

## ARTICLE OPEN



# Optimizing cancer immunotherapy response prediction by tumor aneuploidy score and fraction of copy number alterations

Tian-Gen Chang<sup>1</sup>✉, Yingying Cao<sup>1</sup>, Eldad D. Shulman<sup>1</sup>, Uri Ben-David<sup>2</sup>, Alejandro A. Schäffer<sup>1</sup> and Eytan Ruppin<sup>1</sup>✉

Identifying patients that are likely to respond to cancer immunotherapy is an important, yet highly challenging clinical need. Using 3139 patients across 17 different cancer types, we comprehensively studied the ability of two common copy-number alteration (CNA) scores—the tumor aneuploidy score (AS) and the fraction of genome single nucleotide polymorphism encompassed by copy-number alterations (FGA)—to predict survival following immunotherapy in both pan-cancer and individual cancer types. First, we show that choice of cutoff during CNA calling significantly influences the predictive power of AS and FGA for patient survival following immunotherapy. Remarkably, by using proper cutoff during CNA calling, AS and FGA can predict pan-cancer survival following immunotherapy for both high-TMB and low-TMB patients. However, at the individual cancer level, our data suggest that the use of AS and FGA for predicting immunotherapy response is currently limited to only a few cancer types. Therefore, larger sample sizes are needed to evaluate the clinical utility of these measures for patient stratification in other cancer types. Finally, we propose a simple, non-parameterized, elbow-point-based method to help determine the cutoff used for calling CNAs.

npj Precision Oncology (2023)7:54; <https://doi.org/10.1038/s41698-023-00408-6>

## INTRODUCTION

Although various studies have shown that high tumor mutation burden (TMB) may predict immunotherapy response, at least in some cancer types<sup>1,2</sup>, more precise identification of patients that are likely to respond to cancer immunotherapy is still a challenging unmet clinical need. One promising approach to identify responders of immunotherapy has been to study the predictive ability of other measures of genomic alterations in cancer in these patients. Two natural candidates are scores based on copy-number alterations (CNAs): (a) tumor aneuploidy, which measures chromosome-level CNAs, and (b) global genomic CNAs, which quantifies the extents of both chromosomal and focal copy-number events<sup>3</sup>. Both tumor aneuploidy and genomic CNAs have been shown to play a role in cancer progression and to be predictive for cancer prognosis<sup>3–5</sup>.

Recently, Spurr et al. reported that the tumor *aneuploidy score* (AS), defined as the fraction of chromosome arms with arm-level CNAs in a sample, which was called using a loose cutoff of  $|\log_2 \text{copy ratio}| > 0.1$ , is significantly predictive of survival following immunotherapy in low-TMB patients, but not in high-TMB patients, in a pan-cancer analysis<sup>6</sup>. In addition, they reported that AS had stronger predictive power than another metric conceptually related to the AS, the *fraction of genome encompassed by copy-number alterations* (FGA) which quantifies the extent of both chromosomal and focal copy-number events<sup>6</sup>. As FGA combines both chromosomal and focal CNAs, if the association between CNAs and immunotherapy response is driven by the overall genomic instability, one would expect FGA to perform at least as well as AS in predicting immunotherapy response. Therefore, the conclusion in ref. <sup>6</sup> that AS is a better predictor than FGA in low-TMB patients is non-intuitive. Intrigued by these potentially clinically impactful findings, we set out to explore several related fundamental questions: (1) Does the choice of cutoff during CNA

calling influence the predictive power? (2) Are AS and FGA also predictive of survival for high-TMB patients? (3) Are AS and FGA predictive of survival of patients following immunotherapy in individual cancer types?

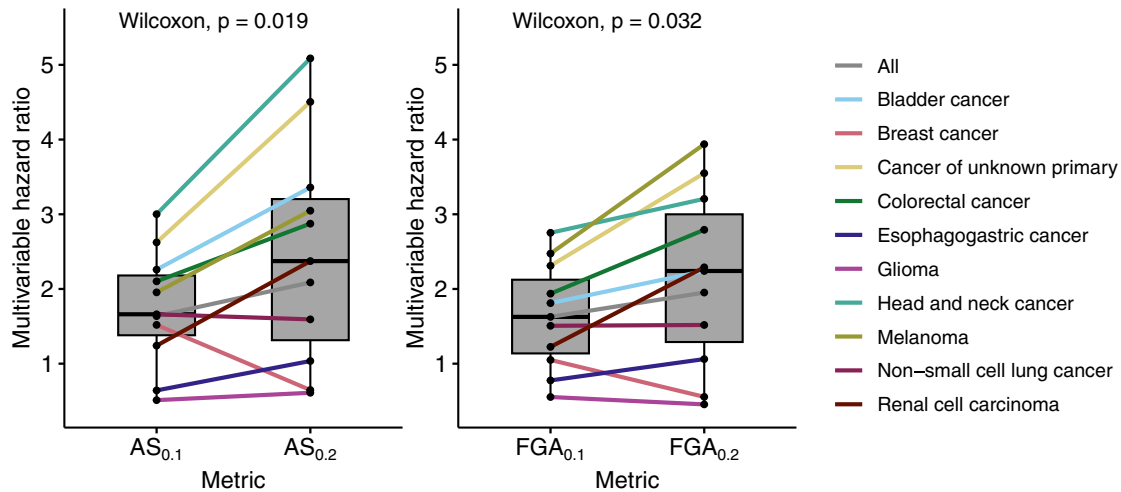
## RESULTS

### The choice of cutoff during CNA calling markedly influences the predictive power of AS and FGA for patient survival following immunotherapy

We first re-analyzed the same data used in ref. <sup>6</sup>, i.e., the Samstein et al.'s cohort<sup>1</sup> from MSK-IMPACT. This study analyzed a published cohort of 1660 advanced cancer patients from ten different cancer types treated with immune checkpoint blockade (ICB). Their results show that, at the pan-cancer level, a higher AS was associated with worse survival following immunotherapy among patients with low TMB (defined as the bottom 80% of TMB in each cancer type). However, their study did not explicitly identify the individual cancer types in which AS is predictive. As a pan-cancer Kaplan–Meier survival analysis (as performed in ref. <sup>6</sup>) may be confounded by the cancer-type composition of the overall dataset, and as most clinical trials usually focus on individual cancer types, we first set out to compare the Kaplan–Meier survival curves of low-TMB patients with high versus low AS for each of the ten cancer types individually.

The initial cancer-type-specific analysis was performed by using the AS values provided in ref. <sup>6</sup> (which calls chromosome-level CNAs using the cutoff of  $|\log_2 \text{copy ratio}| > 0.1$ ; denoted as  $AS_{0.1}$ ). Unexpectedly, a Kaplan–Meier survival analysis of low-TMB patients identified a statistically significantly worse survival following immunotherapy in a single individual cancer type, i.e., *cancer of unknown primary*, which refers to a group of cancers with unknown origin, often due to metastasis making it difficult to

<sup>1</sup>Cancer Data Science Laboratory, Center for Cancer Research, National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD 20892, USA. <sup>2</sup>Department of Human Molecular Genetics and Biochemistry, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. ✉email: [tiangen.chang@nih.gov](mailto:tiangen.chang@nih.gov); [eytan.rupp@nih.gov](mailto:eytan.rupp@nih.gov)



**Fig. 1 The choice of cutoff during CNA calling markedly influences the predictive power of AS and FGA for patient survival following immunotherapy.** Comparison of HRs using  $AS_{0.1}$  or  $AS_{0.2}$  or  $FGA_{0.1}$  or  $FGA_{0.2}$  in a multivariate Cox model with TMB and ICB drug class. Paired Wilcoxon test  $P$  values are displayed. In the plot, the upper and lower boundaries signify the first and third quartiles, correspondingly, the central line denotes the median, and the whiskers stretch to the most distant data points not classified as outliers (within 1.5 times the interquartile range). The data are from the Samstein et al.'s cohort<sup>1</sup>.

locate the primary site ( $n = 70$ , hazard ratio  $HR = 2.27$ ,  $P = 0.031$ ; Supplementary Fig. 1). Here, the HR denotes the relative risk of the  $AS_{0.1}$ -high individuals compared to the  $AS_{0.1}$ -low set as the reference.

Aiming to improve on these results, we observed that while the cutoff used to determine a CNA event in ref. <sup>6</sup> was  $|\log_2 \text{copy ratio}| > 0.1$ , the cutoff of  $|\log_2 \text{copy ratio}| > 0.2$  in calculating AS and/or FGA was more frequently used (e.g., refs. <sup>7–9</sup>). Our first hypothesis tested whether the choice of cutoff during CNA calling affects the predictive power of AS and FGA for survival following immunotherapy. To this end, we re-calculated AS and FGA for each sample using the CNA calling cutoff of  $|\log_2 \text{copy ratio}| > 0.2$ . We then compared the HRs of AS and FGA in individual cancer types using AS and FGA, respectively, in a multivariable Cox proportional hazards regression of overall survival with TMB and ICB drug class, as had been done in ref. <sup>6</sup>. Remarkably, HRs were significantly increased for both AS ( $P = 0.019$ ) and FGA ( $P = 0.032$ ) by using the CNA calling cutoff of  $|\log_2 \text{copy ratio}| > 0.2$  (denoted by  $AS_{0.2}$  and  $FGA_{0.2}$ , respectively) compared to that calculated by using a cutoff of 0.1 in ref. <sup>6</sup> (denoted by  $AS_{0.1}$  and  $FGA_{0.1}$  respectively; Fig. 1).

### **$AS_{0.2}$ and $FGA_{0.2}$ predict pan-cancer survival following immunotherapy for both high-TMB and low-TMB patients**

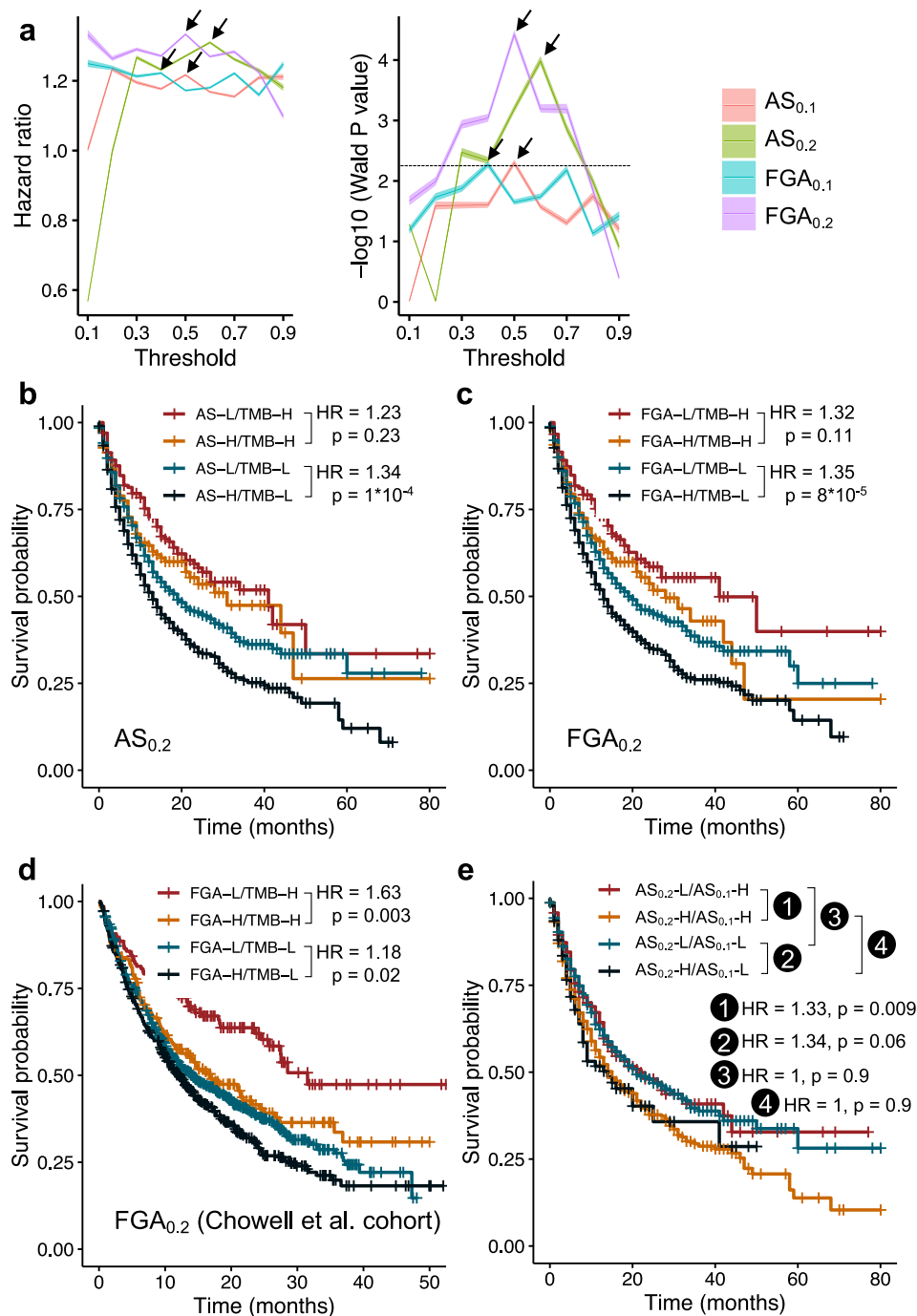
Although  $AS_{0.2}$  and  $FGA_{0.2}$  are continuous variables, a binary score (based on high or low scores) is preferred in clinical decision-making. Following ref. <sup>6</sup>, we determined the percentile to partition the  $AS_{0.2}$  scores into low and high so that they are optimally synergized with TMB to risk-stratify patients following immunotherapy by testing every tenth quantile within each cancer type, moving in increments from the 20th to 80th percentile, using a multivariate model with TMB (binned at 80th percentile) and ICB drug class. We identified the 60th percentile in each cancer type as the optimal binarization threshold to classify patients into high  $AS_{0.2}$  and low  $AS_{0.2}$  groups because it yielded highest multivariate HR with significant Bonferroni-corrected  $P$  value (Fig. 2a). Similarly, the optimal percentiles to binarize the  $AS_{0.1}$ ,  $FGA_{0.1}$ , and  $FGA_{0.2}$  scores are 50th, 40th, and 50th, respectively (Fig. 2a).

Then, we tested our second hypothesis, examining whether  $AS_{0.2}$  and  $FGA_{0.2}$  can predict survival outcomes for both high-TMB patients and low-TMB patients. Strikingly, both  $AS_{0.2}$  and  $FGA_{0.2}$  had similar effect size in predicting survival in high-TMB patients

compared with that in low-TMB patients, respectively. Specifically, the HRs between high and low  $AS_{0.2}$  groups were 1.23 and 1.34, respectively, among high-TMB versus low-TMB patients (Fig. 2b); Similarly, the HRs between high and low  $FGA_{0.2}$  groups were 1.32 and 1.35, respectively, among high-TMB versus low-TMB patients (Fig. 2c). Overall, by using TMB and  $AS_{0.2}$  (or  $FGA_{0.2}$ ) together, we can classify patient survival following immunotherapy into four groups: high TMB & low  $AS_{0.2}$  (or  $FGA_{0.2}$ ) > high TMB & high  $AS_{0.2}$  (or  $FGA_{0.2}$ ) > low TMB & low  $AS_{0.2}$  (or  $FGA_{0.2}$ ) > low TMB & high  $AS_{0.2}$  (or  $FGA_{0.2}$ ). In addition,  $FGA_{0.2}$  was found to have consistently slightly higher HRs and lower  $P$  values than  $AS_{0.2}$  (Fig. 2a–c), which suggests that  $FGA_{0.2}$  is better or performs at least as well as  $AS_{0.2}$  in predicting pan-cancer ICB response.

To further test this finding in other datasets, we analyzed another MSK-IMPACT cohort published recently by Chowell et al.<sup>8</sup>. In the Chowell et al.'s cohort, there are in total 15 cancer types, 8 of them are in common with the above-used Samstein et al.'s cohort (we merged gastric and esophageal cancers in the Chowell et al.'s cohort into esophagogastric cancer to keep in line with the tumor type classification in the Samstein et al.'s cohort). We note that we could not use the Chowell et al. data to validate the AS analysis because these data do not include AS values and it is not possible to calculate the AS values based on the publicly available information. Consistently,  $FGA_{0.2}$  was found to predict survival following immunotherapy for both high-TMB and low-TMB patients. Specifically, the HRs between high and low  $FGA_{0.2}$  groups were 1.63 and 1.18, respectively, among high-TMB versus low-TMB patients (Fig. 2d).

We hypothesized that CNA calling cutoff  $|\log_2 \text{copy ratio}| > 0.1$  is a too low cutoff, which introduced noise in calculating patient AS, and thus dampened its predictive power of survival following immunotherapy. To test this hypothesis, we divided patients in the Samstein et al.'s cohort into four groups by their high/low  $AS_{0.1}/AS_{0.2}$  scores and compared the Kaplan–Meier survival curves (Fig. 2e). We found that, among high  $AS_{0.2}$  or among low  $AS_{0.2}$  patients, there was no significant survival difference between patients that had high or low  $AS_{0.1}$  values. In contrast, among high  $AS_{0.1}$  patients, a subset of patients, i.e., the low  $AS_{0.2}$  patients, had much better survival rates than high  $AS_{0.2}$  patients ( $HR = 1/1.33 = 0.75$ ,  $P$  value = 0.009); they actually achieved similar survival rates as the low  $AS_{0.1}/low AS_{0.2}$  patients ( $HR = 1$ ,  $P$  value = 0.9). On the other hand, among low  $AS_{0.1}$  patients, a subset of patients, i.e., the high  $AS_{0.2}$  patients, had significantly



**Fig. 2** AS<sub>0.2</sub> and FGA<sub>0.2</sub> predict pan-cancer survival following immunotherapy for both high-TMB and low-TMB patients. **a** The x axis shows candidate binarization proportions 0.1 through 0.9 corresponding to 10th through 90th percentiles to partition patient scores into high score versus low score at each percentile. In total, 1660 multivariate Cox models as part of the leave-one-out cross-validation analysis are constructed with AS<sub>0.1</sub> or AS<sub>0.2</sub> or FGA<sub>0.1</sub> or FGA<sub>0.2</sub> (binned at the candidate binarization percentile), TMB (binned at the 80th percentile), and ICB drug class. The Wald *P* values and multivariate HRs with 95% confidence intervals are displayed, respectively. Black arrows indicate Wald *P* values and multivariable HRs at the optimal percentiles, respectively. Dashed line denotes the Bonferroni-corrected *P* = 0.05. **b** Pan-cancer Kaplan–Meier analysis of AS<sub>0.2</sub> binned at the 60th percentile and TMB binned at the 80th percentile in the Samstein et al.’s cohort. **c, d** Pan-cancer Kaplan–Meier analysis of FGA<sub>0.2</sub> binned at the 50th percentile and TMB binned at the 80th percentile in the Samstein et al.’s cohort (**c**) and in the Chowell et al.’s cohort (**d**). **e** Pan-cancer Kaplan–Meier analysis of AS<sub>0.2</sub> binned at the 60th percentile and AS<sub>0.1</sub> binned at the 50th percentile in the Samstein et al.’s cohort. HR and *P* values of pairwise comparisons between different groups are shown. H high, L low.

worse survival rates than low AS<sub>0.2</sub> patients (HR = 1.34, *P* value = 0.06); they actually had similar survival rates as the high AS<sub>0.1</sub>/high AS<sub>0.2</sub> patients (HR = 1, *P* value = 0.9). This result testifies that the AS<sub>0.1</sub> indeed mis-classifies a number of patients as a result of the loose CNA calling cutoff used. Further investigation into the

patients that were misclassified by AS<sub>0.1</sub> showed that the “low AS<sub>0.2</sub>, high AS<sub>0.1</sub>” patients had significantly lower tumor purity than the “high AS<sub>0.2</sub>, high AS<sub>0.1</sub>” patients; and similarly, the “high AS<sub>0.2</sub>, low AS<sub>0.1</sub>” patients had significantly higher tumor purity than the “low AS<sub>0.2</sub>, low AS<sub>0.1</sub>” patients (Supplementary Fig. 2).

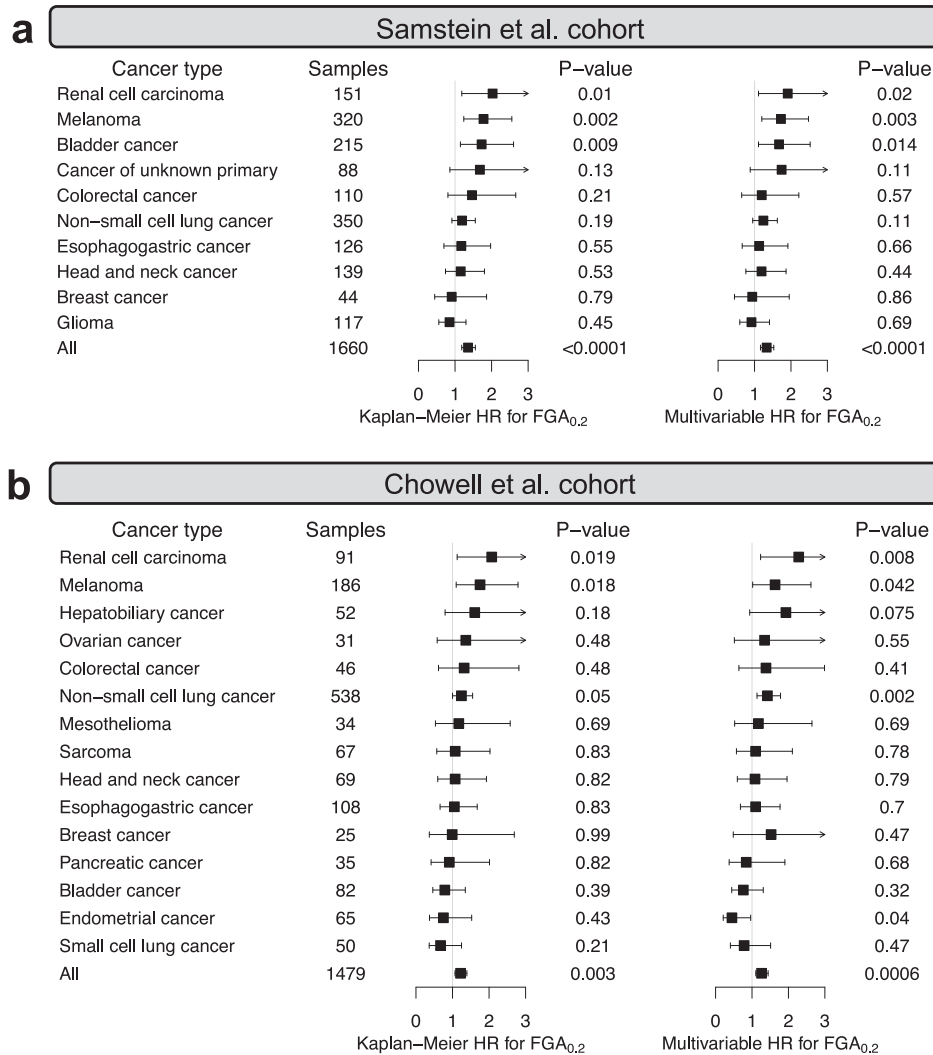
These findings suggest that tumor purity may, at least partially, explain the switch of some samples from high/low  $AS_{0.1}$  to low/high  $AS_{0.2}$ . However, further studies are needed to fully understand the relationship between AS and tumor purity and to determine the optimal cutoff for AS in predicting patient response to immunotherapy when such data are available.

### $AS_{0.2}$ FGA $_{0.2}$ predict survival following immunotherapy in certain individual cancers

Having demonstrated that  $AS_{0.2}$  and FGA $_{0.2}$  predict survival following immunotherapy for both high-TMB and low-TMB patients at the pan-cancer level, we next asked whether these scores could also predict survival in *individual* cancer types. As a result, in the Samstein et al.'s cohort, FGA $_{0.2}$  had significant HRs for pan-cancer (HR = 1.36,  $P < 0.0001$ ) and in three individual cancer types in Kaplan–Meier survival analysis, i.e., renal cell carcinoma (HR = 2.03,  $P = 0.01$ ), melanoma (HR = 1.78,  $P = 0.002$ ), and bladder cancer (HR = 1.73,  $P = 0.009$ ; Fig. 3a). In comparison,  $AS_{0.2}$  yielded significant Kaplan–Meier univariable HRs in bladder cancer and renal cell carcinoma, and marginally significant

multivariate HRs in melanoma (Supplementary Fig. 3). Comparison of HRs using  $AS_{0.2}$  or FGA $_{0.2}$  in a multivariable Cox model with TMB (binned at the 80th percentile) and ICB drug class yielded very similar result (Fig. 3a and Supplementary Fig. 3). Overall, we conclude that FGA performs comparable to or better than AS in predicting immunotherapy response in individual cancers, suggesting that it is the overall genome affected by CNAs (rather than the individual CNA length or mechanism of formation) that drives the observed CNA-immunotherapy response associations.

We further tested the robustness of FGA $_{0.2}$  in predicting survival following immunotherapy in the other dataset, i.e., the Chowell et al.'s cohort. As a result, FGA $_{0.2}$  had significant Kaplan–Meier univariable HRs for pan-cancer (HR = 1.22,  $P = 0.003$ ) and in renal cell carcinoma (HR = 2.07,  $P = 0.019$ ) and melanoma (HR = 1.75,  $P = 0.018$ ). Again, multivariable Cox model with adjustment of TMB (binned at the 80th percentile) and ICB drug class yielded similar result (Fig. 3b). However, FGA $_{0.2}$  did not predict worse survival for high FGA $_{0.2}$  in bladder cancer in this cohort as what in the Samstein et al.'s cohort (HR = 0.79,  $P = 0.39$ ; Fig. 3b), which might be due to the small sample size in the Chowell et al.'s cohort ( $n = 82$ ; Fig. 3b), and/or, due to tumor heterogeneity. For



**Fig. 3** FGA $_{0.2}$  predicts survival following immunotherapy in certain individual cancers. Univariable Kaplan–Meier survival analysis and multivariable survival analysis using Cox proportional hazards regression of overall survival with FGA $_{0.2}$  (binned at the 50th percentile), TMB (binned at the 80th percentile), and ICB drug class in the Samstein et al.'s cohort (**a**) and in the Chowell et al.'s cohort (**b**). In the plot, squares positioned at midpoints symbolize point estimates of HRs, and the accompanying bars indicate 95% confidence intervals. Wald  $P$  values are displayed.

example, further comparison analysis showed that bladder cancer samples in the Samstein et al.'s cohort had slightly lower mean  $FGA_{0.2}$  value (0.19 versus 0.23,  $P=0.14$ ) and better survival ( $HR=0.75$ ,  $P=0.098$ ) than that in the Chowell et al.'s cohort (Supplementary Fig. 4).

Interestingly,  $FGA_{0.2}$  predicted significant HRs for one more cancer type, non-small cell lung cancer, although with modest HR values ( $HR=1.25$ ,  $P=0.05$ ; Fig. 3b). Given that the HR values in the Samstein et al.'s cohort show a similar trend but are non-significant ( $HR=1.19$ ,  $P=0.19$ ; Fig. 3a), we wondered whether the difference in statistical significance might be due to the difference of sample size. Therefore, we performed a power analysis to estimate the sample size needed for achieving statistical significance  $P$  value less than 0.05. The estimated sample size for non-small cell lung cancer in the Samstein et al.'s cohort to achieve  $P<0.05$  is about 1600 (Supplementary Table 1). Similarly, it was found that colorectal cancer might also achieve significant  $HR>1$  with a sample size of ~600 patients in both cohorts (Supplementary Table 1). Our data analysis revealed that significant Kaplan–Meier survival analysis always corresponded to significant multivariate analysis in individual cancer types and vice versa. Therefore, it is unlikely that the limited efficacy of AS/FGA in certain cancer types is due to the choice of univariate or multivariate analysis. Instead, sample size limitation may be a critical factor for specific cancer types, as suggested by our power analysis. In contrast, in some cancer types, extremely large estimated sample size is needed for achieving statistical significance, e.g., esophagogastric cancer (Supplementary Table 1). It is more plausible that AS/FGA may not work in those cancer types due to specific underlying biological factors.

In addition, to investigate whether mutation of specific genes may contribute to the survival difference following immunotherapy, we performed a differential gene mutation frequency analysis among high  $FGA_{0.2}$  (or  $AS_{0.2}$ ) group versus low  $FGA_{0.2}$  (or  $AS_{0.2}$ ) group in the two cancer types in the Samstein et al.'s cohort with the largest sample size, i.e., melanoma and non-small cell lung cancer. We found that none of the genes had significantly differential mutation frequencies between the high  $FGA_{0.2}$  (or  $AS_{0.2}$ ) versus low  $FGA_{0.2}$  (or  $AS_{0.2}$ ) patients after multiple testing correction (Supplementary Table 2).

### The elbow-point-based method offers one systematic way to determine the cutoff used for calling CNAs

Finally, as shown above, the cutoff used for calling CNAs is critical for calculating  $AS_{0.2}$  and  $FGA_{0.2}$ . A low cutoff of  $|\log_2$  copy ratio| in calling CNA events might introduce noise (false positives), whereas a high cutoff might result in missing true events (false negatives). There are a number of parameters that may affect the optimal cutoff, e.g., cancer type, tumor purity, and the platform used for CNA calling (e.g., whole-exome sequencing, single nucleotide polymorphism arrays, and shallow whole genome sequencing)<sup>10–14</sup>. The variance of these parameters in different cancer types is likely to explain why AS and FGA scores have very different predictive power in distinct cancer types. We hence reasoned that an arbitrary threshold could never be optimal for all datasets and searched for an unbiased approach for threshold calling. We used the *elbow method*, which was developed to identify a cutoff point that optimally distinguishes between two qualitative, discrete states<sup>15</sup>. This method has been found to be effective in determining optimal parameter thresholds in a variety of data-driven optimization tasks including the determination of the number of clusters, determination of the number of principal components, and with relevance to our goal, determination of the threshold on a receiver operating characteristic curve<sup>16–18</sup>.

We calculated the elbow points of CNA calling cutoff  $|\log_2$  copy ratio| for AS for all ten individual cancer types (exemplified as in Fig. 4a), which are in the range of 0.14–0.22 with 95% confidence interval in the range of 0.12–0.27 (Fig. 4b). Therefore, the cutoff of

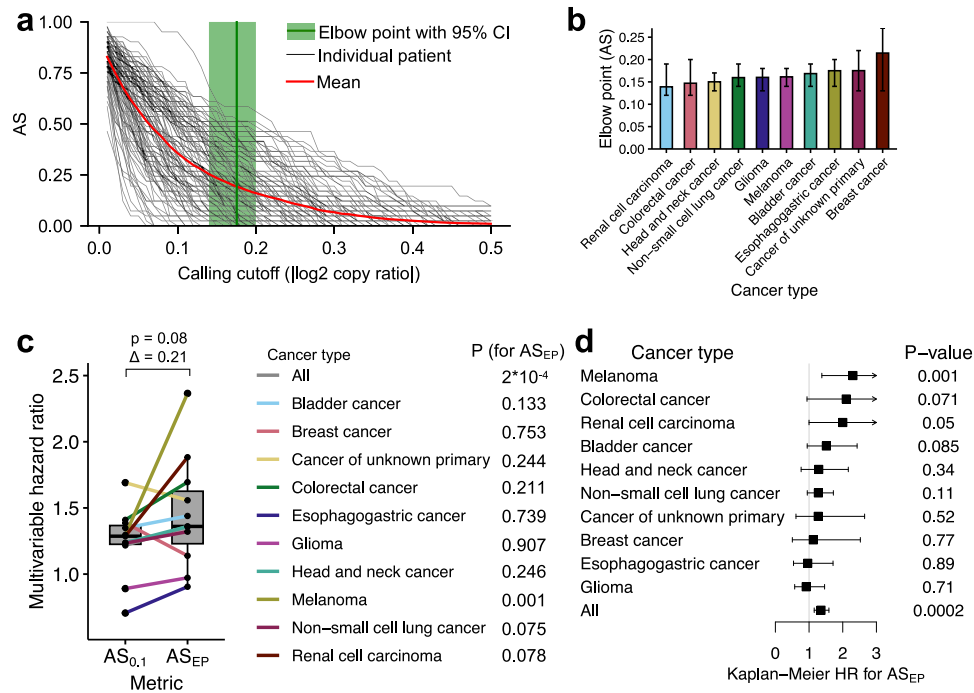
0.1 used in ref. <sup>6</sup> is well-below the elbow points for all individual cancer types. However, on the other hand, the average values of elbow points across different cancer types of both AS and FGA are 0.17, which is very close to the cutoff of 0.2 used above. These facts may explain why the 0.2 cutoff performs much better than the 0.1 cutoff. We further re-evaluated the predictive power of AS by calculating AS using the elbow points as the CNA calling cutoff per cancer types (denoted as  $AS_{EP}$ ). We identified the 30th percentile as the optimal binarization threshold to classify patients into high  $AS_{EP}$  and low  $AS_{EP}$  groups (Supplementary Fig. 5). The multivariable HRs of binarized  $AS_{EP}$  (with adjustment of TMB and ICB drug class) in individual cancer types were, on average, greater than those obtained using  $AS_{0.1}$  ( $\Delta$  mean  $HR=0.21$ ,  $P=0.08$ ; Fig. 4c). Furthermore,  $AS_{EP}$  predicted significant HR in melanoma and marginally significant HRs in two other cancer types, i.e., non-small cell lung cancer and renal cell carcinoma, tested by both Kaplan–Meier univariable survival analysis and multivariable Cox model with adjustment for TMB and ICB drug (Fig. 4c, d). The elbow-point-based method to determine the cutoff used for calling CNAs yielded similar result in FGA (Supplementary Figs. 5 and 6). To test if differential tumor purity across different cancer types may contribute to the variation of elbow points in individual cancer types, we investigated the relationship between elbow points and average tumor purity. A weak negative but statistically non-significant correlation was found (Supplementary Fig. 7).

In addition, we also tested using another method to determine the cutoff, testing a Gaussian mixture model. However, the Gaussian mixture model gave unrealistically high cutoff values ranging from 0.35 to 0.39 for individual cancer types (Supplementary Fig. 8a), which resulted in  $AS=0$  for nearly half of the samples (Supplementary Fig. 8b).

Taken together, these results suggest that the elbow method, a simple and non-parametric method, is robust and superior to some arbitrarily chosen cutoffs (e.g., the 0.1 cutoff used in ref. <sup>6</sup>). However, it was not possible to further test the elbow method in the Chowell et al.'s cohort due to the inaccessibility of some of the data. In the future, the elbow-point-based method needs to be tested in more cohorts to further validate it. Moreover, as tumor purity and ploidy information of samples per tumor type are important factors in detecting CNAs, more sophisticated methods (e.g., iChorCNA<sup>12</sup>, Accurity<sup>14</sup>) are needed to take this information into consideration before determining the cutoff for CNA calling when such data are available.

### DISCUSSION

In summary, we have comparatively assessed the power of AS and FGA in predicting patient survival following immunotherapy in pan-cancer and individual cancer types. Addressing our research questions, we first show that choice of cutoff during CNA calling greatly influences the predictive power of AS and FGA for patient survival following immunotherapy. Specifically, the AS measure defined in ref. <sup>6</sup> ( $AS_{0.1}$ ) cannot significantly predict survival benefit following immunotherapy in low-TMB patients in any single cancer type (Supplementary Fig. 1).  $AS_{0.2}$  and  $FGA_{0.2}$ , re-calculated using a more appropriate pan-cancer CNA calling cutoff of  $|\log_2$  copy ratio|  $> 0.2$ , have a considerably stronger predictive power of survival following immunotherapy (Fig. 1). Second, we show that  $AS_{0.2}$  and  $FGA_{0.2}$  predict pan-cancer survival following immunotherapy for both high-TMB and low-TMB patients, rather than in low-TMB patients only, as was claimed in ref. <sup>6</sup>; as evidence, the arbitrary cutoff of  $|\log_2$  copy ratio|  $> 0.1$  used in ref. <sup>6</sup> is found to misclassify many patients (Fig. 2). Finally, from a translational standpoint, the currently available data suggest that both AS and FGA can significantly predict survival following immunotherapy in only a few cancer types (Figs. 3 and 4). Therefore, larger sample sizes are required to evaluate, and ultimately use these measures within individual cancer types.



**Fig. 4** The elbow-point-based method offers one systematic way to determine the cutoff used for calling CNAs. **a** The elbow method for determining the cutoff of  $|\log_2$  copy ratio| was used in calling AS for individual cancer types (exemplified by esophagogastric cancer here). The AS for each patient with different calling cutoffs are shown in black curves. The mean value of all patients is shown in the red curve. The mean elbow point is shown with 95% confidence intervals, which are calculated using 1000-replicate bootstrapping. **b** The elbow-point values of the cutoff of  $|\log_2$  copy ratio| in calculating AS in individual cancer types. The bars represent 95% confidence intervals of the elbow-point values calculated using a 1000-replicate bootstrapping. **c** Comparison of HRs using AS<sub>0.1</sub> or AS<sub>EP</sub> in a multivariate Cox model with TMB (binned at the 80th percentile) and ICB drug class. The difference of mean HRs of AS<sub>0.1</sub> and AS<sub>EP</sub> and paired Wilcoxon test *P* value are displayed. Wald *P* values for HRs of AS<sub>EP</sub> in individual cancer types are displayed at the right side of the plot. The upper and lower boundaries signify the first and third quartiles, correspondingly, while the central line denotes the median. Whiskers stretch to the most distant data points not classified as outliers (within 1.5 times the interquartile range), and outliers are illustrated as points above and below the box-and-whisker diagram. **d** Univariable Kaplan–Meier survival analysis and multivariable survival analysis using Cox proportional hazards regression of overall survival with AS calculated using cancer-type-specific elbow-point-based CNA calling cutoff (AS<sub>EP</sub>; binned at the 30th percentile), TMB (binned at the 80th percentile), and ICB drug class. Wald *P* values are displayed. Squares positioned at midpoints symbolize point estimates of HRs, and the accompanying bars indicate 95% confidence intervals. The data are from the Samstein et al.'s cohort<sup>1</sup>.

## METHODS

### Patient samples

Data for the Samstein et al.<sup>1</sup> cohort (MSK-IMPACT) were downloaded from cBioPortal at [http://www.cbioportal.org/study?id=tmb\\_mskcc\\_2018](http://www.cbioportal.org/study?id=tmb_mskcc_2018). Segmented copy-number data were downloaded from AACR Project GENIE v.7.1. Note that one sample of *skin cancer-nonmelanoma* was excluded from the analyses as it was the only sample representing this specific histology. Data for the Chowell et al.'s cohort were obtained from the Supplementary Table of ref.<sup>8</sup>.

### Copy-number alteration and tumor aneuploidy assessment

AS were calculated using ASCETS v.1.1 at <https://github.com/beroukhim-lab/ascets> with the following command line:

```
script run_ascets.R -i genie_msk_cna_hg19_seg -c genomic_
arm_coordinates_hg19.txt -o ./output/output -t x
```

where *x* is the CNA calling cutoff, defined as the threshold at which a CNA event is counted if  $|\log_2$  copy ratio| > *x*.

FGA was calculated as the ratio between the sum of the lengths of the genomic segments with  $|\log_2$  copy ratio| > *x*, and the sum

of the lengths of all measured segments:

$$\text{FGA} = \frac{\text{sum}(\text{seg\_length}[\text{abs}(\text{seg.mean}) \geq x])}{\text{sum}(\text{seg\_length})}$$

### CNA calling cutoff point determination

**The elbow-point-based method.** CNA events, which were used to calculate AS and FGA, were first called using  $|\log_2$  copy ratio| cutoffs ranging from 0.01 to 0.5 with a step size of 0.01. Then, to calculate the cancer-type-specific elbow points of cutoffs, mean values of AS/FGA across samples in individual cancer types were calculated under each cutoff to generate the AS/FGA-cutoff curves. Finally, the cancer-type-specific elbow point in each bootstrap replication was calculated using Python package *kneed* v.0.8.1; and 95% confidence intervals of elbow points were determined from 1000-replicate bootstrapping.

**The Gaussian mixture model.**  $|\log_2$  copy ratio| cutoffs were calculated in a similar way as the elbow method, except that in the final step, the cutoff point was determined by the Gaussian mixture model with two components using the *GaussianMixture()* function in the Python package *sklearn* v.1.2.1. This model assumes that the data is generated from two Gaussian

distributions with different means and variances, and that each data point belongs to one of the two distributions with a certain probability.

### Binarization of TMB, AS, and FGA

The patient TMB, AS, and FGA values were binarized into score-high versus score-low groups in a cancer-type-specific manner. Specifically, in each cancer type, the patients who had the top 20% of the TMB values were classified into high-TMB group, while others were classified into low-TMB group following<sup>1</sup>. To determine the optimal binarization of AS and FGA that effectively synergized with TMB for risk stratification of patients undergoing immunotherapy, we performed a comprehensive analysis. This involved testing every tenth quantile within each cancer type, ranging from the 20th to 80th percentile, using a multivariate model that incorporated TMB (binned at the 80th percentile) and ICB drug class following<sup>6</sup>. Leave-one-out cross-validation was conducted to identify the optimal threshold for defining high versus low AS (or FGA). In detail, for each threshold, we constructed a Cox proportional hazards survival model incorporating binarized AS (or FGA), TMB, and drug class. This process was repeated iteratively for the cohort size ( $n = 1660$ ), with one unique patient left out in each iteration. The goal was to identify the threshold that yielded the highest multivariate HR in synergy with TMB, while maintaining a significant Bonferroni-corrected  $P$  value (see Fig. 2a).

### Statistical analysis

**Survival analysis.** Kaplan–Meier survival analysis was performed using the R packages *survminer* v.0.4.9 and *survival* v.3.3.1, and HR and  $P$  values were calculated with univariable Cox proportional hazard regression using the *coxph()* function<sup>19</sup>. Multivariable analysis was performed with Cox proportional hazard regression in individual cancer types, with inclusion of covariates including FGA (or AS), TMB and ICB drug class.

**Power analysis.** The power analysis of minimum sample size estimation for achieving statistically significant survival difference (Kaplan–Meier HR  $> 1$ ,  $P < 0.05$ ) in individual cancer types in the Samstein et al.'s cohort was performed using the R package *powerSurvEpi* v.0.1.3 with parameter “power = 0.8”, which means that there is an 80% chance of correctly detecting a statistically significant effect if one exists.

**Gene mutation frequency analysis.** We defined the gene mutation frequency in a group of patients as the fraction of patients with mutations in the gene of interest. To identify genes with significantly different mutation frequencies between AS (or FGA) high and low groups, we compared the gene mutation frequencies in the two groups using the chi-squared test. We used the *chi2\_contingency()* function from the Python package *scipy* v.1.10.1 to perform the chi-squared test. To correct for multiple testing, we applied the Bonferroni correction.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

Data for the Samstein et al.'s cohort are available at [https://www.cbiportal.org/study/summary?id=tmb\\_mskcc\\_2018](https://www.cbiportal.org/study/summary?id=tmb_mskcc_2018) and the GENIE<sup>20</sup> v.7.1 release: <https://www.synapse.org/#!Synapse:syn7222066/wiki/405659>. Data for the Chowell et al.'s cohort are available from the Supplementary Table of ref. <sup>8</sup> at [https://static-content.springer.com/esm/art%3A10.1038%2F541587-021-01070-8/MediaObjects/41587\\_2021\\_1070\\_MOESM3\\_ESM.xlsx](https://static-content.springer.com/esm/art%3A10.1038%2F541587-021-01070-8/MediaObjects/41587_2021_1070_MOESM3_ESM.xlsx), where FGA, TMB, ICB drug class, and overall survival information are provided. Aneuploidy scores were called using ASCETS at

<https://github.com/beroukhim-lab/ascets> and values for each sample are provided in the GitHub repository at <https://github.com/rootchang/Aneuploidy-FGA-ICB>.

### CODE AVAILABILITY

All code necessary to replicate these analyses is provided in the following GitHub repository: <https://github.com/rootchang/Aneuploidy-FGA-ICB>.

Received: 1 March 2023; Accepted: 25 May 2023;

Published online: 03 June 2023

### REFERENCES

- Samstein, R. M. et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **51**, 202–206 (2019).
- McGrail, D. J. et al. High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann. Oncol.* **32**, 661–672 (2021).
- Ben-David, U. & Amon, A. Context is everything: aneuploidy in cancer. *Nat. Rev. Genet.* **21**, 44–62 (2020).
- Hieronimus, H. et al. Tumor copy number alteration burden is a pan-cancer prognostic factor associated with recurrence and death. *eLife* **7**, <https://doi.org/10.7554/eLife.37294> (2018).
- Sansregret, L. & Swanton, C. The role of aneuploidy in cancer evolution. *Cold Spring Harb. Perspect. Med.* **7**, <https://doi.org/10.1101/cshperspect.a028373> (2017).
- Spurr, L. F., Weichselbaum, R. R. & Pitroda, S. P. Tumor aneuploidy predicts survival following immunotherapy across multiple cancers. *Nat. Genet.* **54**, 1782–1785 (2022).
- Spurr, L. F. et al. Quantification of aneuploidy in targeted sequencing data using ASCETS. *Bioinformatics* **37**, 2461–2463 (2021).
- Chowell, D. et al. Improved prediction of immune checkpoint blockade efficacy across multiple cancer types. *Nat. Biotechnol.* **40**, 499–506 (2022).
- Rizvi, H. et al. Molecular determinants of response to anti-programmed cell death (PD)-1 and anti-programmed death-ligand 1 (PD-L1) blockade in patients with non-small-cell lung cancer profiled with targeted next-generation sequencing. *J. Clin. Oncol.* **36**, 633–641 (2018).
- Woo, X. Y. et al. Conservation of copy number profiles during engraftment and passaging of patient-derived cancer xenografts (vol 53, pg 86, 2021). *Nat. Genet.* **53**, 761–761 (2021).
- Beroukhim, R. et al. The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
- Adalsteinnsson, V. A. et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat. Commun.* **8**, <https://doi.org/10.1038/s41467-017-00965-y> (2017).
- Hoge, A. C. H. et al. DNA-based copy number analysis confirms genomic evolution of PDX models. *NPJ Precis. Onc.* **6**, <https://doi.org/10.1038/s41698-022-00268-6> (2022).
- Luo, Z. H., Fan, X. P., Su, Y. & Huang, Y. S. Accurity: accurate tumor purity and ploidy inference from tumor-normal WGS data by jointly modelling somatic copy number alterations and heterozygous germline single-nucleotide-variants. *Bioinformatics* **34**, 2004–2011 (2018).
- Satopaa, V., Albrecht, J., Irwin, D. & Raghavan, B. Finding a “kneedle” in a haystack: detecting knee points in system behavior. in *2011 31st International Conference on Distributed Computing Systems Workshops* 166–171 (IEEE, 2011).
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S. & Satoto, B. D. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conf. Ser.: Mater. Sci. Eng.* **336**, 012017 (2018).
- Linting, M., Meulman, J. J., Groenen, P. J. F. & van der Kooij, A. J. Nonlinear principal components analysis: Introduction and application. *Psychol. Methods* **12**, 336–358 (2007).
- Oh, J. H., Hong, J. Y. & Baek, J. G. Oversampling method using outlier detectable generative adversarial network. *Expert Syst. Appl.* **133**, 1–8 (2019).
- Therneau, T. M. A package for survival analysis in R. R package version 4.2-0, <https://CRAN.R-project.org/package=survival> (2020).
- Andre, F. et al. AACR Project GENIE: powering precision medicine through an International Consortium. *Cancer Discov.* **7**, 818–831 (2017).

### ACKNOWLEDGEMENTS

This research was supported in part by the NIH Intramural Research Program, National Cancer Institute. This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>). The authors would like to acknowledge the American Association for Cancer Research and its financial and material support in

the development of the AACR Project GENIE registry, as well as members of the consortium for their commitment to data sharing. Interpretations are the responsibility of the study authors.

### AUTHOR CONTRIBUTIONS

T.-G.C. and E.R. conceived and designed the study. T.-G.C. and Y.C. collected and managed the data. T.-G.C., Y.C., and E.D.S. performed the statistical analyses. U.B.-D. and A.A.S. provided statistical advice. All authors critically revised the manuscript for important intellectual content.

### FUNDING

Open Access funding provided by the National Institutes of Health (NIH).

### COMPETING INTERESTS

E.R. is a co-founder of MedAware, Metabomed and Pangea Biomed (divested), and an unpaid member of Pangea Biomed's scientific advisory board. U.B.-D. receives grant funding from Novocure. The remaining authors declare no competing interests.

### ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41698-023-00408-6>.

**Correspondence** and requests for materials should be addressed to Tian-Gen Chang or Eytan Ruppin.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023