DeepTMB: An uncertainty-aware deep calibration of tumor mutational burden with a synthetic tumor-only dataset

Jordan Anaya[1] **and Alexander S. Baras**[1,2,3]

[1]**Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA**
[2]**The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA**
[3]**Bloomberg Kimmel Institute for Cancer Immunotherapy, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA**

Corresponding author:
Alexander S. Baras[1,2,3]

Email address: baras@jhmi.edu

ABSTRACT

Background: Tumor mutational burden (TMB) is commonly investigated as a biomarker for immune checkpoint blockade (ICB) therapy, but lack of uniform standards in its calculation may be preventing consensus recommendations. Gene panels have mostly supplanted exome sequencing for calculating TMB, but these panels have a strong tendency to overestimate TMB. With TMB cutoffs being developed with panel-based estimates and suggestions to use The Cancer Genome Atlas (TCGA) exome data to model panel data, there is a need to understand how to properly compare values from different assays.

Design: Our approach to TMB calibration allows for a nonlinear fit and a heteroscedastic error distribution, utilizes a per-batch read depth augmentation to upsample the exomic data to what would be seen at panel depth, and is trained on a novel tumor-only synthetic dataset generated by utilizing germline as well as somatic calls.

Results: We were able to accurately model the distribution of tumor-normal data. Removing hotspots and including synonymous mutations resulted in better regression metrics, as did including NCIt label. Correcting for read depth resulted in a slight increase in the spread of the data but did not affect the model fit. When applying a model trained on tumor-normal data to synthetic tumor-only input there was a dramatic drop in the positive predictive value (PPV) at a TMB of 10. This was largely mitigated by training with synthetic tumor-only input (up to over a 200% improvement in PPV).

Conclusions: A deep learning model is capable of accurately modeling the tumor-normal TMB distribution. Using a synthetic tumor-only input for training appears to be a viable strategy for calibrating tumor-only panels.

INTRODUCTION

Immune checkpoint blockade (ICB) has received FDA approval for several cancer types along with any cancer with mismatch-repair deficiency[1]. However, not all patients respond and may experience adverse events[2], necessitating the search for prognostic indicators of response[3]. Because the cancer types where ICB has been effective are generally characterized by high tumor mutational burden (TMB)[4] it was reasonable to explore TMB as a biomarker[5–9], and for both survival and response TMB has been predictive across multiple cancers[10,11].

Although an appealing metric, TMB values can be influenced by assay type, sample quality, and bioinformatics pipeline[12,13]. Exome sequencing is viewed as the gold standard assay for determining TMB, but smaller gene-focused assays which only cover approximately 1 Mb of the ≈30 Mb coding exome are often used instead since they are already part of standard of care to identify potential drug targets, require less tumor material, and are less expensive. When synthetic panel-restriction experiments have been performed with exome data, the values from synthetic panel-restriction of exome data generally correlate well with what was derived from the full exome data[14,15], justifying the use of panels

for this purpose. Importantly, the use of some panel-based assays for characterizing likelihood of ICB response has been approved by the FDA[16,17].

It is important to note that most such gene panels were not specifically designed to accurately measure TMB; rather, they were simply repurposed for this metric, and the approval of gene-panel derived TMB as a biomarker in this context has not come without controversy[18]. Panels are biased to include genes commonly mutated in cancer, have higher depth of sequencing than exome data, and often don't have a paired normal sample from the patient to remove germline variants. Each of these factors bias panels towards reporting a higher value of TMB than would be seen with exome sequencing, which unfortunately creates the potential for a lower positive predictive value (PPV) than expected based on exome restriction experiments. Furthermore, it is unclear how to properly translate TMB cutoff(s) determined from a given gene panel assay to another. These issues are widely recognized and an international consortium has been created with the goal of harmonizing TMB values[19].

In an ideal scenario a reference dataset which contained matching high-depth exome data and panel sequencing for every commonly used assay would be used for calibration. But this data does not exist at scale, therefore various calibration efforts have been performed by computationally restricting variable-depth exome data to a given gene panel's genomic footprint, thereby generating the input for modeling exome derived TMB. Using this approach and The Cancer Genome Atlas (TCGA) data, researchers have modeled the uncertainty of TMB estimation that is inherent to subsampling and gene biases[20,21], but have not directly considered the contributions of read depth and tumor-only sequencing. In an attempt to account for data processing variability across assays, various laboratories were asked to process TCGA variant calls into their own gene panel TMB estimates[22]. This processing may have included hotspot removal, synonymous exclusion, variant allele frequency (VAF) restriction, germline filtering, or potentially internal calibration. However, because the starting data used in many of these studies already has germline mutations subtracted via a matched normal, we believe these synthetic panel estimates may not accurately reflect actual tumor-only panel data and that it would be problematic to apply models calibrated on tumor-normal data to data from tumor-only assays.

When TMB is calculated without a matched-normal sample laboratories rely on various population databases to filter germline variants. Laboratories may also use an algorithm that tries to distinguish somatic and germline using the VAFs[23]. Regardless of the approach, there will usually be some error in the ability to accurately identify all germline variants and these additional so-called "private" germline variants erroneously contribute to TMB estimates in the context of tumor-only sequencing, an artifact which cannot be accounted for when using germline-subtracted tumor-normal data.

The sequencing depth of the TCGA data is primarily at 50-100X whereas panel data is often >300X. Mutation callers are known to have limited sensitivity at lower sequencing depths[24,25], and when panel data has been downsampled to 100X only 83.5% of mutations were re-called[26]. Because of this we previously modeled how many more mutations would be seen in the TCGA data if sequenced at panel depth[27]. We estimated that on average the TCGA MC3 working group called about 83% of the mutations that would have been called at panel depth. While the shape of the regression curves won't change if the independent and dependent variables undergo the same linear transformation, if the MC3 mutation calls are viewed as one possible outcome of sequencing these samples it adds uncertainty in the regression that is currently unaccounted for.

Here we address three sources of bias in gene panel data: (a) regional subsampling of the genome in areas more commonly mutated in cancer (b) artifactual contribution of "private" germline variants to TMB in tumor-only sequencing (c) increased sensitivity of variant detection due to higher read depth in gene panel sequencing. In contrast to other approaches, in this study we leverage a deep learning regression framework that can model a heteroscedastic, asymmetric error distribution[28]. At very low TMB panel data initially underestimates TMB because it is bounded at 0, then it overestimates TMB because of the enrichment of genes commonly mutated in cancer, and then at higher values is a close approximation to exome TMB. This results in a nonlinear relationship between panel TMB and exomic TMB. Moreover, our regression framework models the error distribution in a manner that is dependent on the value of the input and as such can adapt—a lognormal error model may be applied at lower input values, and a more normally distributed model may be applied at larger values. To add in the uncertainty due to the low-depth sequencing we augment the data every batch, treating each sample as coming

from a distribution of possible values. To account for the error associated with removing germline variants we generate tumor-only input by applying population database filtering to TCGA germline and somatic data.

MATERIALS AND METHODS

Further details are available as supplemental material and code to generate the results of this publication is publicly available (https://doi.org/10.5281/zenodo.6478645).

TMB calculation

The TCGA exomic data was generated by different laboratories across multiple years with different technologies and protocols. This has the effect that there is variation in the quality of each sample, including depth of sequencing and exome coverage. Given that TMB is defined as mutations per Mb, the different exome coverages is potentially problematic. The majority of samples were processed with the Broad's exome kit, however even in these cases there's no guarantee that the coverage is identical between samples. As a result, instead of assuming a common size of each sample's captured exome we utilized the Broad's coverage WIGs to give each sample a unique exomic footprint, defining exomic TMB as the number of nonsynonymous mutations per Mb of covered coding sequence. Samples with a TMB (as defined above) larger than 40 were excluded. To better characterize the histopathology of the TCGA samples beyond the TCGA project codes (BRCA, PRAD, etc.) we mapped the samples to NCI thesaurus (NCIt) ontology, and only included samples which we could map and only used NCIt labels which had at least 50 mapped samples.

The TCGA MC3 reports all dinucleotide substitutions as two separate single base substitutions which may individually be reported as silent or nonsynonymous. The Pan-Cancer Analysis of Whole Genomes (PCAWG) consortium previously assumed all of these mutations are in phase[29]; we similarly treated adjacent mutations with similar variant allele counts as being in phase (supplemental methods). We assumed the new merged mutations were nonsynonymous mutations.

To generate estimates of the panel input and exomic TMB which would be observed at panel depth we used our previously described method[27]. To generate synthetic tumor-only panel data we filtered the TCGA PanCanAtlas germline calls along with the MC3 calls with the commonly used gnomAD population databases.

Data sources

TCGA somatic[30] and germline[31] data were used to model panel data utilizing coordinates from the Genomics, Evidence, Neoplasia, Information, Exchange (GENIE) project[32]. Hotspots were taken from Chang et al.[33]. gnomAD 2.1 exome and genome[34] were used as population databases.

Software

PyRanges[35] was used to perform intersections. BCFtools was used to perform VCF operations[36]. All the analyses are performed in Python 3 and the model was built with TensorFlow 2 and TensorFlow Probability.

RESULTS

Model fit

Regression models traditionally try to predict the y label such that the predictions and labels are as close as possible. Instead, we model the probability distribution of the label with the goal of maximizing the likelihood of the observed data given the predicted distribution. TensorFlow Probability provides numerous options for the assumed data distribution, including normal, lognormal, and binomial. The normal distribution is assumed when minimizing mean squared error (MSE), an assumption which is immediately broken when the y values are strictly positive. If the relationship between panel TMB and exomic TMB was strictly a genomic subsampling question we would expect the regression to follow a binomial distribution; however, the notion of a constant probability of success cannot be assumed with gene biases and different synonymous / nonsynonymous ratios. Like the binomial distribution the lognormal distribution is strictly positive and defined by 2 parameters, but doesn't make any assumptions about the probability of observing each individual mutation. We found lognormal to be a reasonable approximation of the observed label distribution across a range of input values.

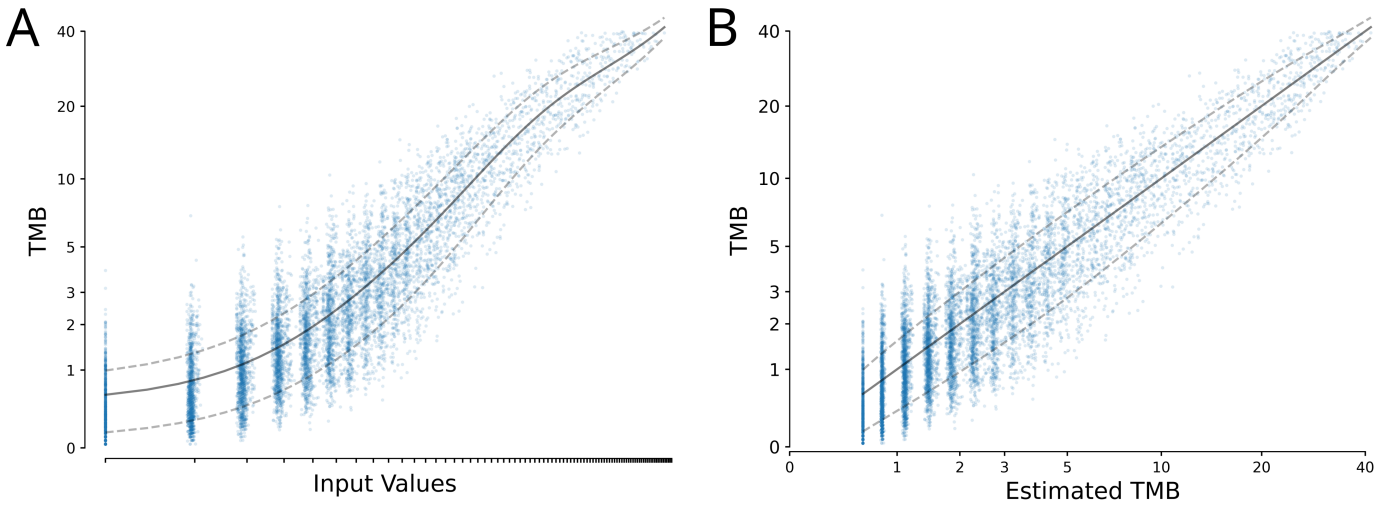

Figure 1. Model fit. (A) The fit of the model along with the estimated 68% confidence interval. Input and target data are also plotted. (B) Same plot as A but with the x axis passed through the model and the model fit replaced with the 45 degree line.

Figure 1A shows the exomic TMB values for TCGA MC3 data on the y axis with all mutations within the coding MSK-468 panel coordinates per Mb of covered-panel-CDS on the x axis. The model outputs the parameters of a lognormal distribution at a given input value which can then be used to calculate a point estimate of the target exomic TMB along with any desired confidence interval. Note, the input is all coding mutations normalized by covered panel size and as such is not a direct estimate of exomic nonsynonymous TMB.

The log-log transformed plots highlight the features of lower TMB values which otherwise would be compressed and indistinguishable. At each input value the model attempts to fit the distribution of the output, and can be seen to pass through the means of each distribution. After fitting the model, as can be seen in Figure 1B, the model output is now centered along the identity line (y=x) when plotted against the labels. Our assumption of the lognormal distribution can be appreciated by observing that the 68% confidence intervals estimated by the model contained 68% of the data.

Effect of panel inputs

A commonly employed method of estimating TMB is to normalize the number of nonsynonymous mutations observed in the panel by the panel size. In other protocols for calculating TMB, synonymous mutations may be included and common mutations may be excluded. Using our model we investigated these combinations of possible inputs across 5 different panels. While not currently employed by any protocols to our knowledge, given that each tumor type has a distinct distribution of TMB values and tumor type may also inform the synonymous/nonsynonymous ratio, we also investigated the use of tumor type as an input.

As seen in Table 1, simply applying our model to a nonsynonymous input results in a significant improvement in the regression metrics. Removing hotspots marginally improved the regression metrics while adding synonymous mutations significantly improved the regressions. NCIt labels provided an even further improvement upon the best combination of panel inputs. Using PCAWG data we also investigated adding noncoding mutations and saw a further improvement, suggesting that for panels with significant noncoding coverage the best data for modeling would include noncoding mutations from whole genome data restricted by the panel genomic coordinates (Supplemental Table 1).

Effect of read depth

Exomic TMB is viewed as the gold standard measure of TMB and reported values are currently treated as if they have no error associated with them. However, if the exome sequencing is performed at 50-100X as opposed to the 300-600X typically seen in panel sequencing, then a reported value should be interpreted as one possible outcome of sequencing the sample. We make the assumption that the hidden uncertainty in exomic TMB is due to missing mutations rather than falsely called mutations. As a result we take reported exomic values to be the lower bound of the true TMB.

Panel	Model	Synonymous	Hotspots	NCIt	RMSE	MAE	R-squared	PPV
DFCI-ONCOPANEL-3	Counting	-	+	-	2.94	2.00	0.74	0.68
	DeepTMB	-	+	-	2.33	1.44	0.83	0.83
	DeepTMB	-	-	-	2.29	1.41	0.84	0.83
	DeepTMB	+	+	-	2.03	1.27	0.87	0.85
	DeepTMB	+	-	-	2.00	1.25	0.88	0.85
	DeepTMB	+	-	+	1.66	0.97	0.92	0.85
MDA-409-V1	Counting	-	+	-	2.81	1.95	0.76	0.65
	DeepTMB	-	+	-	1.85	1.17	0.90	0.86
	DeepTMB	-	-	-	1.82	1.15	0.90	0.86
	DeepTMB	+	+	-	1.67	1.05	0.92	0.88
	DeepTMB	+	-	-	1.58	1.01	0.92	0.88
	DeepTMB	+	-	+	1.48	0.87	0.93	0.88
MSK-IMPACT341	Counting	-	+	-	3.80	2.59	0.56	0.57
	DeepTMB	-	+	-	2.39	1.50	0.83	0.82
	DeepTMB	-	-	-	2.36	1.46	0.83	0.82
	DeepTMB	+	+	-	2.10	1.33	0.87	0.84
	DeepTMB	+	-	-	2.07	1.30	0.87	0.84
	DeepTMB	+	-	+	1.74	1.02	0.91	0.84
MSK-IMPACT468	Counting	-	+	-	3.28	2.20	0.67	0.63
	DeepTMB	-	+	-	2.24	1.39	0.85	0.83
	DeepTMB	-	-	-	2.22	1.36	0.85	0.83
	DeepTMB	+	+	-	1.96	1.23	0.88	0.85
	DeepTMB	+	-	-	1.92	1.20	0.89	0.85
	DeepTMB	+	-	+	1.62	0.95	0.92	0.85
VICC-01-R2	Counting	-	+	-	2.78	1.87	0.76	0.70
	DeepTMB	-	+	-	2.17	1.35	0.86	0.84
	DeepTMB	-	-	-	2.15	1.34	0.86	0.84
	DeepTMB	+	+	-	1.89	1.20	0.89	0.86
	DeepTMB	+	-	-	1.86	1.18	0.89	0.86
	DeepTMB	+	-	+	1.63	0.96	0.92	0.86

Table 1. TMB regression metrics for TCGA MC3 samples. Regression metrics for different panels, inputs, and models. All metrics are from cross validation. RMSE: root mean squared error. MAE: mean absolute error. The positive predictive value was calculated at a TMB of 10.

We use a stochastic iterative process to estimate the number of mutations that would have been observed in a sample at higher depth. Each iteration in the process pulls from a read depth distribution for the sample and decides whether or not to call a mutation based on empirically determined probabilities that accounted for clonality and purity. This has the effect that samples with lower read depth, or lower purity, or lower clonality have higher probabilities of a mutation not being called each iteration. The process is repeated until the number of called mutations reaches the number of observed mutations, and the number of iterations is our estimate of the true TMB. We then get a distribution of the true TMB of a sample by repeating this process multiple times.

If the same linear transformation is applied equally across the inputs and outputs then the shape of a regression curve will not change, but given the different factors involved in calculating the estimates of true TMB it's possible that there's an interaction between the percent increase of the augmented TMB and the baseline TMB. For each sample we generated 100 estimates of the true TMB and plotted the average fold change per sample against the original TMB values (Figure 2A). The fold changes were relatively consistent across all TMB values, with the regression line having minimal slope and intersecting the y axis near 1.2, which corresponds to our overall average estimate of the percent of mutations observed in the MC3 dataset (83%).

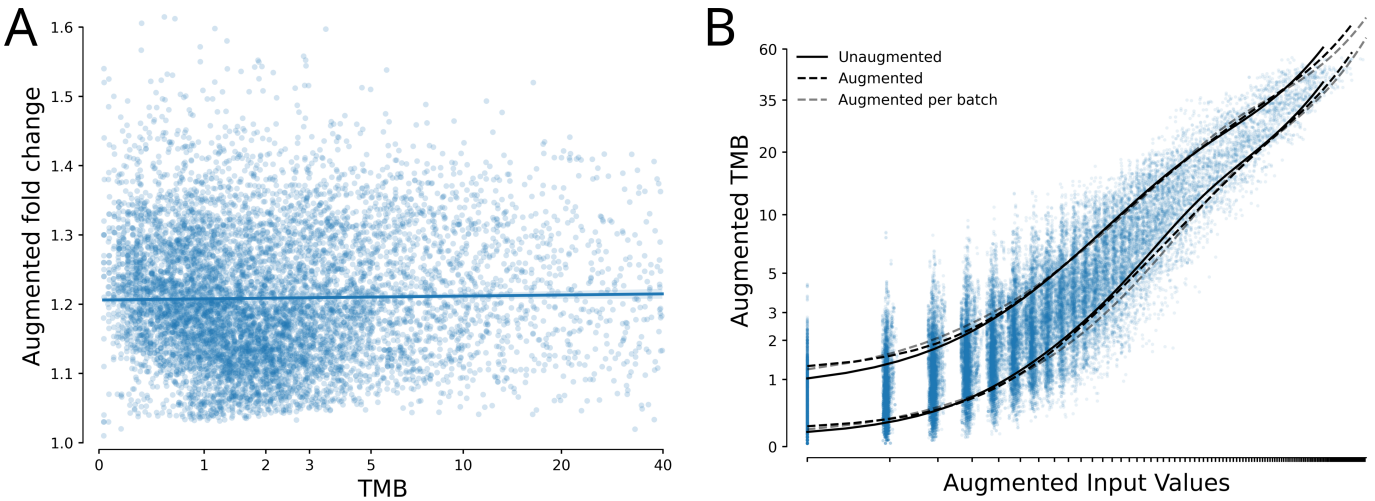

Figure 2. Read depth correction and model fit. (A) The point estimate of the augmented fold change of each sample plotted against TMB. (B) The 68% confidence intervals of three different model fits with 5 possible augmented data points plotted per sample.

Although useful for visualization, using the average values of estimates of the true TMB ignores the uncertainty of the values. Because we can control the exact data that the model sees every weight update, we are able to send the model one of the possible estimates of the input and label at a time. This will not change the average value of distributions predicted by the model since over many weight updates the average of what the model sees will be near the average of the estimates, but this will increase the spread of the data the model sees. We demonstrated this by recreating Figure 1A, but now plotting 5 augmented values per sample along with the confidence intervals for 3 different training regimens (Figure 2B). The distribution of the data is now wider due to both the augmentation and the randomness of the augmentations. The primary differences in the shapes of the regressions occurred at low TMB because we do not augment an input of 0, which results in larger y labels without a concurrent increase in the inputs.

Effect of tumor-only sequencing

While the issue of read depth had limited effects on the shape of the regression curves since the inputs and outputs were both increased, in the case of tumor-only sequencing additional variants are inadvertently included in the TMB calculation, which introduces additional variability. Of particular concern, when a model is trained on tumor-normal data (such as the TCGA MC3 data), and is applied to tumor-only panel inputs that include private variants, there is likely to be a significant effect on the outputs and regression metrics. To investigate this effect we constructed synthetic tumor-only inputs by reintroducing the germline variants into the TCGA somatic mutation profiles, and applying common filtering strategies across a spectrum of stringencies to the combined TCGA germline and somatic data in an attempt to mimic the filtering processes used with real-world tumor-only data.

We used gnomAD 2.1 non-cancer exome and gnomAD 2.1 genome as population databases, and considered the TCGA itself to be a population database of sorts. Using a self-cohort as a population database has the additional benefit of removing potential artifacts in the data. We applied a spectrum of stringency criteria for filtering out germline variants using the maximum population specific allele frequency (popmax_AF greater than 1%, 0.5%, and 0.1%) in the gnomAD 2.1 non-cancer exome and gnomAD 2.1 genome population databases. Each of these was paired with an overall allele frequency (AF greater than 0.1%, 0.05%, and 0.01%), in respective order. The popmax_AF thresholds were also applied to the TCGA "self-cohort" with known hotspots whitelisted[33]. To save on computation we only looked at variants within CDS GENIE coordinates, which came out to about a 4 Mb region.

Defining somatic as the positive class, our 3 different filtering cutoffs resulted in sensitivities / specificities of 99.9% / 98.2%, 99.6% / 98.6%, and 96.5% / 99.1% from least to most stringent. For samples used in our regression models we started with an average of 2,451 germline variants per sample in the 4Mb region represented by the union of all GENIE panels, and after filtering ended with 43, 34, and 20 germline variants per sample for each filtering stringency. Studies

examining gnomAD have estimated that on average a given genome has approximately 200 coding variants present at an AF less than 0.1% (private germline variants)[37]. Assuming a 32 Mb coding region, this would come out to about 25 variants in a 4 Mb region, which is consistent with our results.

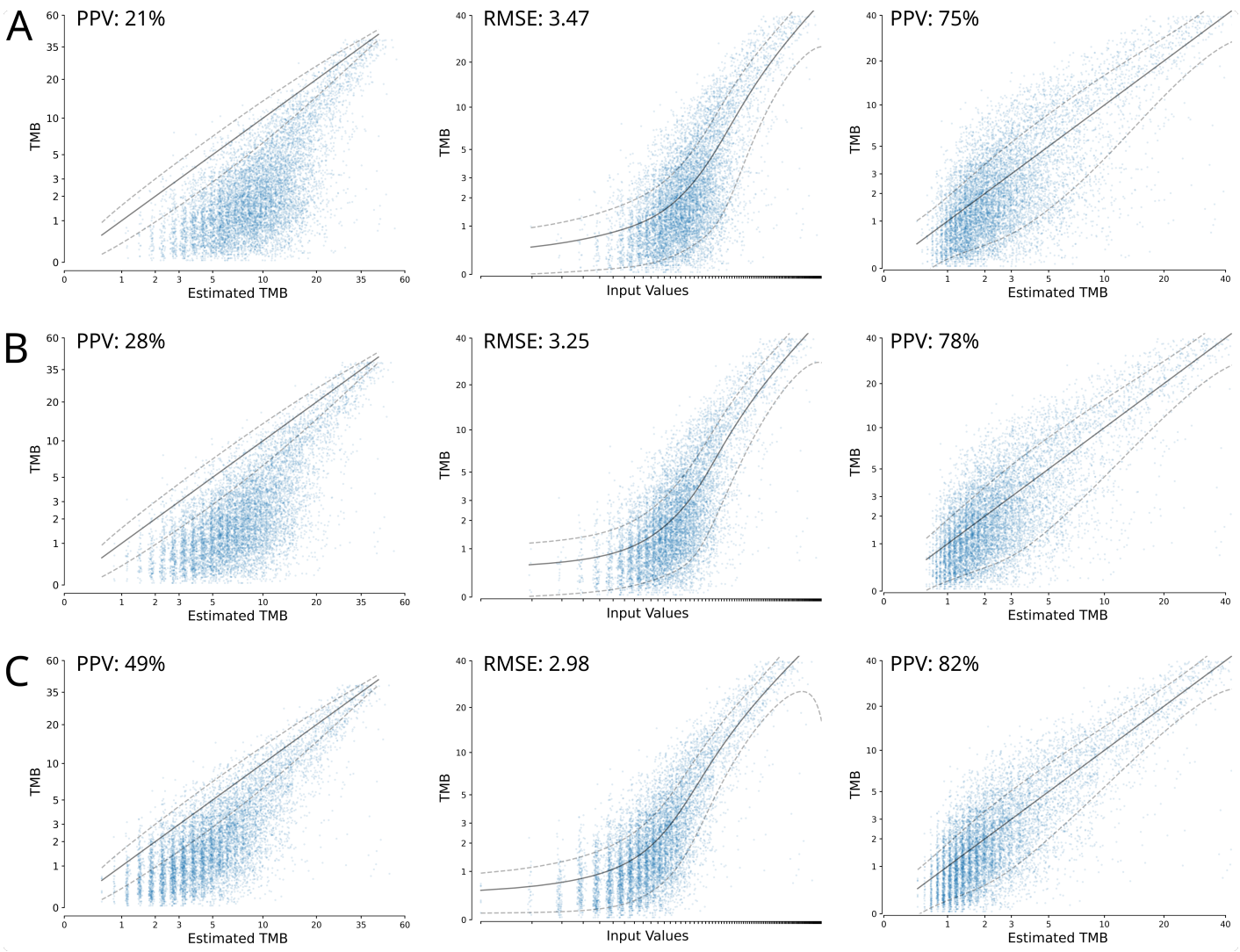

Figure 3. Effect of germline on model fits. The first column shows the effect of making predictions on data containing private germline variants with a model trained on tumor-normal data. The second column shows the model fits of our tumor-only data, and the third column shows the prediction vs. true plots of these fits. The first row (A), is the loosest filtering (1% popmax_AF, 0.1% overall AF). The second row (B) is moderate filtering (0.5% popmax_AF, 0.05% overall AF). The third row (C) is strict filtering (0.1% popmax_AF, 0.01% overall AF).

We trained a model as in Figure 1A but made predictions with the tumor-only data restricted to MSK-468 panel coordinates. As one might expect, the additional private germline variants in the input resulted in a significant bias / overestimation of the predicted exome TMB values. The effect was largest for the loosest filtering criteria (Figure 3), since these data would include the most number of additional private germline variants. Whereas the PPV was previously 85% for this fit, when applied to these inputs it drops to 21%, 28%, or 49% (across the spectrum of filtering criteria examined, respectively). We then naturally retrained our models with the synthetic TCGA tumor-only gene panel data as input to predict the exomic TMB derived from germline-subtracted tumor-normal data. The private germline variants have the effect of shifting samples which previously had low input values to the right and adding variability, which causes a wider range of possible TMB values for a given input. Due to this large spread we switched from the lognormal distribution to

the normal distribution for our model. Despite this additional noise in the regression, the PPVs were largely restored to the value seen with the tumor-normal data.

DISCUSSION

Given the nonlinear relationship between panel input and exomic TMB along with the complex error distributions, a more flexible model (such as deep learning approaches) is better suited for modeling the data. In this study we developed a deep learning probabilistic model that allows for joint modeling of both the point estimate of exomic TMB and the measurement uncertainty / distribution about that estimate, which stems from factors including but not limited to genomic regional sampling, degree of germline private variants, and depth of sequencing. Although it's true the most appropriate model should be used when possible, it is even more important that every effort be made to ensure the training data resembles the data the model will be applied to. In this study we underscored how significant this can be in the case of tumor-only data and characterized this by reintroducing private germline variants to the TCGA MC3 samples.

Previous studies have investigated the performance of different filtering approaches for tumor-only sequencing data, often with the goal of accurately filtering out germline variants[38–45]. Our goal was not to find the optimal filtering algorithm, but to simply use a reasonable filtering approach and to subsequently investigate its effect on TMB calibration. Our findings suggest that the optimal filtering strategy for TMB calibration would favor exclusion of additional germline variants even at the cost of losing some true somatic variants. This does not mean that those somatic variants are invalid, but simply for this purpose may have to be filtered in the context of estimating TMB from tumor-only data. It should be noted that the germline VAFs in our synthetic samples do not perfectly resemble what would be seen in true tumor-only data since they were derived from the matched-normal sample by the TCGA PanCanAtlas group, samples which do not contain the copy number alterations that may be present in the tumor samples. If a more faithful tumor-only dataset is required a tumor-only pipeline could potentially be run on the TCGA tumor samples, and then the variants could be labeled as somatic or germline by comparing to the germline data and the MC3 data.

At least for the TCGA dataset correcting for read depth did not have much effect on the regression or the uncertainty estimates. When using tumor-normal data for calibration of tumor-only panels there is an inherent assumption that no private germline variants will be present in samples the model will be applied to. However, current population databases are not large enough to perfectly filter germline variants, and a variable number of private variants will remain per sample even with very strict filtering. Either additional strategies would need to be applied to remove these rare variants (and not remove somatic variants), or the training data needs to include them. We opted for the second option and made the tumor-only input for every GENIE panel available (Data_S1) to allow any laboratory that uses one of these panels the ability to train with these counts.

ACKNOWLEDGMENTS

The results here are in whole or part based upon data generated by the TCGA Research Network.

FUNDING

This research was supported by the Mark Foundation for Cancer Research (19-035-ASP), and the philanthropy of Susan Wojcicki and Dennis Troper in support of Computational Pathology at Johns Hopkins.

REFERENCES

1. Wei, S. C., Duffy, C. R. & Allison, J. P. Fundamental Mechanisms of Immune Checkpoint Blockade Therapy. *Cancer Discovery* **8,** 1069–1086 (2018).
2. Michot, J. *et al.* Immune-related adverse events with immune checkpoint blockade: a comprehensive review. *European Journal of Cancer* **54,** 139–148 (2016).
3. Khagi, Y., Kurzrock, R. & Patel, S. P. Next generation predictive biomarkers for immune checkpoint inhibition. *Cancer and Metastasis Reviews* **36,** 179–190 (2016).

4. Yarchoan, M., Hopkins, A. & Jaffee, E. M. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *New England Journal of Medicine* **377,** 2500–2501 (2017).

5. Snyder, A. *et al.* Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *New England Journal of Medicine* **371,** 2189–2199 (2014).

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

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

8. Hellmann, M. D. *et al.* Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *New England Journal of Medicine* **378,** 2093–2104 (2018).

9. Hellmann, M. D. *et al.* Tumor Mutational Burden and Efficacy of Nivolumab Monotherapy and in Combination with Ipilimumab in Small-Cell Lung Cancer. *Cancer Cell* **33,** 853–861.e4 (2018).

10. Samstein, R. M. *et al.* Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nature Genetics* **51,** 202–206 (2019).

11. Chowell, D. *et al.* Improved prediction of immune checkpoint blockade efficacy across multiple cancer types. *Nature Biotechnology* (2021).

12. Allgäuer, M. *et al.* Implementing tumor mutational burden (TMB) analysis in routine diagnostics—a primer for molecular pathologists and clinicians. *Translational Lung Cancer Research* **7,** 703–715 (2018).

13. Büttner, R. *et al.* Implementing TMB measurement in clinical practice: considerations on assay requirements. *ESMO Open* **4,** e000442 (2019).

14. Chalmers, Z. R. *et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Medicine* **9,** 34 (2017).

15. Endris, V. *et al.* Measurement of tumor mutational burden (TMB) in routine molecular diagnostics: in silico and real-life analysis of three larger gene panels. *International Journal of Cancer* **144,** 2303–2312 (2019).

16. Marabelle, A. *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. *The Lancet Oncology* **21,** 1353–1365 (2020).

17. FDA. *FDA approves pembrolizumab for adults and children with TMB-H solid tumors* 2020. https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-pembrolizumab-adults-and-children-tmb-h-solid-tumors.

18. Prasad, V. & Addeo, A. The FDA approval of pembrolizumab for patients with TMB >10 mut/Mb: was it a wise decision? No. *Annals of Oncology* **31,** 1112–1114 (2020).

19. Stenzinger, A. *et al.* Tumor mutational burden (TMB) standardization initiative: Establishing a consistent methodology for TMB measurement in clinical samples. *Annals of Oncology* **29,** viii45 (2018).

20. Buchhalter, I. *et al.* Size matters: Dissecting key parameters for panel-based tumor mutational burden analysis. *International Journal of Cancer* **144,** 848–858 (2019).

21. Budczies, J. *et al.* Optimizing panel-based tumor mutational burden (TMB) measurement. *Annals of Oncology* **30,** 1496–1506 (2019).

22. Merino, D. M. *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. *Journal for Immunotherapy of Cancer* **8.** ISSN: 2051-1426 (2020).

23. Sun, J. X. *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 Computational Biology* **14** (ed Dunbrack, R. L.) e1005965 (2018).

24. Spencer, D. H. *et al.* Performance of common analysis methods for detecting low-frequency single nucleotide variants in targeted next-generation sequence data. *The Journal of Molecular Diagnostics* **16,** 75–88 (2014).

25. Alioto, T. S. *et al.* A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. *Nature communications* **6,** 1–13 (2015).

26. Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nature Medicine* **23,** 703–713 (2017).

27. Anaya, J. & Baras, A. S. Read depth correction for somatic mutations. *bioRxiv*. https://doi.org/10.1101/2022.02.16.480761 (2022).

28. Barnes, E. A., Barnes, R. J. & Gordillo, N. *Adding Uncertainty to Neural Network Regression Tasks in the Geosciences* 2021. arXiv: 2109.07250 [physics.ao-ph].

29. Bailey, M. H. *et al.* Retrospective evaluation of whole exome and genome mutation calls in 746 cancer samples. *Nature Communications* **11** (2020).

30. Ellrott, K. *et al.* Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Systems* **6,** 271–281 (2018).

31. Huang, K.-l. *et al.* Pathogenic germline variants in 10,389 adult cancers. *Cell* **173,** 355–370 (2018).

32. AACR Project GENIE Consortium and others. AACR Project GENIE: powering precision medicine through an international consortium. *Cancer Discovery* **7,** 818–831 (2017).

33. Chang, M. T. *et al.* Accelerating discovery of functional mutant alleles in cancer. *Cancer discovery* **8,** 174–183 (2018).

34. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581,** 434–443 (2020).

35. Stovner, E. B. & Sætrom, P. PyRanges: efficient comparison of genomic intervals in Python. *Bioinformatics* **36,** 918–919 (2020).

36. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10** (2021).

37. Gudmundsson, S. *et al.* Variant interpretation using population databases: Lessons from gnomAD. *Human Mutation* (2021).

38. Jones, S. *et al.* Personalized genomic analyses for cancer mutation discovery and interpretation. *Science Translational Medicine* **7** (2015).

39. Garofalo, A. *et al.* The impact of tumor profiling approaches and genomic data strategies for cancer precision medicine. *Genome Medicine* **8** (2016).

40. Chang, H. *et al.* Bioinformatic Methods and Bridging of Assay Results for Reliable Tumor Mutational Burden Assessment in Non-Small-Cell Lung Cancer. *Molecular Diagnosis & Therapy* **23,** 507–520 (2019).

41. Sukhai, M. A. *et al.* Somatic Tumor Variant Filtration Strategies to Optimize Tumor-Only Molecular Profiling Using Targeted Next-Generation Sequencing Panels. *The Journal of Molecular Diagnostics* **21,** 261–273 (2019).

42. Parikh, K. *et al.* Tumor Mutational Burden From Tumor-Only Sequencing Compared With Germline Subtraction From Paired Tumor and Normal Specimens. *JAMA Network Open* **3,** e200202 (2020).

43. Stenzinger, A. *et al.* Harmonization and Standardization of Panel-Based Tumor Mutational Burden Measurement: Real-World Results and Recommendations of the Quality in Pathology Study. *Journal of Thoracic Oncology* **15,** 1177–1189 (2020).

44. Asmann, Y. W. *et al.* Inflation of tumor mutation burden by tumor-only sequencing in under-represented groups. *NPJ precision oncology* **5,** 1–4 (2021).

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