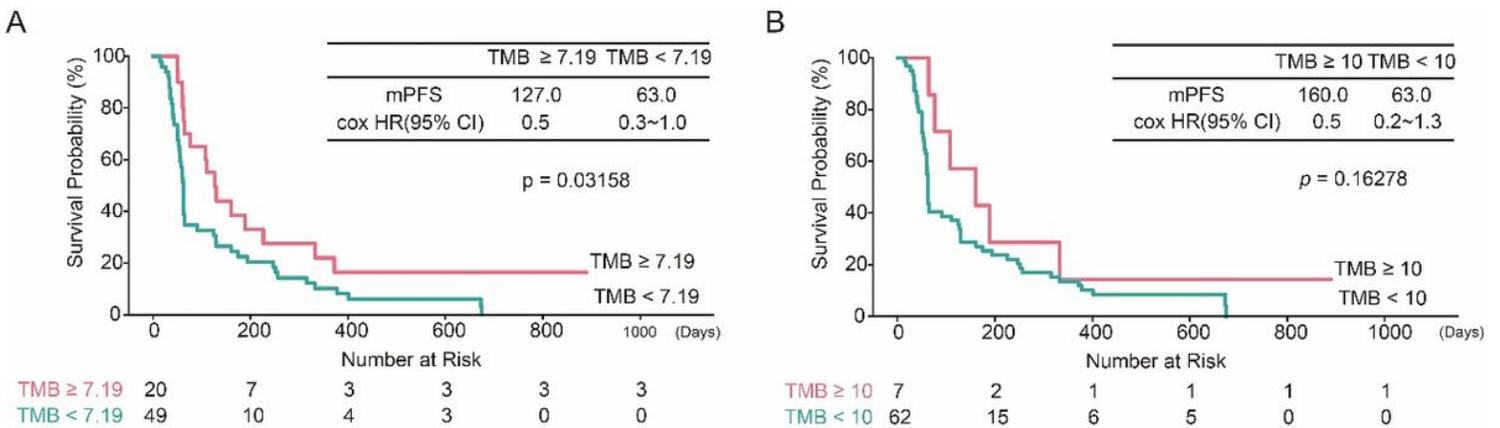


Figure 5

Validation of TMB cut-off of 7.19 mut/Mb in the test cohort. (A) Kaplan–Meier curves of OS comparing TMB ≥ 7.19 group and TMB < 7.19 group. (B) Kaplan–Meier curves of OS comparing TMB ≥ 10 group and TMB < 10 group.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarytable1.xlsx](#)
- [supplementarytable2.xlsx](#)
- [supplementarytable3.xlsx](#)
- [supplementarytable4.xlsx](#)
- [supplementarytable5.xlsx](#)