eLife

Computational and Systems Biology
Cancer Biology

Is tumor mutational burden predictive of response to immunotherapy?

Reviewed Preprint
v2 • October 14, 2024
Revised by authors

Reviewed Preprint
v1 • June 20, 2023

Carino Gurjao, Dina Tsukrov, Maxim Imakaev, Lovelace J Luquette, Leonid A Mirny

Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA • Broad Institute of MIT and Harvard, Cambridge, MA, USA • Institute for Medical Engineering and Science, and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA • Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

Abstract

Cancer immunotherapy by checkpoint blockade (ICB) is effective for various cancer types, yet its clinical use is encumbered by a high variability of patient response. Several studies have reported that the number of non-synonymous mutations (Tumor Mutational Burden, TMB), can predict patient response to ICB. This belief has become widespread and led to the FDA approval of immunotherapy patient prioritization based on TMB levels. The notion that TMB is predictive of response to immunotherapy is rooted in the neoantigen theory. It stipulates that cancer-specific mutations can form neoantigens recognized by the immune system; the more mutations a tumor has, the more likely the immune response is triggered. Here we revisit the data underlying the reported association of TMB with response, and the neoantigen theory. First we assembled the largest pan-cancer dataset of immunotherapy patients with sequencing and clinical data. Surprisingly, we find little evidence that TMB is predictive of response to ICB. We demonstrate that associations similar to the ones reported previously can be observed in shuffled data, suggesting that previous studies suffered from a lack of correction for multiple hypotheses testing and confounding disease subtypes.

Second, we develop a model that expands the neoantigen theory and can be consistent with both immunogenicity of neoantigens and the lack of association between TMB and response. Our analysis shows that the use of TMB in clinical practice is not supported by available data and can deprive patients of treatment to which they are likely to respond.

eLife Assessment

This **useful** study examines relationships between tumor mutational burden and the response to immunotherapy, using new data sets along with publicly available data sets. The authors conclude that tumor mutational burden cut-offs are unreliable proxies for predicting the response to therapy, underpinned by **solid** evidence, but with several caveats and assumptions that leave the central question subject to further inquiry. In summary, this is an interesting study that adds to a growing body of work investigating the particular conditions governing the effectiveness of immunotherapy.

https://doi.org/10.7554/eLife.87465.2.sa2

Introduction

Immune checkpoint blockade (ICB) treatments such as anti-CTLA-4 and anti-PD1, which target regulatory pathways in T-lymphocytes to enhance anti-tumour immune responses, have already proven to elicit durable clinical responses for some patients (1–4). However, the genetic determinants of response to immunotherapy have yet to be found. Several studies (5–9) suggested that Tumour Mutational Burden (TMB), computed as the total number of nonsynonymous somatic mutations, is correlated with response to immunotherapy in cancer. The underlying hypothesis posits that a fraction of nonsynonymous mutations become exposed as epitopes and constitute neoantigens, which can trigger an anticancer response by the immune system. The association between high mutational burden and response to immunotherapy, within and across cancer types (10–14), has been widely reported in the scientific literature and the media. As a result, TMB is currently discussed as the most clinically advanced biomarker of response to immune checkpoint blockade (15, 16), and the FDA approved the use of TMB to identify patients most likely to derive clinical benefit (17). These studies also triggered a search for inexpensive assays to evaluate TMB directly (18), as well as TMB-derived measures, such as neoantigens, neoepitopes, and mutation clonality (19), which are all currently under investigation to further stratify patients most likely to respond to immunotherapy. Our analysis focuses on TMB itself, as this is the most widely used and only FDA-approved measure.

Using a biomarker to stratify and prioritize patients for treatment runs a risk of depriving patients who have a chance to respond to a life-saving treatment. High variability of response makes relying on a predictor particularly risky. Hence, we revisit original data that were used to establish correlation between TMB and response. We tested TMB as a predictor of both binary responder/non-responder labels from original clinical studies, as well as continuous survival data. We also investigated whether a TMB threshold could distinguish patients with high and low survival after multiple hypothesis testing. We find that no TMB threshold performs better on the clinical data than on randomized ones. We further show that irrespective of the strategy to choose the threshold, even if we were to employ the optimal TMB cutoff, it would still lead to about 25% of responders falling below the treatment prioritization threshold. In addition, we re-examine the pan-cancer association of TMB with response rate to ICB.

Finally we revisit the neoantigen theory that was the rationale for using TMB as a predictor of response to immunotherapy. The theory stipulates that non-synonymous mutations can lead to the production of unique antigens (*neo*antigens) that are recognized by the immune system as foreign, triggering the immune response against cancer cells. The theory further assumes that the more mutations a cancer has, the more likely it triggers the immune system, and the more likely it will

benefit from immunotherapy. We develop a simple model that is based on the neoantigen theory and find that it has two regimes. In one regime, the probability of response increases gradually with TMB, as commonly believed. Yet in the other, the probability of response saturates after a few mutations, making a chance to respond independent of TMB. Our analysis of the clinical data is consistent with the latter regime. Thus our model shows that the neoantigen theory is fully consistent with the lack of association between TMB and response.

Results

Data aggregation

To evaluate the association of TMB with response to ICB across a broader range of cancer types, we aggregated and analysed data for 882 immunotherapy patients with publicly available pre-treatment whole-exome sequencing data (referred below as CPI800+, **Table S1** and **Material and Methods**). We included patient-level data from an aggregate of early seminal studies (20) as well as a clear cell renal cell cancer (21, 22), non-small cell lung cancer (9), bladder cancer (23) and melanoma (5, 24, 25) ICB-treated cohorts. For every dataset examined, we retrieved TMB levels and survival data (Progression-Free Survival (PFS) or Overall Survival (OS)) for each patient. The original studies provided response classification for most patients.

We also leveraged 1283 patients (termed CPI1000+) who underwent immunotherapy (26), have unified TMB (n=1083) and response definition, as well as survival measures for some patients (n=545 with OS data). Furthermore, we obtained gene panel data (MSK-IMPACT) for 1662 patients (**Table S1**) who underwent immunotherapy. To the best of our knowledge, together this dataset constitutes the largest pan-cancer aggregate of ICB-treated patients with sequencing and clinical data, which allow a robust unified statistical assessment of TMB as a predictor of ICB response

Is TMB associated with response after treatment?

First, we compared the TMB in patients that have been classified as responders and non-responders based on a number of clinical characteristics. All datasets show not only a considerable overlap in TMB between responders and non-responders, but the lack of considerably elevated TMB for responders and a large range of TMB values in each group.

Consistent with previous studies we find no difference in TMB between responders and nonresponders for major cancer types (clear cell Renal Cell Carcinoma, Head and Neck Squamous Cell Carcinoma and Breast Cancer), with only melanoma (mel1 and mel2, p=0.026 p=0.041) and non-small cell lung cancer datasets (lung1 and lung2, $p=8.3\times10^{-6}$ and $p=7.7\times10^{-3}$) yielding significant differences as reported earlier (9, 20, 21, 24, 27) (**Figure 1 and Figure S1A**). Analysis in CPI1000+ revealed similar results, yet colorectal cancer, and bladder cancers showed significance of elevated TMB for responders (p=0.044 and $p=5.4\times10^{-7}$, respectively) (**Figure S1B**). We also found, consistent with previous studies, that this small elevation of TMB for responders is due to confounding effects of cancer subtypes, i.e. due to the different response rates of cancer subtypes with different TMB ranges (see **Supplemental Text)**. When we revisit a study that reported pan-cancer correlation between response rate and TMB (12, 14), we found that association was driven largely by overall high response in melanoma and subtypes of colorectal cancer (MSI+) with extreme differential response (see **Supplemental Text**). No association between TMB and response rate across all other cancer types is present in available data.

We also examined potential correlations between TMB and survival, rather than using a binary response variable (**Figure 2**). Strikingly, plots of survival versus TMB do not show a visible correlation, trend or TMB cutoff that could differentiate longer and shorter surviving patients. There is a considerable and non-diminishing fraction of patients with long survival, even for lower

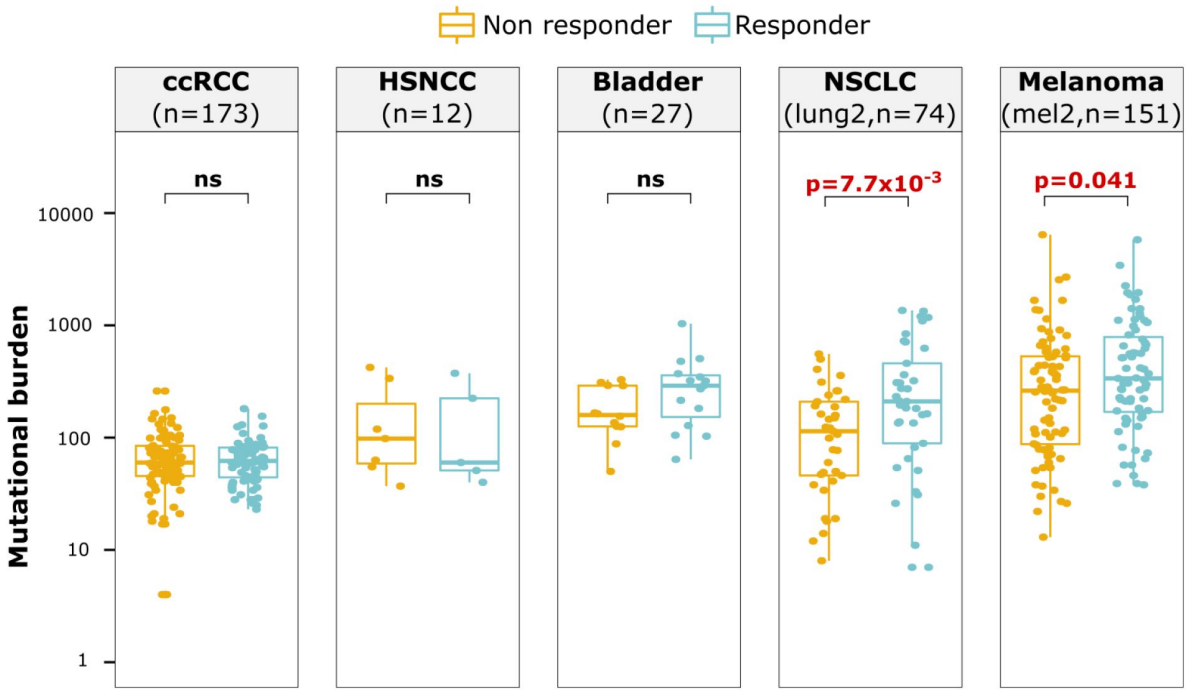

Figure 1

TMB and response to ICB

Association of TMB with response to ICB across five cancer types from CPI800+ (the largest cohorts of each cancer type are plotted here, the others are shown in **Figure S1A**). Only melanoma and non-small cell lung cancer have a significantly different TMB between responders and non-responders. ccRCC: clear cell Renal Cell Carcinoma, HNSCC: Head and Neck Squamous Cell Carcinoma; NSCLC: Non Small Cell Lung Carcinoma

ranges of TMB values. As we demonstrate below, attempts to find a TMB cutoff value to differentiate long- and short-survivals on such data can be prone to misinterpretation and require careful correction for multiple hypothesis testing.

Next we investigated whether (i) small differences in TMB between responders and non-responders for some cancer types can be of clinical use, even when p-value>0.05; (ii) there is a potential TMB cutoff that can predict groups of patients with a significant difference in survival; (iii) these clinical data can be reconciled with the neoantigen theory.

TMB is a poor predictor of response

The key component for validating a biomarker is acceptable classification accuracy, i.e. the biomarker's capacity to correctly classify a patient's response (28). ROC curves analysis (**Figure 3A** and **3B**) is a standard tool used across disciplines for measuring the quality of a predictor; it provides a comprehensive quantification of specificity and sensitivity over all possible cutoffs, with the Area Under the ROC Curve (AUC) being an aggregate measure of predictor performance (AUC=0.5 for a predictor performing as well as random). Our ROC-curve analysis shows the (i) lack of a clear TMB cutoff that could be used in the clinic; (ii) poor performance of the TMB-based predictor of response to ICB, as evident from the low AUC in most datasets: mel1 and mel2 yielding of 0.62 and 0.59, and lung2 has an AUC of 0.68. Lung1, however, has the highest AUC of 0.85, which, as we show below, is still insufficient to select patients for ICB. ROC curve analysis on CPI1000+ cohort, with unified TMB, also shows a similarly poor AUC of 0.6 (**Figure S2**).

Using a poor predictor for treatment decisions can lead to patient misclassification, i.e. patients who could benefit from the therapy would be deprived of it (responders below the TMB threshold), and patients who get the treatment but don't benefit from it (non-responders above the threshold). To quantify the shortcomings of TMB-based selection of patients for treatment we computed the proportion of misclassified patients based on the FDA approval of 10 mutations/Mb threshold to select patients for ICB (**Figure 3C** and **Figure 3D**). We find that, on average across the non-small cell lung cancer and melanoma datasets, 62% of responders were below the treatment prioritization threshold and 19% of non-responders were above. While these misclassification rates vary across datasets, fractions of potential responders under the TMB threshold remain high. Moreover, the poor predictive power of TMB indicates that current efforts of choosing a single TMB measure for all cancer types (so call "harmonizing" TMB) would not address fundamental limitations of TMB as the biomarker of response. Indeed, our ROC analysis shows that even the optimal cutoff (Youden index associated cutoff) for each dataset would result in approximately 25% of responders ending up below the treatment prioritization threshold and thus discouraged from receiving a potentially efficacious and life-extending treatment (**Figure 3D**). As such, the main challenge in using TMB in the clinic is the inherently poor association between TMB and response to treatment.

TMB cannot detect groups of patients with different survival post-immunotherapy

To evaluate the use of TMB for prioritizing patients, and to go beyond the binary response classification, we examined an association between TMB and survival time (Overall Survival, OS and Progression-free Survival PFS). Since survival data is "censored" i.e. only lower bound on survival is known for some patients that didn't show progression or dropped out of the study, standard correlation-based methods cannot be used to evaluate such association. Nevertheless, groups of patients can be compared using survival analysis methods. Hence we tested whether it is possible to find a TMB threshold that can separate patients into groups with significantly distinct survival.

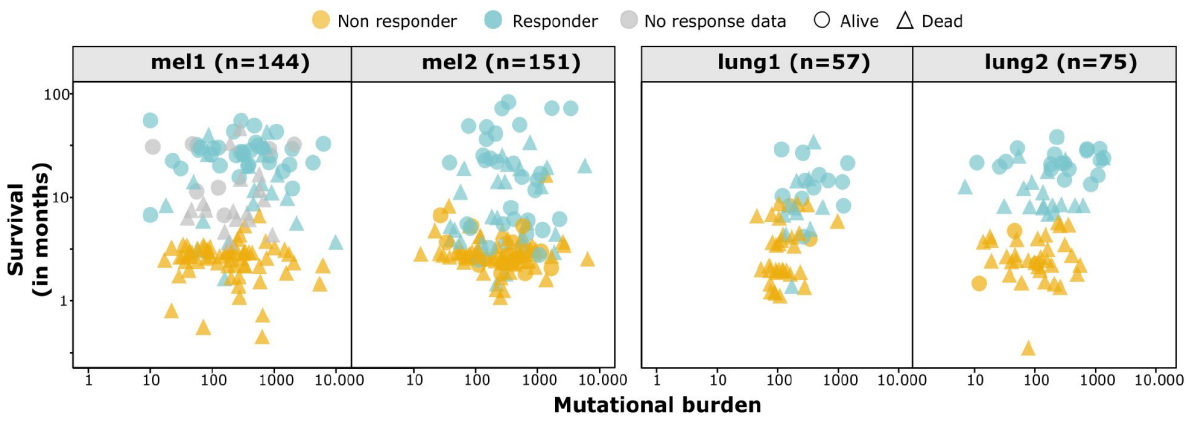

Figure 2

TMB association with progression-free survival post-immunotherapy

Plots of progression-free survival and TMB for melanoma and lung cancer ICB cohorts show the lack of correlation or of an obvious TMB cutoff. Computing a simple correlation for survival and censored data cannot correctly represent the dependence since patients who are alive live longer than the reported survival, and limiting correlation to patients who are dead would bias the analysis. Thus other survival statistics are used through the paper.

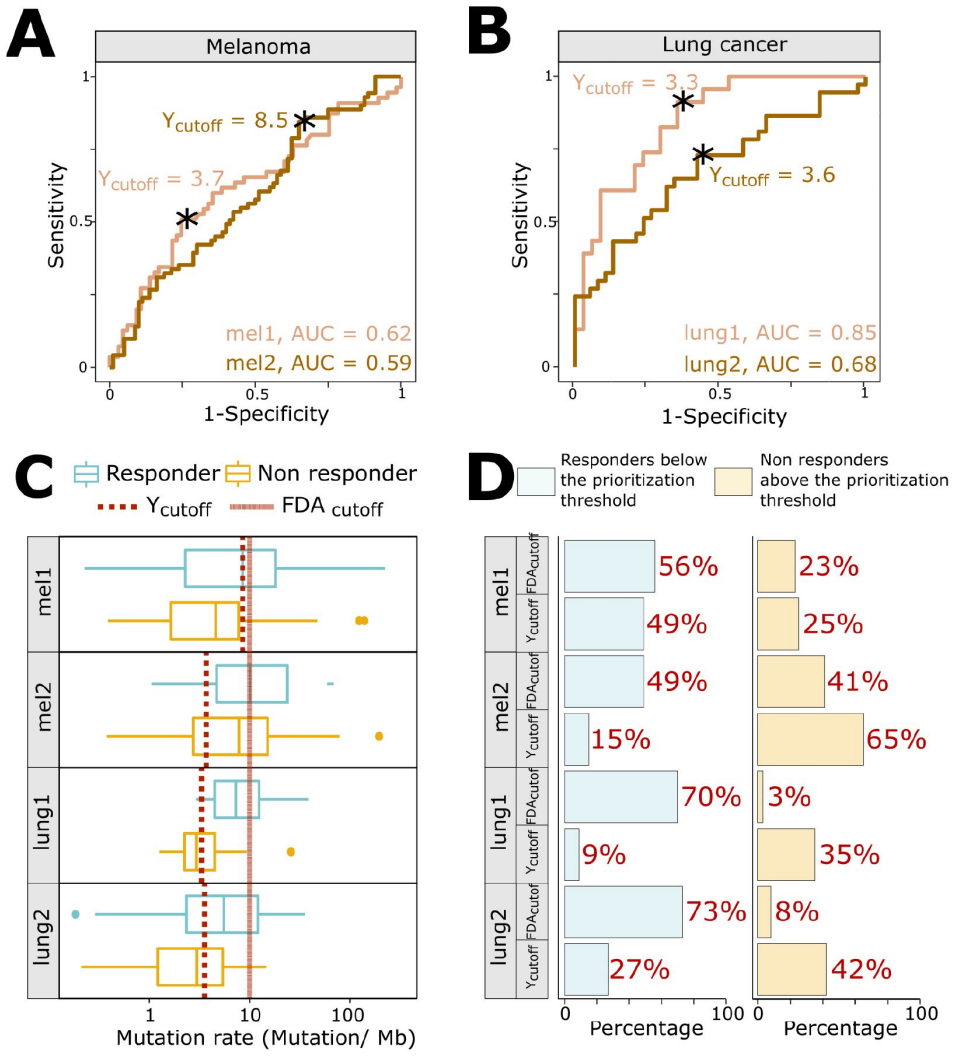

Figure 3

TMB as a biomarker of response to immunotherapy

(A)(B) ROC curves for the melanoma and lung cancer cohorts. The asterisk represents the Youden index cutoff. **(C)** Boxplots of nonsynonymous mutation rates across responders and non responders in the melanoma and lung cancer cohorts. The FDA-approved cutoff (10 mutations/Mb) and the best cutoff (Youden index associated cutoff) are shown by vertical lines. **(D)** Proportion of misclassified patients based on the FDA-approved cutoff, as well as the Youden index cutoff for each dataset. The use of either cutoff leads to substantial fraction of misclassified patients (potential responders below the treatment cutoff, or non-responders above the cutoff).

Scatterplots of survival versus TMB (**Figure 2**, **Figure S3A** and **S3B**) show no evidence of such TMB threshold. Nevertheless, several studies reported (6, 8) a seemingly statistically significant difference in survival between patients below and above a particular threshold. One caveat of this approach is that choosing the threshold value suffers from inherent multiple hypothesis testing, i.e. when the TMB thresholds have been selected among numerous possible alternatives multiple hypotheses are being tested. This inherent multiple hypothesis testing would require further correction of the p-values; a step that is missing in other studies. However, standard approaches (e.g., Bonferroni correction, FDR correction) for multiple hypotheses testing could be too stringent because the hypotheses generated by comparing survival in two groups at multiple TMB thresholds are not independent.

To address this challenge we used a randomization approach (29, 30), similar to earlier studies that examined cutoff selection for dose-response in epidemiological studies (31). We generate 1000 randomized datasets by shuffling TMB among patients, while keeping survival and censored labels unchanged for each patient. First, for each dataset (real or randomized) we determined the optimal threshold that maximizes the difference in survival for groups of patients above and below the threshold. This was done by trying all possible threshold values, computing the difference in survival by logrank test for the groups above and below the threshold; and selecting the optimal threshold that maximizes the difference in survival (i.e. minimize the logrank p-value). Second, we compared the optimal p-value for the real data, p_{real}, with the distribution of those for shuffled datasets $f(p_{shuf})$ and computed the *corrected p-value* as the fraction of shuffled datasets below the real ($p_{shuf} < p_{real}$) (**Figure 4A**).

Applied to the melanoma and lung cancer data, this approach (**Figure 4B** and **Figure 4C**) shows that the majority (~60-70%) of randomly shuffled datasets produced p_{shuf} below the standard 0.05 threshold, creating a seemingly significant TMB-survival association and emphasizing the need for multiple hypothesis correction. Applied to lung, melanoma and across CPI1000+ dataset (**Figure S4**), our correction for multiple hypothesis testing reveals the lack of a significant TMB threshold that can classify patients into groups with different survival. In particular for lung cancer, for which we previously observed a significant association between TMB and response (**Figure 1A**), we obtained no significant threshold for TMB. We also ran our analysis using OS (for datasets where both OS and PFS are available: mel1, mel2 and lung1) instead of PFS as an endpoint and showed similar results, suggesting that survival definitions do not drive the results of our analysis (**Figure S5A** and **S5B**).

We further obtained consistent results for 1662 patients of MSK-IMPACT cohort treated with ICB but genotyped with gene panels rather than whole-exome sequencing (**Figure S6**). Most of the 10 cancer types tested had a non-significant p-value including colorectal cancer (p=0.088) and melanoma (p=0.093) which have marginally significant p-values, except for non-small cell lung cancer (p=0.034). This study did not provide additional information such as tumor location for melanoma, Microsatellite Instability (MSI) status for colorectal cancer, or COPD for non-small cell lung tumors, which, as we showed above, can confound the association of TMB with response (24, 32). Taken together, our analysis shows the lack of TMB thresholds that can establish a high-TMB group with a significantly longer survival.

Model reconciles neoantigen theory and data

Neoantigen theory is widely used to argue that cancers with high TMB are more likely to elicit an immune response upon ICB. Although our results show the lack of such dependence, we demonstrate that the effect we observe can nevertheless be explained by a simple mathematical model of neoantigens and immunogenicity (**Figure 5A**).

Response to ICB treatment requires that cancers elicits an immune response, hence the probability of clinical response can be written as $P\{clinical\ response\} = P\{clinical\ response | clinical\ response\} \cdot P\{immune\ response\}$, where $P\{clinical\ response | clinical\ response\}$ is the probability of clinical

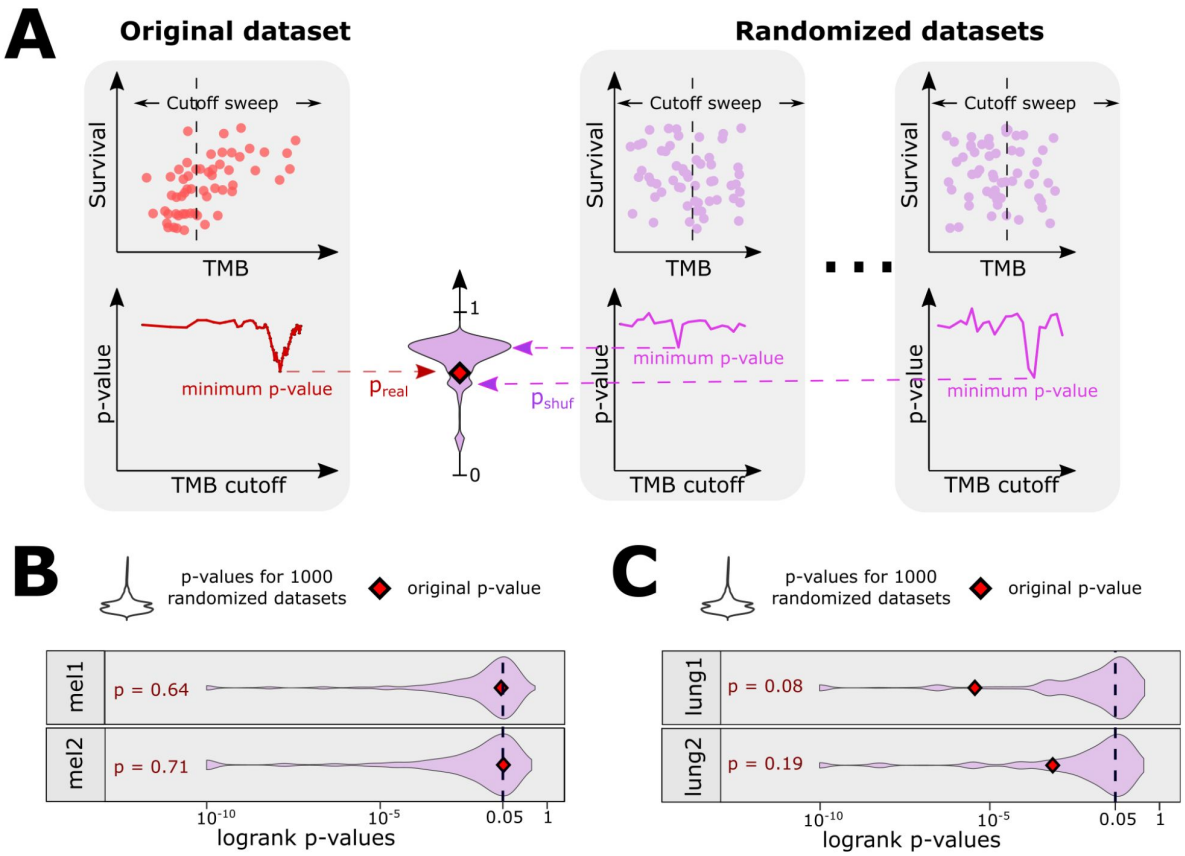

Figure 4

TMB association with progression-free survival post-immunotherapy

(A) Overview of the randomization analysis. Left: the optimal cutoff is found to maximize the difference between survival between groups above and below the cutoff (i.e to minimize the logrank p-value, yielding p_{real}). Right: the same procedure for shuffled data yields p_{shuf}. The fraction of $p_{shuf} < p_{real}$ produces a p-value corrected for multiple hypothesis testing for non-independent tests. **(B)** Results of the randomization analysis in the melanoma cohorts and stratification by subtypes (p-values < 10^{-10} not shown) **(C)** Randomization analysis results in the lung cancer cohorts and stratification by subtypes (p-values < 10^{-10} not shown).

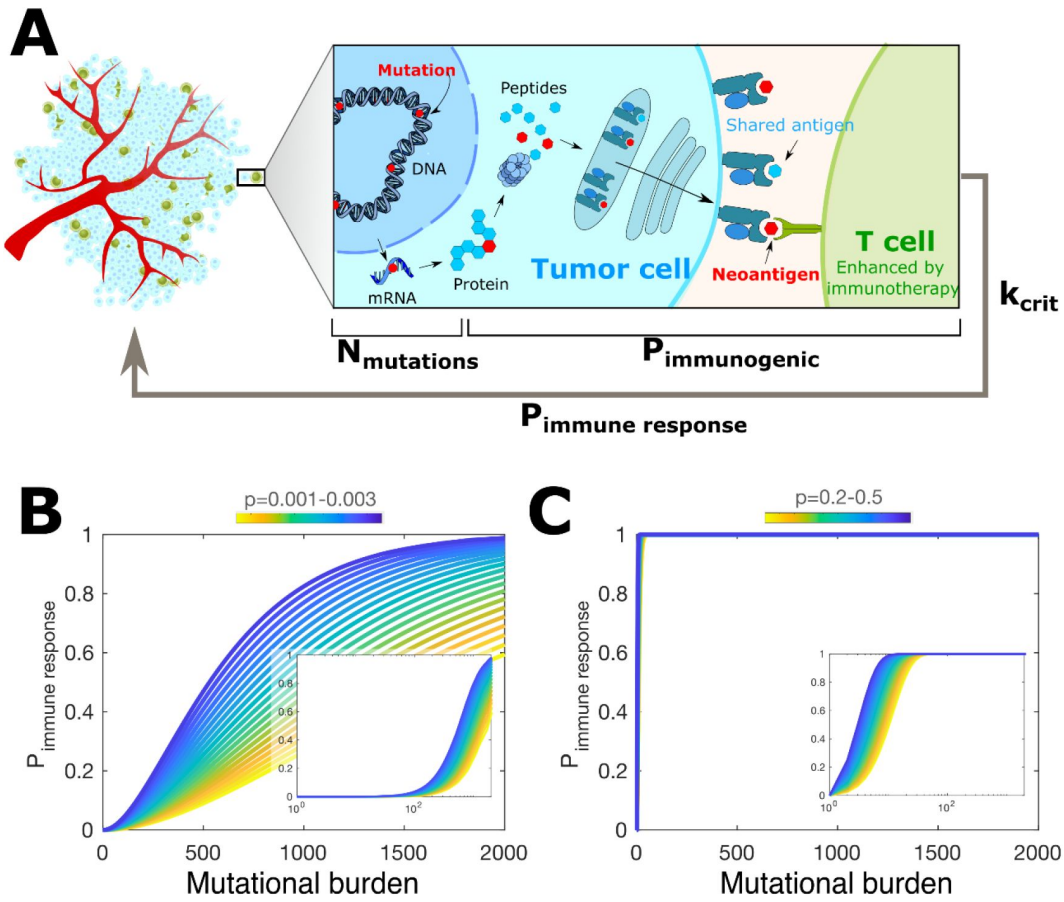

Figure 5

TMB and cancer immunogenicity

(A) Our model of cancer immunogenicity coarse-grains several cellular processes into the probability that a mutation becomes immunogenic ($P_{immunogenic}$). If the number of immunogenic mutations reaches k_{crit}, the cancer triggers an immune response. The probability of immune response $P_{immune\ responce}$ as a function of TMB for k_{crit} =2 for low and **(B)** high **(C)** ranges of p, the probability of a mutation to be immunogenic.

response, given that cancer elicits an immune response which is complex and depends on many factors including tumor immune microenvironment. Yet the prerequisite for the clinical response is the immune response $P\{immune\ response\}$ that we focus on. For simplicity assume that every non-synonymous mutation has the same probability p to be immunogenic (i.e. to be expressed, presented, interact with MHC, and trigger an immune response; we generalize to different values of p below). Due to immunodominance, only k_{crit} immunogenic mutations are sufficient to elicit a full immune response. Hence, the probability for a cancer with N mutations (=TMB) to elicit an immune response is then the probability of having k_{crit} or more immunogenic mutations among N:

$$P\{immune\ response\} = \sum_{k=k_{crit}}^{N} \binom{N}{k} p^{k} (1-p)^{N-k} = 1 - \sum_{k=0}^{k_{crit}-1} \binom{N}{k} p^{k} (1-p)^{N-k}$$

which is the CDF of a binomial distribution. In the case of $k_{crit} = 1$, when a single immunogenic mutation can trigger the response the expression simplifies

$$P\{immune\ response\} = 1 - (1-p)^{N} \approx 1 - exp(-pN).$$

Figure 5B, **Figure 5C** and **Figure S7** show $P\{immune\ response\}$ as a function of N (TMB) for a range of p and k_{crit} values. The model shows two different behaviors. If individual mutations are unlikely to be immunogenic ($p \ll 1$), e.g. due to a low probability of being presented, the probability of response increases gradually with TMB (**Figure 5B**). The neoantigen theory generally expects such gradual increase in immunogenicity of cancer with TMB. Yet, available data (**Figure 2**) don't show such a trend.

On the contrary, if mutations are more likely to be immunogenic p~0. 1, the probability of response quickly saturates (**Figure 5C**), making such tumors respond to ICB irrespective of TMB, as we observed in clinical data. In the case of $k_{crit} = 1$, $P\{immune\ response\}$ saturates for N~1 and p ~ 10. In the data we observe about the same probability of response to ICB for cancers with more than ~10-20 mutations (**Figure 1**). Our model shows that achieving about constant $P\{immune\ response\}$ for $N > 10 - 20$ mutations, requires $p \sim 0.1$ for $k_{crit} = 1$, and $p \sim 0.2$ for $k_{crit} > 1$. The same argument holds when each mutation has its own probability to be immunogenic $p(i)$, then $P\{immune\ response\} = 1 - \Pi_{i=1}^{N}(1 - p(i)) = 1 - exp(-\bar{p}N)$, where $\bar{p}$ is the mean probability of a mutation to be immunogenic. Thus only the average probability of a mutation to be immunogenic matters. In summary, we find that the model agrees with clinical data if individual non-synonymous mutations have, on average, p ~10 – 20% chance for triggering an immune response.

These estimates are consistent with *in silico* neoantigen predictions which showed (33) that 64% of mutations are strong, and 22% weak binders. For these values the model shows that the probability of eliciting a response quickly approaches 1 for TMB ≥ 10 and stays constant and independent of TMB. **(Materials and Methods).**

The model further suggests that for the regime consistent with the data ($P_{immunogenic}$=0.2-0.6; k_{crit} = *1-2*) (i) >90% of tumours with as little as 10 non-synonymous mutations are immunogenic; (ii) when 90% of tumours are immunogenic they have on average as few as 2 immunogenic mutation, the vast majority of mutations present in cancers are non-immunogenic. Such saturation of immunogenicity with low TMB in our model suggests that further immunogenic mutations experience weak negative selection (i.e. threshold epistasis), i.e. weak -if any- immunoediting (34). These results are also consistent with observed immunodominance hierarchies of the T cell responses (35): low TMB tumours can mount responses as robust as high TMB tumours since only a small subset of neoantigens are targeted by T cells.

Furthermore, our model also explains a puzzling observation that immunoediting, i.e. negative selection against immunogenic mutations, is inefficient, allowing tumors to accumulate a high TMB (36, 37). Indeed, once a cancer accumulates mutations making it immunogenic, additional mutations incur at no additional selective disadvantage i.e. show "the epistasis of diminishing return", and hence accumulate as neutral or weakly damaging passenger mutations (34, 38, 39). Moreover, according to this argument, a cancer should develop means to suppress the immune response early in its development, a prediction that can be tested in future studies of cancer clonal evolution. Taken together, our model and analysis of the available data indicate that cancer with even very few mutations can be immunogenic, suggesting that patients with low TMB might also mount robust immune responses, as has been shown for paediatric patients with acute lymphoblastic leukaemia (35).

Discussion

Tumor Mutational Burden, a measure of the total somatic nonsynonymous mutations in a tumor, became a popular biomarker of response to ICB, notably because of its relative simplicity to assess.

However, this paradigm is largely based on a series of early papers that examined response in melanoma and lung cancer that we show here to be potentially problematic statistically and further confounded by tumor subtype. Several studies have also reported poor association of TMB with response for specific cancer types, and highlighted TMB and its expression/presentation-based derivatives as problematic for clinical cohort classification (27). In particular for melanoma, published analyses (24) and our results indicate that the site location can explain the observed association between TMB and response to ICB. For lung cancer, our analysis points to the possibility that co-occurrence with COPD may explain the association between TMB and response to ICB among smokers. Overall we demonstrate that while most cohorts and cancer types show the lack of association of TMB and response or survival, the remaining statistical signal in some cohorts can arise due to confounders such as clinical subtypes. Future studies can examine the underlying biology of TMB and neoantigens, aiming to explain the better responses to ICB in certain clinical and cancer subtypes.

Critically, even if responders show significantly but slightly elevated TMB, such associations do not imply the suitability of TMB as a biomarker of response. In particular, we show that no TMB cutoff can distinguish groups of patients with significantly different survival rates. Besides, we show that TMB has poor accuracy as a classifier of response, even in the best-case scenario (Youden optimal cutpoint). This result challenges the FDA approval of TMB for prioritizing patients for ICB. If implemented, such TMB-based clinical decision making would deprive many patients who can benefit from ICB from receiving a life-extending treatment.

An ICB clinical trial that used FDA-approved TMB threshold (KEYNOTE-158) (40) has focused on rare cancers, excluding melanoma and lung cancer. While claiming a higher response rate among high-TMB patients, the trial observed little, if any, difference in overall survival of high-TMB and other patients, putting in question the clinical use of TMB-based prioritization.

We also put forward a simple model that reconciles our findings with the neoantigen theory. Our model shows that if each mutation has a high chance of triggering an immune response, then only a few new mutations make a cancer immunogenic, consistent with the observed immunodominance i.e. the immune response is mounted against only a few of the neoantigens. This result is also consistent with the observed lack of association between antigen density and T-cell presence previously reported (41). Moreover, our model suggests that most cancers are immunogenic, arguing that failures of ICB likely arise due to factors independent of cancer immunogenicity. Quantitative measurements (42) and modelling of neo-antigenic effects can deepen our understanding of cancer development and response to immunotherapy.

Although attractive and scalable, TMB does not consider the effect of specific mutations (missense, frameshift etc), their presentation and clonality (19), nor the state of the tumour, its microenvironment, and interactions with the immune system that can be integrated into potentially better predictors of response to ICB (43, 44). In addition, another major limitation of TMB is the lack of standardized measures. This includes the lack of standard sequencing methods to assess TMB: TMB can be measured from Whole-Exome sequencing, Whole-Genome sequencing, targeted panel and even RNA sequencing(45). This also includes biases introduced by using different mutation calling pipelines resulting in different TMB, sequencing depth and different characteristics of the samples (e.g. low purity samples typically yield lower TMB). For the biology of oncoimmunity, our analysis suggests that, contrary to the neoantigen theory, cancer immunogenicity does not increase with the growing load of neoantigens, and that clinical subtypes can underlie better response to ICB.

Altogether, our analysis indicates that low TMB should not be used to deprive otherwise eligible patients for immunotherapy treatment, and stimulates further research into other determinants of response to immunotherapy.

Material and Methods

Immunotherapy study population

CPI800+ was formed of eight independent WES cohorts (n=882, detailed in **Table S1**). The TMB and clinical annotations were not modified from the original studies. Post ICB sequenced samples were excluded from our analysis. In addition, gene panel datasets (n=1662, detailed in **Table S1**) were identified from cbioportal (46).

TCGA data

Lung cancer TCGA data were also retrieved from cbioportal (46), and additional clinical annotations were downloaded from The Cancer 3' UTR Atlas (47). COPD status was assessed based on the standard spirometric classification, i.e. post-bronchodilator ratio of forced expiratory volume in one second (FEV1) and forced vital capacity (FVC) below 70%.

Statistical analysis

We used R version 3.6.2 to perform statistical analyses. Two-group comparisons were evaluated by a two-sided Mann–Whitney U test unless otherwise indicated. $P < 0.05$ was considered statistically significant.

Code availability

The R code and data used to reproduce the analysis and figures from the paper are available on GitHub *https://github.com/mirnylab/TMB_analysis*

Acknowledgements

We are grateful to Krupa Thakkar, Kevin Litchfield and Charles Swanton for running our analysis on their “CPI1000+” meta-aggregate of immunotherapy patients (26).

We are thankful for many productive and exciting discussions of this work to Christopher McFarland, Johannes Berg, Donate Weghorn, Eli van Allen, Kenneth Kehl, Shamil Sunyaev, Martha Luksza, Michael, Lassig, Boris Reizis, Virginia Savova, Gregory Kryukov, Sebastian Amigorena, Sean McGrath, Baptiste Boisson, Toni Choueiri, Paul B Robbins and all members of the Mirny lab. We are grateful to the organizers and participants of "Physicists working on Cancer" workshop at the Weizmann Institute of Science, Schwartz/Reisman Institute for Theoretical Physics, particularly, Eytan Domany, Herbert Levine, Caterina La Porta and Stefano Zapperi. We are also grateful to the organizers and participants of the workshop "Tumors and Immune Systems: From Theory to Therapy" seminar at Institut d'Etudes Scientifique de Cargèse, particularly, Alexandra Walczak, Thierry Mora, Vassili Soumelis, Paul Thomas and Jason George.

LJL was supported by the Training Program in Bioinformatics and Integrative Genomics (NIH T32HG002295, PI: P.Park). This project grew from the qualifying examination problem in Bioinformatics and Integrative Genomics given by LAM to LJL. We acknowledge support of the MIT-France Seed Fund, and The Chicago Region Physical Science Oncology Center (PS-OC, National Cancer Institute U54CA193419).

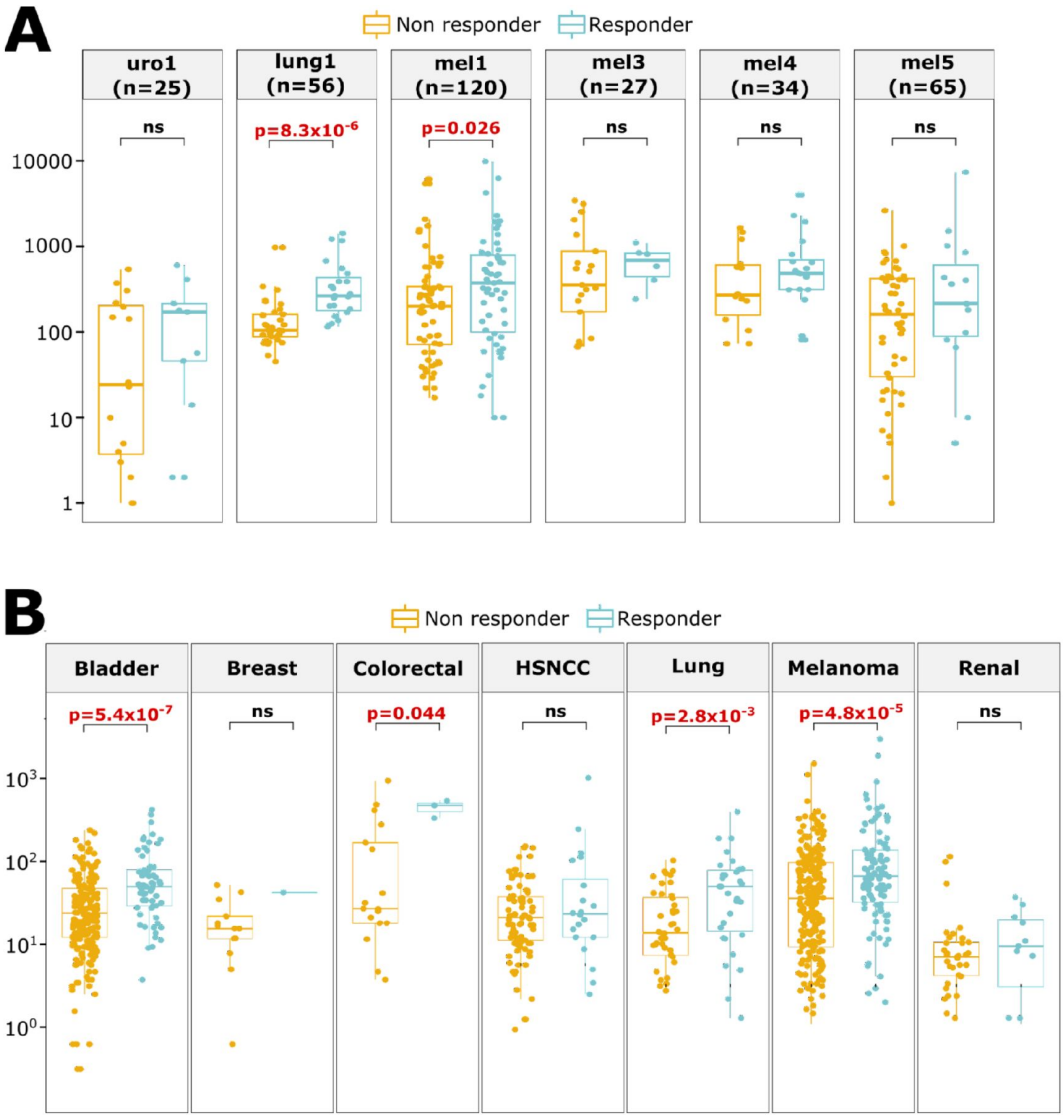

Figure S1

TMB association with clinical benefit from ICB across cancers

(A) Association of TMB with response to ICB across three cancer types from CPI800+. Only melanoma and non-small cell lung cancer have a significantly different TMB between responders and non-responders. **(B)** Association of TMB with response to ICB across seven cancer types from CPI1000+. Melanoma, non-small cell lung, bladder and colorectal cancer have a significantly different TMB between responders and non-responders.

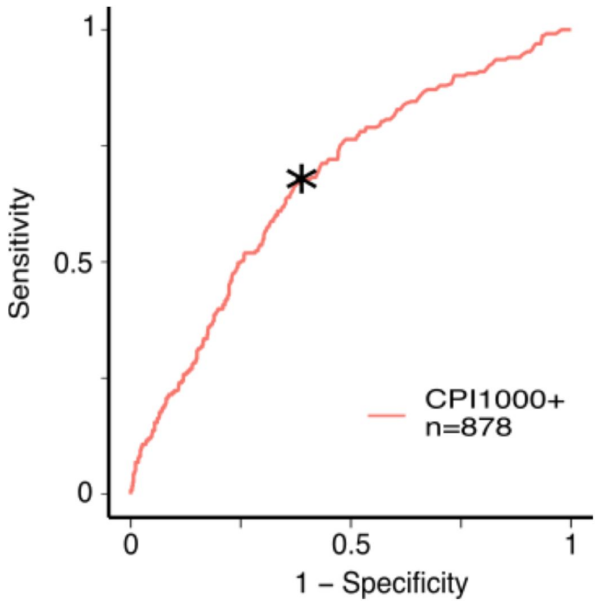

Figure S2

TMB as a biomarker of response to immunotherapy

ROC curve for CPI1000+. The asterisk represents the Youden index cutoff

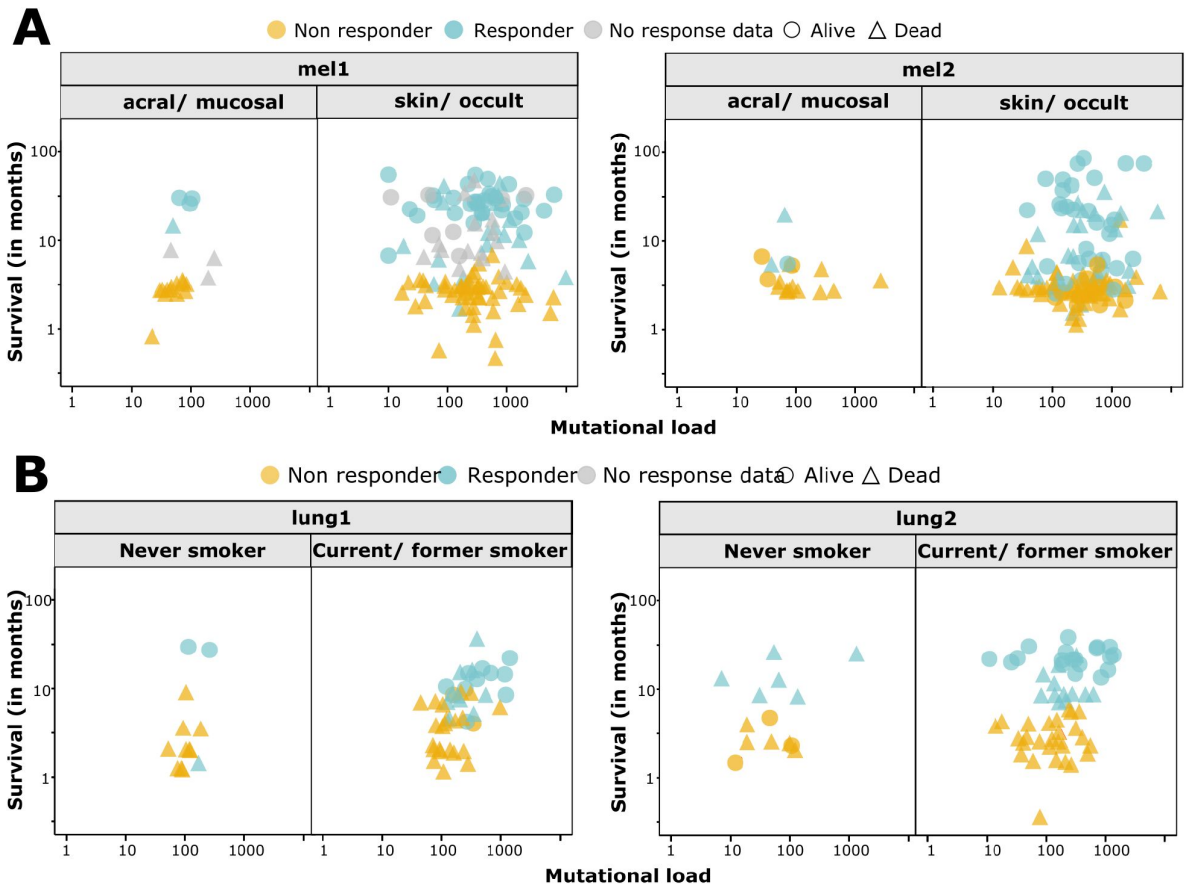

Figure S3

TMB association with progression-free survival post-immunotherapy

(A) (B) Plots of progression-free survival and TMB for melanoma and lung cancer ICB cohorts labeled by cancer subtype, showing the lack of correlation or of an obvious TMB cutoff.

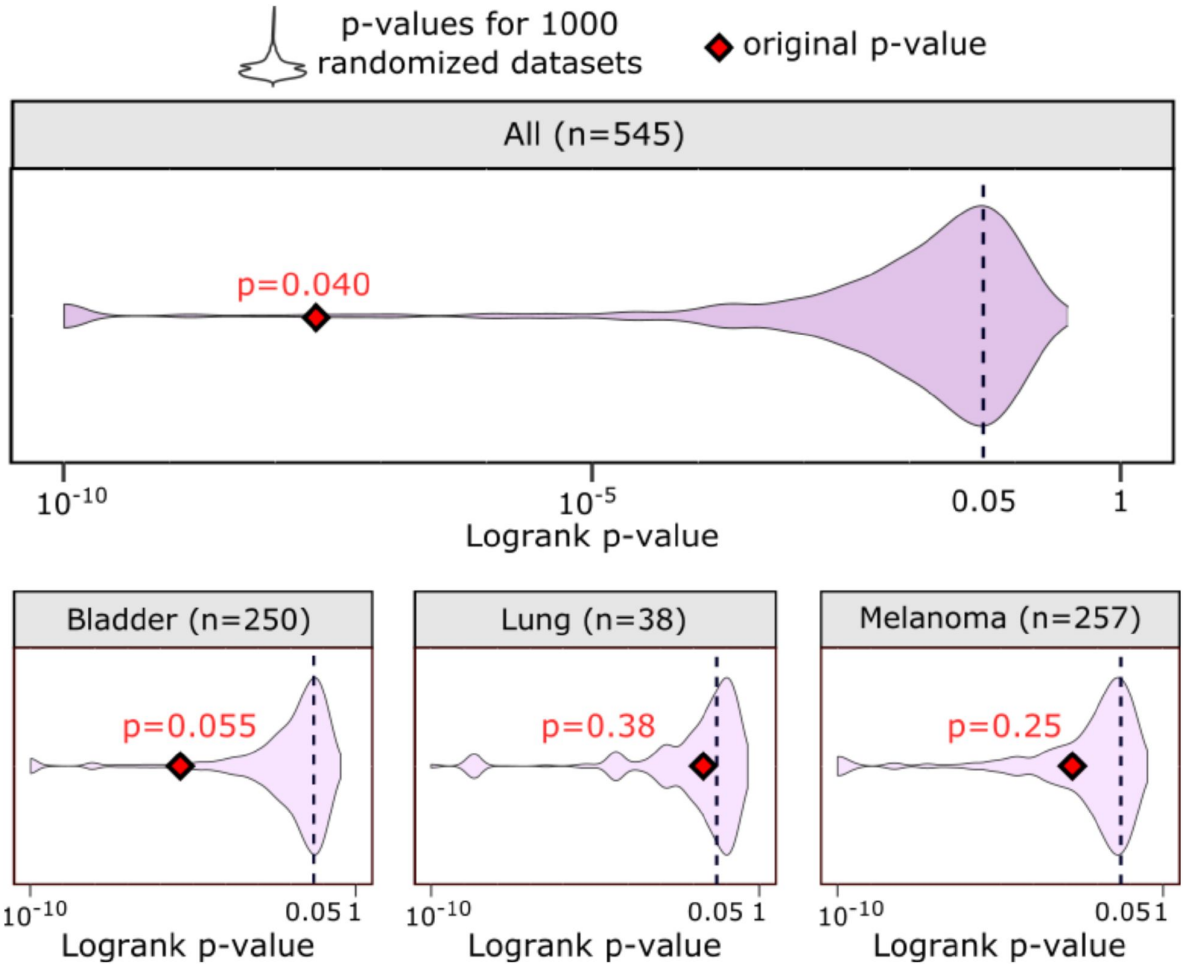

Figure S4

TMB association with overall survival post-immunotherapy

Results of the randomization analysis in CPI1000+ (p-values < 10-10 not shown).When cancer types of CPI1000+ were combined, a nominally significant p-value (p=0.04) arises, likely due to cancer types with different TMB ranges showing significantly different survival rates to ICB.

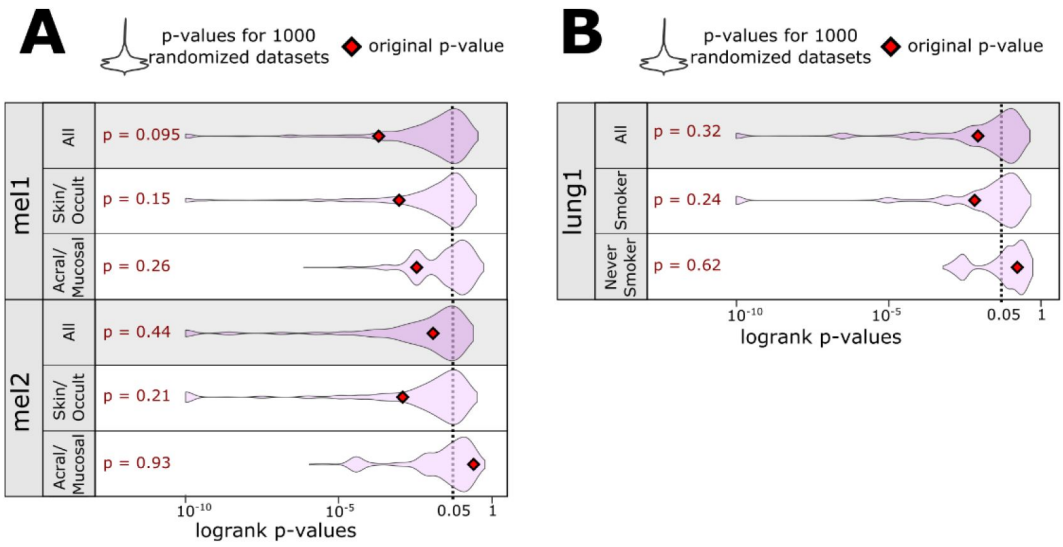

Figure S5

TMB association with overall survival post-immunotherapy

(A) Randomization analysis results in mel1 and mel2 and stratification by subtypes (p-values < 10^{-10} not shown) **(B)** Randomization analysis results in lung1 and stratification by subtypes (p-values < 10^{-10} not shown). When corrected for multiple hypotheses all cohorts fail to provide a statistically significant cutoff.

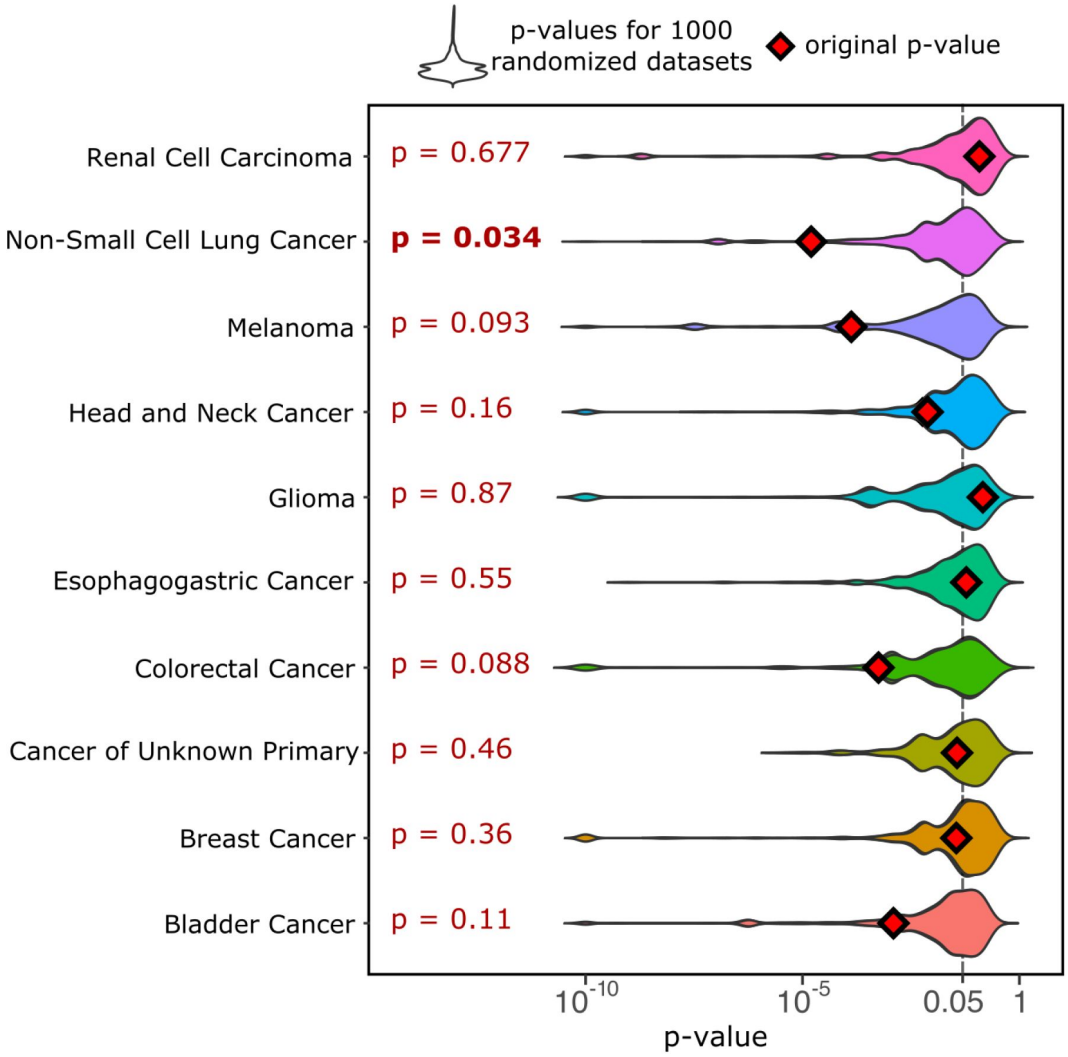

Figure S6

TMB association with overall survival post-immunotherapy

Randomization analysis results in multiple cancer types with MSK-IMPACT targeted next-generation sequencing data (p-values < 10^{-10} not shown)

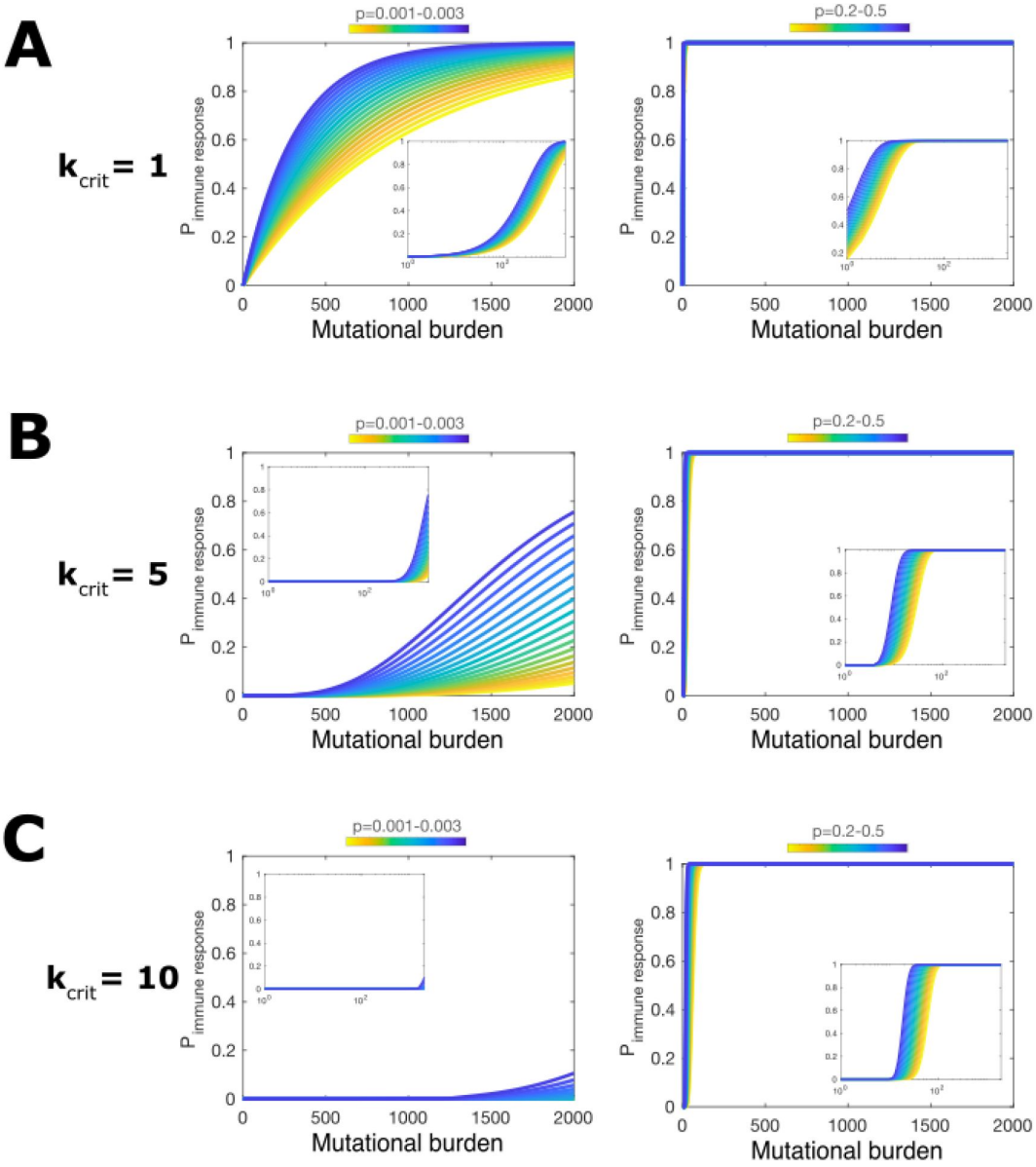

Figure S7

Components of cancer immunogenicity

The probability of immune response $P_{immune\ responce}$ as a function of TMB for **(A)** k_{crit} =1, **(B)** k_{crit} = 5 and **(C)** k_{crit} =10 for low (left panels) and high (right panels) ranges of p, the probability of a mutation to be immunogenic.

References

1. Postow M. A., Callahan M. K., Wolchok J. D. (2015) **Immune Checkpoint Blockade in Cancer Therapy** *J. Clin. Oncol* **33**:1974–1982

2. Hodi F. S., et al. (2010) **Improved survival with ipilimumab in patients with metastatic melanoma** *N. Engl. J. Med* **363**:711–723

3. Dang T. O., Ogunniyi A., Barbee M. S., Drilon A. (2016) **Pembrolizumab for the treatment of PD-L1 positive advanced or metastatic non-small cell lung cancer** *Expert Rev. Anticancer Ther* **16**:13–20

4. Ribas A., Wolchok J. D. (2018) **Cancer immunotherapy using checkpoint blockade** *Science* **359**:1350–1355

5. Hugo W., et al. (2017) **Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma** *Cell* **168**

6. Snyder A., et al. (2014) **Genetic basis for clinical response to CTLA-4 blockade in melanoma** *N. Engl. J. Med* **371**:2189–2199

7. Van Allen E. M., et al. (2015) **Genomic correlates of response to CTLA-4 blockade in metastatic melanoma** *Science* **350**:207–211

8. Rizvi N. A., et al. (2015) **Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer** *Science* **348**:124–128

9. Hellmann M. D., et al. (2018) **Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer** *Cancer Cell* **33**:843–852

10. Cristescu R., et al. (2018) **Pan-tumor genomic biomarkers for PD-1 checkpoint blockade–based immunotherapy** *Science* **362** https://doi.org/10.1126/science.aar3593

11. Samstein R. M., et al. (2019) **Tumor mutational load predicts survival after immunotherapy across multiple cancer types** *Nat. Genet* **51**:202–206

12. Yarchoan M., Hopkins A., Jaffee E. M. (2017) **Tumor Mutational Burden and Response Rate to PD-1 Inhibition** *N. Engl. J. Med* **377**:2500–2501

13. Chan T. A., et al. (2019) **Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic** *Ann. Oncol* **30**:44–56

14. Osipov A., et al. (2020) **Tumor Mutational Burden, Toxicity, and Response of Immune Checkpoint Inhibitors Targeting PD(L)1, CTLA-4, and Combination: A Meta-regression Analysis** *Clin. Cancer Res.* https://doi.org/10.1158/1078-0432.CCR-20-0458

15. Sharma P., Hu-Lieskovan S., Wargo J. A., Ribas A., Primary Adaptive (2017) **and Acquired Resistance to Cancer Immunotherapy** *Cell* **168**:707–723

16. Heeke S., Hofman P. (2018) **Tumor mutational burden assessment as a predictive biomarker for immunotherapy in lung cancer patients: getting ready for prime-time or not?** *Translational Lung Cancer Research* **7**

17. Food and Drug Administration (2020) **Food and Drug Administration, Highlights of prescribing information: KEYTRUDA. https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/125514s068lbl.pdf. 2020**

18. Sicklick J. K., et al. (2019) **Molecular profiling of cancer patients enables personalized combination therapy: the I-PREDICT study** *Nat. Med* **25**:744–750

19. McGranahan N., Swanton C. (2019) **Neoantigen quality, not quantity** *Sci. Transl. Med* **11**

20. Miao D., et al. (2018) **Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors** *Nat. Genet* **50**:1271–1281

21. Miao D., et al. (2018) **Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma** *Science* **359**:801–806

22. Braun D. A., et al. (2020) **Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma** *Nat. Med* **26**:909–918

23. Snyder A., et al. (2017) **Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: An exploratory multi-omic analysis** *PLoS Med* **14**

24. Liu D., et al. (2019) **Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma** *Nat. Med* **25**:1916–1927

25. Riaz N., et al. (2017) **Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab** *Cell* **171**:934–949

26. Litchfield K., et al. (2021) **Meta-analysis of tumor and T cell intrinsic mechanisms of sensitization to checkpoint inhibition** *Cell* **184**:596–614

27. Wood M. A., Weeder B. R., David J. K., Nellore A., Thompson R. F. (2020) **Burden of tumor mutations, neoepitopes, and other variants are weak predictors of cancer immunotherapy response and overall survival** *Genome Med* **12**

28. Pepe M. S., Feng Z., Janes H., Bossuyt P. M., Potter J. D. (2008) **Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design** *J. Natl. Cancer Inst* **100**:1432–1438

29. Westfall P. H., Stanley Young S. (1993) **Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment** John Wiley & Sons

30. Dudoit S., Shaffer J. P., Boldrick J. C. (2003) **Multiple Hypothesis Testing in Microarray Experiments** *Stat. Sci* **18**:71–103

31. Shlyakhter A., Mirny L., Vlasov A., Wilson R. (1996) **Monte Carlo modeling of epidemiological studies** *Human and Ecological Risk Assessment: An International Journal* **2**:920–938

32. Xiao Y., Freeman G. J. (2015) **The microsatellite instable subset of colorectal cancer is a particularly good candidate for checkpoint blockade immunotherapy** *Cancer Discov* **5**:16–18

33. Turajlic S., et al. (2017) **Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis** *Lancet Oncol* **18**:1009–1021

34. Weghorn D., Sunyaev S. (2017) **Bayesian inference of negative and positive selection in human cancers** *Nat. Genet* **49**:1785–1788

35. Zamora A. E., et al. (2019) **Pediatric patients with acute lymphoblastic leukemia generate abundant and functional neoantigen-specific CD8+ T cell responses** *Sci. Transl. Med* **11**

36. Efremova M., et al. (2018) **Targeting immune checkpoints potentiates immunoediting and changes the dynamics of tumor evolution** *Nat. Commun* **9**

37. Van den Eynden J., Jiménez-Sánchez A., Miller M. L., Larsson E. (2019) **Lack of detectable neoantigen depletion signals in the untreated cancer genome** *Nat. Genet* **51**:1741–1748

38. McFarland C. D., Mirny L. A., Korolev K. S. (2014) **Tug-of-war between driver and passenger mutations in cancer and other adaptive processes** *Proc. Natl. Acad. Sci. U. S. A* **111**:15138–15143

39. Persi E., Wolf Y. I., Leiserson M. D. M., Koonin E. V., Ruppin E. (2018) **Criticality in tumor evolution and clinical outcome** *Proc. Natl. Acad. Sci. U. S. A* **115**:E11101–E11110

40. Marabelle A., et al. (2020) **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* **21**:1353–1365

41. Spranger S., et al. (2016) **Density of immunogenic antigens does not explain the presence or absence of the T-cell–inflamed tumor microenvironment in melanoma** *Proceedings of the National Academy of Sciences* **113**:E7759–E7768 https://doi.org/10.1073/pnas.1609376113

42. Sarkizova S., Hacohen N. (2017) **How T cells spot tumour cells** *Nature* **551**:444–446

43. Łuksza M., et al. (2017) **A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy** *Nature* **551**:517–520

44. Gurjao C., et al. (2019) **Intrinsic Resistance to Immune Checkpoint Blockade in a Mismatch Repair–Deficient Colorectal Cancer** *Cancer Immunol Res* **7**:1230–1236

45. Vega D. M., et al. (2021) **Aligning tumor mutational burden (TMB) quantification across diagnostic platforms: phase II of the Friends of Cancer Research TMB Harmonization Project** *Ann. Oncol* **32**:1626–1636

46. Gao J., et al. (2013) **Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal** *Sci. Signal* **6**

47. Feng X., Li L., Wagner E. J., Li W. (2018) **TC3A: The Cancer 3' UTR Atlas** *Nucleic Acids Res* **46**:D1027–D1030

Editors

Reviewer #2 (Public review):

The manuscript points out that TMB cut-offs are not strong predictors of response to immunotherapy or overall survival. By randomly shuffling TMB values within cohorts to simulate a null distribution of log-rank test p-values, they show that under correction, the statistical significance of previously reported TMB cut-offs for predicting outcomes is questionable. There is a clinical need for a better prediction of treatment response than TMB alone can provide. However, the analysis does not convincingly refute the validity of the well-known pan-cancer correlation between TMB and immunotherapy response. (In a supplemental analysis, the authors attempt to demonstrate a lack of correlation by specifically removing the most supportive cancer types from a pan-cancer correlation test.) The failure to detect significant TMB cut-offs may be due to insufficient power, as the examined cohorts have relatively low sample sizes. A power analysis would be informative of what cohort sizes are needed to detect small to modest effects of TMB on immune response.

The manuscript provides a simple model of immunogenicity that is tailored to be consistent with a claimed lack of relationship between TMB and response to immunotherapy. Under the model, if each mutation that a tumor has acquired has a relatively high probability of being immunogenic (~10%, they suggest), and if 1-2 immunogenic mutations is enough to induce an immune response, then most tumors produce an immune response, and TMB and response should be uncorrelated except in very low-TMB tumors. The question then becomes whether the response is sufficient to wipe out tumor cells in conjunction with immunotherapy, which is essentially the same question of predicting response that motivated the original analysis. While TMB alone is not an excellent predictor of treatment response, the pan-cancer correlation between TMB and response/survival is highly significant, so the model's only independent prediction is wrong. Additionally, experiments to predict and validate neoepitopes suggest that a much smaller fraction of nonsynonymous mutations produce immune responses (1,2).

A key idea that is overlooked in this manuscript is that of survivorship bias: self-evidently, none of the mutations found at the time of sequencing have been immunogenic enough to provoke a response capable of eliminating the tumor. While the authors suggest that immunoediting "is inefficient, allowing tumors to accumulate a high TMB," the alternative explanation fits the neoepitope literature better: most mutations that reach high allele frequency in tumor cells are not immunogenic in typical (or patient-specific) tumor environments. Of course, immunotherapies sometimes succeed in overcoming the evolved immune evasion of tumors. Higher-TMB tumors are likely to continue to have higher mutation rates after sequencing; increased generation of new immunogenic mutations may partially explain their modestly improved responses to therapy.

References:
(1) Wells, D. K. et al. Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. Cell 183, 818-834.e13 (2020).

(2) Yadav, M. et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. Nature 515, 572-576 (2014).

https://doi.org/10.7554/eLife.87465.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***Reviewer #1 (Public Review):***
>
> *In their manuscript entitled: "Is tumor mutational burden predictive of response to immunotherapy?", Gurjao and colleagues discuss the use of tumor mutational burden (TMB) as a predictive biomarker for cancer patients to respond to immune checkpoint blockage (ICB). By analyzing a large cohort of 882 patient samples across different tumor types they find either little or no association of TMB to the response of ICB. In addition, they showed that finding the optimal cutoff for patient stratification lead to a severe multiple testing problem. By rigorously addressing this multiple testing problem only non-small cell lung cancer out of 10 cancer types showed a statistically significant association of TMB and response to ICB. Nevertheless, it is clearly shown that in any case the rate of misclassification is too high that TMB alone would qualify as a clinically suitable biomarker for ICB response. Finally, the authors demonstrate with a simple mathematical model that only a few strong immunogenic mutations would be sufficient for an ICB response, thereby showing that also patients with a low TMB score could benefit from immunotherapy. The manuscript is clearly written, the results are well presented and the applied methods are state-of-the-art.*

We would like to thank the reviewer for their thoughtful suggestions and efforts towards improving our manuscript. We address below the reviewer's recommendations.

> ***Reviewer #1 (Recommendations For The Authors):***
>
> *(1) The method used for mutation call can also influence the TMB score. Mutation data was downloaded from public databases and not re-called for this study, a potential caller bias could be present. What was the calling strategy of the used data sets? For the present study, I don't think that this is crucial because different callers or post-call processing would be used at different sites to determine TMB. I think it should the mutation calling bias should also be discussed in the manuscript as another shortcoming for TMB as a biomarker for ICB response.*

We thank the reviewer for this comment. Mutational data was not aggregated across studies and caller bias would thus not have any impact on the results of this manuscript. In addition, we further clarified the role of mutation calling bias in the Discussions section.

"Although attractive and scalable, TMB does not consider the effect of specific mutations (missense, frameshift etc), their presentation and clonality (19), nor the state of the tumour, its microenvironment, and interactions with the immune system that can be integrated into potentially better predictors of response to ICB (43, 44). In addition, another major limitation of TMB is the lack of standardized measures. This includes the lack of standard sequencing methods to assess TMB: TMB can be measured from Whole-Exome sequencing, Whole-Genome sequencing, targeted panel and even RNA sequencing. This also includes biases introduced by using different mutation calling pipelines resulting in different TMB, sequencing depth and different characteristics of the samples (e.g. low purity samples typically yield lower TMB)."

> *(2) In their mathematical model of neoantigens and immunogenicity it is assumed that the probability of a mutation to be immunogenic is constant for all mutations. In reality this is certainly not satisfied. However, the central conclusion from the model still holds. I think that this is important to discuss in the manuscript.*

We thank the reviewer for this suggestion and now consider the case where each mutation has its own probability p(i) of being immunogenic.

"Our model shows that achieving about constant $P\{immune\ response\}$ for $N > 10-20$ mutations, requires $p \sim 0.1$ for $k_{crit} = 1$, and $p \sim 0.2$ for $k_{crit} > 1$. The same argument holds when each mutation has its own probability to be immunogenic $p(i)$, then $P\{immune\ response\} = 1 - \Pi_{i=1}^{N}(1 - p(i)) = 1 - exp(-\bar{p}N)$, where is the mean probability of a mutation to be immunogenic. Thus only the average probability of a mutation to be immunogenic matters. In summary, we find that the model agrees with clinical data if individual non-synonymous mutations have, on average, $p \sim 10-20\%$ chance for triggering an immune response."

> *(3) In the mathematical formula on page 8, C_N^k is the binomial coefficient. This should be stated or written out.*

Thank you for pointing this out. Corrected.

"Due to immunodominance, only a few k_{crit} immunogenic mutations are sufficient to elicit a full k_{crit} immune response. Hence, the probability for a cancer with N (=TMB) mutations to elicit an immune response is then the probability of having k or more immunogenic mutations among :

$$P\{immune\ response\} = \sum_{k=k_{crit}}^{N} \binom{N}{k} p^{k}(1-p)^{N-k} = 1 - \sum_{k=0}^{k_{crit}-1} \binom{N}{k} p^{k}(1-p)^{N-k},$$

which is the CDF of a binomial distribution."

> *(4) The mathematical model provides an explanation that tumors with a low TMB can also respond on ICB. It cannot explain tumors with high TMB lacking ICB response. An explanation of this phenomenon is discussed in the paper but I think also the impact of the tumor immune microenvironment should be mentioned here.*

As we explained in the presentation of the model, even immunogenic tumors elicit response to ICB with some probability. In the revision we write:

"$P\{clinical\ response\} = P\{clinical\ response \mid immune\ response\} \cdot P\{immune\ response\}$, where $P\{clinical\ response \mid immune\ response\}$ is the probability of clinical response, given that cancer elicits an immune response which is complex and depends on many factors including tumor immune microenvironment. Yet the prerequisite for the clinical response is the immune response $P\{immune\ response\}$ that we focus on."

> ***Reviewer #2 (Public Review):***
>
> *The manuscript points out that TMB cut-offs are not strong predictors of response to immunotherapy or overall survival. By randomly shuffling TMB values within cohorts to simulate a null distribution of log-rank test p-values, they show that under correction, the statistical significance of previously reported TMB cut-offs for predicting outcomes is questionable.*

We would like to thank the reviewer for their thoughtful suggestions and efforts towards improving our manuscript.

> *There is a clinical need for a better prediction of treatment response than TMB alone can provide. However, no part of the analysis challenges the validity of the well-known pan-cancer correlation between TMB and immunotherapy response.*

We address the pan-cancer correlation in the supplemental text and Figure S3. We realized the supplemental text was missing in eLife submission and included in the bioRxiv only. We apologize for this oversight. In particular, we show that the "well-known pan-cancer correlation" is largely based on a few outlier cancer subtypes - MSI colorectal cancers and uveal/ ocular melanomas. We show that when we remove these cancer types from the pan-cancer dataset, the correlation becomes non-significant for the remaining 15 cancer types.

> *The failure to detect significant TMB cut-offs may be due to insufficient power, as the examined cohorts have relatively low sample sizes. A power analysis would be informative of what cohort sizes are needed to detect small to modest effects of TMB on immune response.*

Since we see no effect, we cannot perform a power analysis. Moreover, increasing cohort sizes cannot increase the effect -- dramatic misclassification of responders (the fraction of responders below the treatment cutoff) would remain the same, making TMB unsuitable for clinical decision-making.

> *The manuscript provides a simple model of immunogenicity that is tailored to be consistent with a claimed lack of relationship between TMB and response to immunotherapy. Under the model, if each mutation that a tumor has acquired has a relatively high probability of being immunogenic (~10%, they suggest), and if 1-2 immunogenic mutations is enough to induce an immune response, then most tumors produce an immune response, and TMB and response should be uncorrelated except in very low-TMB tumors.*

Contrary to reviewer's suggestion, our modeling is not tailored to be consistent with the lack of association between TMB and response. On the contrary, we found the model has two regimes: the first regime (where p<<1) in which higher TMB leads to a higher probability of response, which doesn't agree with the data , and the second regime (p~0.1) in which cancers with TMB>10-20 are immunogenic, consistent with the clinical data.

We further expanded on these key points in the Results:

"The model shows two different behaviors. If individual mutations are unlikely to be immunogenic ($p \ll 1$) , e.g. due to a low probability of being presented, the probability of response increases gradually with TMB (Figure 5B). The neoantigen theory generally expects such gradual increase in immunogenicity of cancer with TMB. Yet, available data (Figure 2) don't show such a trend.

On the contrary, if mutations are more likely to be immunogenic p~0. 1, the probability of response quickly saturates (Figure 5C), making such tumors respond to ICB irrespective of TMB, as we observed in clinical data."

We also expanded on these key points in the Introduction:

"We develop a simple model that is based on the neoantigen theory and find that it has two regimes. In one regime, the probability of response increases gradually with TMB, as commonly believed. Yet in the other, the probability of response saturates after a few mutations, making a chance to respond independent of TMB. Our analysis of the clinical data is consistent with the latter regime. Thus our model shows that the neoantigen theory is fully consistent with the lack of association between TMB and response."

> *The question then becomes whether the response is sufficient to wipe out tumor cells in conjunction with immunotherapy, which is essentially the same question of predicting response that motivated the original analysis. While TMB alone is not an excellent predictor of treatment response, the pan-cancer correlation between TMB and response/survival is highly significant, so the model's only independent prediction is wrong.*

Our study indicates that TMB is a very poor predictor (writing that it's "not an excellent predictor of treatment response" is understatement). Moreover we show that a widely believed "pan-cancer correlation" is shaky as well (Supplemental text and Figure S3). So we don't see any contradictions between the model and the data.

> *Additionally, experiments to predict and validate neoepitopes suggest that a much smaller fraction of nonsynonymous mutations produce immune responses1,2.*

We agree with the reviewer. That's exactly what the model suggests.

> *A key idea that is overlooked in this manuscript is that of survivorship bias: self-evidently, none of the mutations found at the time of sequencing have been immunogenic enough to provoke a response capable of eliminating the tumor. While the authors suggest that immunoediting "is inefficient, allowing tumors to accumulate a high TMB," the alternative explanation fits the neoepitope literature better: most mutations that reach high allele frequency in tumor cells are not immunogenic in typical (or patient-specific) tumor environments. Of course, immunotherapies sometimes succeed in overcoming the evolved immune evasion of tumors. Higher-TMB tumors are likely to continue to have higher mutation rates after sequencing; increased generation of new immunogenic mutations may partially explain their modestly improved responses to therapy.*

We disagree with reviewers' assertion that survivorship bias could explain observed phenomena. If immunogenic mutations that arise during cancer development were eliminated (by purifying selection, i.e. reduced fitness or cellular death) then observed mutations would carry noticeable signatures of purifying selection. On the contrary, cancer genomic data shows incredibly weak signals of purifying selection on non-synonymous mutations (Weghorn and Sunyaev, Nature Genetics 2017), and observed passenger mutations are practically indistinguishable from random in their effect on proteins (McFarland et al PNAS 2013).

We do agree with the statement that "most mutations … in tumor cells are not immunogenic". In fact that's exactly what our model predicts: (1-p)~90% of mutations in the model are non-immunogenic, while remaining p~10% being sufficient to trigger an immune response. We clarify this in the text of the paper: "On the contrary, if mutations are more likely to be

immunogenic p~0. 1, the probability of response quickly saturates (Figure 5C), making such tumors respond to ICB irrespective of TMB, as we observed in clinical data. ”

> ***Reviewer #2 (Recommendations For The Authors):***
>
> *Abstract*
>
> *Defining TMB as "number of non-synonymous mutations": while TMB is not consistently defined throughout the literature, it is usually given as a rate rather than a total count, and sometimes synonymous mutations are included. Consider adopting the definition used by the TMB Harmonization Project: "number of somatic mutations per megabase of interrogated genomic sequence.3"*

We thank the reviewer for their comment,

> *Be more specific about your findings, so that abstract readers can get some understanding of your proposed explanation for the "immunogenicity of neoantigens and the lack of association between TMB and response."*

We thank the reviewer for their comment. We modified the abstract to explain that the theory we developed expands the neoantigen theory yet can be consistent with the observed lack of association between TMB and response:

"Second, we develop a model that expands the neoantigen theory and can be consistent with both immunogenicity of neoantigens and the lack of association between TMB and response. Our analysis shows that the use of TMB in clinical practice is not supported by available data and can deprive patients of treatment to which they are likely to respond.”

> *Introduction*
>
> *Again, consider using a more standard definition of TMB.*

We thank the reviewer for their comment. Our study did not seek to harmonize TMB across the datasets and we thus used the total number of mutations rather than the mutational rate often used for comparison across different datasets.

> *Expand the introduction to provide a preview of the purpose and direction of your analysis. The current draft reveals only that the analysis will relate to TMB.*

We expanded the introduction providing the motivation, the approach, and the summary of main findings.

“Using a biomarker to stratify and prioritize patients for treatment runs a risk of depriving patients who have a chance to respond to a life-saving treatment. High variability of response makes relying on a predictor particularly risky. Hence, we revisit original data that were used to establish correlation between TMB and response. We tested TMB as a predictor of both binary responder/non-responder labels from original clinical studies, as well as continuous survival data. We also investigated whether a TMB threshold could distinguish patients with high and low survival after multiple hypothesis testing. We find that no TMB threshold performs better on the clinical data than on randomized ones.

We further show that irrespective of the strategy to choose the threshold, even if we were to employ the optimal TMB cutoff, it would still lead to about 25% of responders falling below the treatment prioritization threshold. In addition, we re-examine the pan-cancer association of TMB with response rate to ICB.

“Finally we revisit the neoantigen theory that was the rationale for using TMB as a predictor of response to immunotherapy. The theory stipulates that non-synonymous mutations can lead to the production of unique antigens (_neo_antigens) that are recognized by the immune system as foreign, triggering the immune response to cancer. The theory further assumes that the more mutations a cancer has, the more likely it triggers the immune system, and the more likely it will benefit from immunotherapy. We develop a simple model that is based on the neoantigen theory and find that it has two regimes. In one regime, the probability of response increases gradually with TMB, as commonly believed. Yet in the other, the probability of response saturates after a few mutations, making a chance to respond independent of TMB. Our analysis of the clinical data is consistent with the latter regime. Thus our model shows that the neoantigen theory is fully consistent with the lack of association between TMB and response.”

> *Section: Is TMB associated with response after treatment?*
>
> *The claim that after excluding melanoma and some colorectal cancers, there is no relationship between TMB and response rates in pan-cancer studies cites references 12 and 14. In reference 12 (Yarchoan et al.), it is clear from glancing at their Figure 1 that a pan-cancer correlation between TMB and response would remain with these cancer types excluded. This discrepancy requires explanation. "Supplementary text" is cited for this claim, but it was not included in the file that I received.*

We address the pan-cancer correlation in the supplemental text and Figure S3. While the figure was available, we realized the supplemental text was missing in eLife submission. We apologize for this oversight.

> *Plots of survival and TMB do not show "visible correlation": Please strengthen this claim with an appropriate statistical test.*

We expand the figure caption to explain the following:

“Plots of progression-free survival and TMB for melanoma and lung cancer ICB cohorts show the lack of correlation or of an obvious TMB cutoff. Computing a simple correlation for survival and censored data cannot correctly represent the dependence since patients who are alive live longer than the reported survival, and limiting correlation to patients who are dead would bias the analysis. Thus other survival statistics are used through the paper.”

> *Section: Model reconciles neoantigen theory and data*
>
> *Page 8: In the probability formula, the C term is not defined. My guess is that it means choose(N, k).*
>
> *Please clarify.*

Thank you for pointing this out. Corrected using more conventional notation.

$$P\{immune\ response\} = \sum_{k=k_{crit}}^{N} \binom{N}{k} p^k (1-p)^{N-k} = 1 - \sum_{k=0}^{k_{crit}-1} \binom{N}{k} p^k (1-p)^{N-k}$$

which is the CDF of a binomial distribution.

> *Page 8: Assuming the above, P(immune response) = P(X >= k_crit); where X~Bin(N, p). The formula should be explicitly introduced in terms of the CDF of the binomial distribution to prevent readers from thinking the wheel is being re-invented.*

We thank the reviewer for pointing this out, we modified the equation in the text to make it easier to see this point (see above). We refrain from going further since the CDF of a binomial distribution doesn't have a closed form and can only be written as the regularized incomplete beta function.

> *Page 9: Missing word in "allowing cancers with as little as mutations to be"*

We thank the reviewer for pointing this out, we modified the text accordingly.

> *See comments in public review. In brief, I think a convincing case is made regarding the significance of TMB cut-offs as predictors of survival within cancer types, but frankly this elementary model is not compelling.*
>
> *Section: Materials and Methods*
>
> *In the manuscript, it is stated that TMB is accepted as reported by data sources. Since most of the comparisons in the manuscript are within-data-source, that is acceptable. However, it should be ensured that TMB measurements are comparable between samples within each source. For example, when TMB is reported as a total mutation count, it can be verified that all samples have the same coverage, or measurement can be converted to mutations per megabase of coverage. In the same vein, if this manuscript's definition of TMB only includes nonsynonomous mutations, it should be confirmed that the TMB reported by data sources excludes synonymous mutations.*

We thank the reviewer for their comment. We leverage total TMB as reported in the original studies claiming an association between TMB and response/ survival.

> *Figure S2: Instead of writing "the Youden index associated cutoffs is also plotted," it can be stated that the asterisk represents the Youden index cutoff, or a legend can be added that provides this information.*

We thank the reviewer for pointing this out, we modified the text accordingly.

https://doi.org/10.7554/eLife.87465.2.sa0