Original Article

Tumor mutation burden derived from small next generation sequencing targeted gene panel as an initial screening method

Yuan Tang[1], Yuli Li[1], Weiya Wang[1], Analyn Lizaso[2], Ting Hou[2], Lili Jiang[1], Meijuan Huang[3]

[1]Department of Pathology, West China Hospital, Chengdu 610041, China; [2]Burning Rock Biotech, Guangzhou 510300, China; [3]Department of Thoracic Oncology, West China Hospital, Chengdu 610041, China

Contributions: (I) Conception and design: L Jiang, M Huang, Y Tang; (II) Administrative support: L Jiang, M Huang; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: Y Tang, Y Li, W Wang; (V) Data analysis and interpretation: A Lizaso, T Hou; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Lili Jiang, MD; Meijuan Huang, MD. West China Hospital, No. 37 Guo Xue Xiang, Chengdu 610041, China.
Email: 879876047@qq.com; hmj107@163.com.

Background: With the increasing use of immune checkpoint inhibitors, tumor mutation burden (TMB) assessment is now routinely included in reports generated from targeted sequencing with large gene panels; however, not all patients require comprehensive profiling with large panels. Our study aims to explore the feasibility of using a small 56-gene panel as a screening method for TMB prediction.

Methods: TMB from 406 non-small cell lung cancer (NSCLC) patients was estimated using a large 520-gene panel simulated with the prospective TMB status for the small panel. This information was then used to determine the optimal cut-off. An independent cohort of 30 NSCLC patients was sequenced with both panels to confirm the cut-off value.

Results: By comparing sensitivity, specificity, and positive predictive value (PPV), the cut-off was set up as 10 mutations/megabase, yielding 81.4% specificity, 83.6% sensitivity, and 62.4% PPV. Further validation with an independent cohort sequenced with both panels using the same cut-off achieved 95.7% sensitivity, 71.4% specificity and 91.7% PPV. The decreasing trend of sensitivity with the increasing trend of both specificity and PPV with a concomitant increase in the cut-off for the small panel suggests that TMB is overestimated but highly unlikely to yield false-positive results. Hence, patients with low TMB (<10) can be reliably stratified from patients with high TMB (≥10).

Conclusions: The small panel, more cost-effective, can be used as a screening method to screen for patients with low TMB, while patients with TMB ≥10 are recommended for further validation with a larger panel.

Keywords: Non-small cell lung cancer (NSCLC); small gene panel; tumor mutation burden (TMB); TMB in NSCLC

Submitted Nov 27, 2019. Accepted for publication Dec 19, 2019.
doi: 10.21037/tlcr.2019.12.27
View this article at: http://dx.doi.org/10.21037/tlcr.2019.12.27

Introduction

Solid tumors with high tumor mutation burden (TMB), including melanoma and non-small cell lung cancer (NSCLC), demonstrate remarkable responses to immune checkpoint inhibitors (1-4). It has been hypothesized that tumors with high TMB are more likely to harbor tumor-specific antigenic peptides or neoantigens, which make them targets of activated immune cells and results in a positive response with immunotherapy (1,5).

Neoantigen load, although not clinically observable, has been shown to be associated with the burden of mutation and immunotherapy response (2,6). TMB was therefore considered an emerging predictive biomarker for the effectiveness of inhibitors of the immune checkpoint

(1-3,6). The number of mutations present in a tumor sample is estimated to be TMB or mutation load. It is measured as the number of somatic mutations per megabase (Mb) of the interrogated genome, including single nucleotide and nonsense variants and short insertion-deletion variations.

Whole exome sequencing (WES), spanning about 30 to 50 Mb of coding sequences, has been traditionally used to estimate TMB (3). Advanced NSCLC patients with high TMB estimated by WES who received single-agent or combination immune checkpoint inhibitors were more likely to have improved objective response, more durable clinical benefit and longer progression-free survival than patients with low TMB (2,7,8). In contrast to WES, targeted next-generation sequencing (NGS) with gene panels consisting of 300 to 500 genes, spanning about 1 to 3 Mb of the genome, can accurately sequence genes or regions of interest at a much higher depth to reveal therapeutically-relevant mutations (9,10). The high correlation of TMB derived from WES and large targeted gene panels has then prompted the replacement of WES in the routine clinical assessment of TMB (3,4,11,12); however, its use is still limited by high cost and longer turnaround time. Moreover, despite providing a comprehensive mutation profile, information gathered from sequencing with large targeted panels is not needed in the therapeutic decisions for some patients.

Conversely, smaller targeted gene panels have been increasingly utilized in clinical practice, particularly in lung cancer, due to its cost-effectiveness and faster turnaround time as compared with the use of large targeted gene panels and WES (13). Like large gene panels, targeted sequencing using small gene panel can also simultaneously find genomic alterations in cancer-related genes but only in a limited number of genes. With the more common use of small targeted gene panel, we hypothesized that predicting TMB from it can be clinically useful as a first and cost-effective screening method. In this study, we explored the feasibility of using a 56-gene panel covering 0.26 Mb of the human genome to predict TMB.

To reach this goal, we first simulated the target regions of the small gene panel using sequencing data obtained with the large gene panel and derived an optimal TMB cut-off using a training dataset. We then confirmed the utility of this cut-off using an independent cohort sequenced using the small panel and further confirmed with the large panel.

Methods

Patients

Data derived from targeted sequencing using 520-gene panel (OncoScreen Plus, Burning Rock Biotech, China) from a total of 406 NSCLC patients with varied TMB status were used as the training dataset to identify the genes correlated with high TMB, and derive the optimal TMB cut-off through simulations.

An independent cohort comprised of 30 NSCLC patients with various stages and histological types from our institution who were referred for comprehensive molecular testing at Burning Rock Biotech [College of American Pathologists (CAP)-accredited/Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory] between February 2017 and May 2018 were included for the validation stage of this study. Two independent pathologic examinations confirmed tumor histology. Tumors were staged according to the American Joint Committee on Cancer 7th edition TNM staging system of NSCLC (14).

This study was approved by the relevant Institutional Review Board of the West China Hospital and performed following the ethical standards of West China Hospital and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standard. Prior written informed consent was obtained from each of the recruited patients for the use of their plasma and/or tissue samples in further molecular studies.

Tissue DNA isolation and capture-based targeted DNA sequencing

Tissue DNA was extracted from formalin-fixed, paraffin-embedded (FFPE) tumor tissues using QIAamp DNA FFPE tissue kit (Qiagen). A minimum of 50 ng of DNA is required for NGS library construction. Tissue DNA was sheared using Covaris M220 (Covaris, MA, USA), followed by end repair, phosphorylation, and adaptor ligation. Fragments between 200 to 400 base pairs from the sheared tissue DNA were purified (Agencourt AMPure XP Kit, Beckman Coulter, CA, USA), followed by hybridization with capture probes baits, hybrid selection with magnetic beads and PCR amplification. The quality and the size of the fragments were assessed using Qubit 2.0 Fluorimeter with the dsDNA high-sensitivity assay kit (Life Technologies, Carlsbad, CA).

Indexed samples were sequenced on Nextseq500 (Illumina, Inc., USA) with paired end reads and average sequencing depth of 1,000×.

Sequence data analysis

Sequence data were mapped to the reference human genome (hg19) using Burrows-Wheeler Aligner v.0.7.10 (15). Local alignment optimization and variant calling were performed using Genome Analysis Tool Kit v.3.2 (16), and VarScan v.2.4.3 (17). Variants were filtered using the VarScan fpfilter pipeline, loci with depth less than 100 were filtered out. Base-calling in tissue samples required at least 8 and 5 supporting reads for single nucleotide variations (SNVs) and short insertion and deletion variations (INDEL), respectively. Variants with population frequency over 0.1% in the ExAC, 1000 Genomes, dbSNP, or ESP6500SI-V2 databases were grouped as single nucleotide polymorphisms and excluded from further analysis. Remaining variants were annotated with ANNOVAR (18) and SnpEff v.3.6 (19). Analysis of DNA translocation was performed using Factera v.1.4.3 (20).

TMB per patient was computed as a ratio between the total number of mutations detected with the total coding region size of the panel used (i.e., 520-gene OncoScreen Plus panel with 1.26 Mb, 56-gene LungCore panel with 0.25 Mb) using the formula in Eq. [1]. Copy number variations (CNV), fusions, large genomic rearrangements and mutations occurring on the kinase domain of *EGFR* and *ALK* were excluded from the mutation count.

$$\text{TMB} = \frac{\text{mutation count (except for CNV and fusion)}}{\text{the total size of the coding region of the panel used}} \quad [1]$$

Model construction

Various learning methods, including support vector machine (SVM) (21), naive Bayesian, Bayesian network, logistic regression, additive logistic regression (22), random forest, and a multiclass probabilistic classifier were executed to construct the most optimal model in predicting TMB from the small 56-gene panel. These learning algorithms were implemented by WEKA v3.6 (23), with the corresponding packages, including LibSVM, NaïveBayes, BayesNet, Logistic, LogitBoost, RandomForest, MultiClassClassifier, and SVM. Default parameters were obeyed throughout the supervised learning process.

Model performance validation

The sensitivity (Eq. [2]), specificity (Eq. [3]), positive predictive value (PPV) (Eq. [4]) and Matthew's correlation coefficient (MCC) (Eq. [5]) were calculated accordingly using the formula listed below considering the highest proportion of true positives and the lowest number of false negatives. Sensitivity is defined as the percentage of positive data correctly predicted. Specificity is defined as the percentage of negative data correctly predicted. PPV is defined as the percentage of positive results that are a true positive. MCC is a comprehensive indicator that considers both positive and negative data.

$$Sensitivity = \frac{TP}{TP + FN} \quad [2]$$

$$Specificity = \frac{TN}{TN + FP} \quad [3]$$

$$PPV = \frac{TP}{TP + FP} \quad [4]$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad [5]$$

Where TN, TP, FN, FP represent true negative, true positive, false negative, and false positive, respectively.

Statistical analysis

All the data were analyzed using the R statistics package (R v3.4.0; R: The R-Project for Statistical Computing, Vienna, Austria). Differences in the groups were calculated and presented using either Fisher's exact test or two-tailed Student's t-test, as appropriate. P value with $P<0.05$ was considered as statistically significant.

Results

Patient characteristics

A total of 406 NSCLC patients of varied histology and disease stage were included in the training dataset. Among the 406 patients, 62.6% (254/406) of the patients were male, while the remaining 37.4% (152/406) were females. The median age of the cohort was 60 years, ranging from 43 to 88 years. Most of the patients were diagnosed with lung

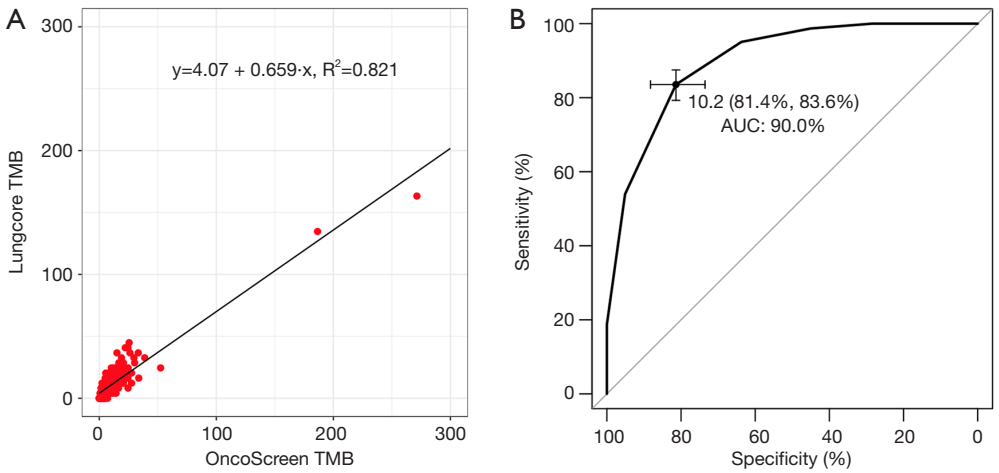

Figure 1 Deriving the TMB cut-off values. (A) Regression analysis revealed the correlation between the TMB estimated from the training dataset using the small gene panel (X-axis) and the large gene panel (Y-axis). (B) Receiver operating characteristic (ROC) plotting the specificity (X-axis) and sensitivity (Y-axis) revealed an area under the curve (AUC) of 90.0% with a TMB cut-off of 10.2 mutations/Mb. TMB, tumor mutation burden.

adenocarcinoma (61.8%, 251/406), while 12.3% (50/406) were diagnosed with squamous cell lung carcinoma. Based on the TMB estimated from the large 520-gene panel, 7.9% (32/406) of the patients have TMB ≥20 mutations/Mb, 17.2% (70/406) have TMB between 10–20 mutations/Mb and 75.4% (306/406) had TMB <10 mutations/Mb.

Among the 30 patients included as an independent cohort for the validation stage of the study, a majority were males (86.7%, 26/30). The median age of the validation cohort was 61 years, ranging from 31 to 79 years. About half of the patients (56.7%, 17/30) were diagnosed with lung adenocarcinoma, 40.0% (12/30) were diagnosed with squamous cell lung carcinoma, and a patient was diagnosed with large cell neuroendocrine lung carcinoma. The cohort was comprised of 30% (9/30) patients with early-stage disease, 53.3% (16/30) patients with advanced-stage disease and the remaining 5 patients had unknown disease stage. Early-stage patients included 4 patients with stage IB, 2 patients each with stage IIA and IIB, respectively and 1 patient with stage IIIA. Advanced-stage patients included 4 patients with stage IIIB, 1 patient with stage IIIC, 10 patients with stage IVA, and 1 patient with stage IVB.

Identifying the optimal cut-off for the small panel

In order to investigate the use of the small 56-gene panel to estimate TMB, by using data simulation, we first extracted the target regions covered by the small gene panel from the sequencing data obtained using the large gene panel for the training cohort composed of 406 NSCLC patients. TMB for both large and small panel is similarly calculated as the ratio of the number of SNVs per coding region of the genome covered by the gene panel as expressed in Eq.1 in the Methods section. We then analyzed the correlation of actual TMB calculated from the large gene panel and the simulated TMB for the small gene panel. Furthermore, we also derived an optimal TMB cut-off that could effectively identify patients with low and high TMB.

The large targeted gene panel (OncoScreen Plus, Burning Rock Biotech, China) includes 520 cancer development-related genes spanning 1.64 Mb of the human genome (1.26 Mb excluding the regions not included in TMB estimation). Previous validation of the 520-gene panel revealed high correlation between actual TMB estimated from the WES data of 8,092 patient samples with 35 types of cancers from the Cancer Genome Atlas (TCGA) and the simulated TMB from target regions included in the 520-gene panel derived from the WES data (Pearson correlation coefficient R^2=0.976; data not shown). In contrast, the small targeted gene panel (LungCore, Burning Rock Biotech, China) includes 56 lung cancer-related genes spanning 0.28 Mb of the human genome (0.25 Mb excluding the regions not included in TMB estimation). The fitted curve illustrating the correlation between the TMB estimated from the large gene panel and the small gene panel is shown in *Figure 1A*. The TMB estimated from small gene panel,

Table 1 Derivation of the optimal TMB cut-off for the small gene panel using TMB cut-off value of 10 mutations/Mb for the 520-gene panel from 406 NSCLC patients

Cutoff for the small panel	TP	FP	TN	FN	Sensitivity	Specificity	PPV
10	83	50	254	19	81.4%	83.6%	62.4%
15	65	15	289	37	63.7%	95.1%	81.3%
20	46	4	300	56	45.1%	98.7%	92.0%
21	29	0	304	73	28.4%	100.0%	100.0%
25	16	0	304	86	15.7%	100.0%	100.0%
30	11	0	304	91	10.8%	100.0%	100.0%
35	8	0	304	94	7.8%	100.0%	100.0%
40	5	0	304	97	4.9%	100.0%	100.0%
41	3	0	304	99	2.9%	100.0%	100.0%

NSCLC, non-small cell lung cancer; TMB, tumor mutation burden; TP, true positive; FP, false positive; TN, true negative; FN, false negative; PPV, positive predictive value.

and large gene panel is highly correlated, with a regression coefficient (R^2) of 0.821 and a Pearson correlation coefficient of 0.906. The genes included in the large and small gene panels are listed in *Tables S1* and *S2*, respectively.

Based on the TMB cut-off of 10 mutations/Mb for large gene panel, we divided the data for the small gene panel into TMB ≥10 and TMB <10. According to the receiver operating characteristic (ROC) curve illustrated in *Figure 1B*, the most optimal cut-off was 10.2 mutations/Mb for the small panel achieving a specificity of 81.4%, sensitivity of 83.6%, and area under the curve (AUC) of 90.0%. Hence, we have set the cut-off as 10 mutations/Mb.

Statistical performance indicators for TMB estimation using the small gene panel

After identifying the optimal TMB cut-off that could distinguish between low and high TMB, we next determined the impact of the TMB cut-off value of 10 mutations/Mb on the performance of TMB estimation using the small panel by calculating statistical performance indicators including sensitivity, specificity and PPV. Furthermore, seven different machine learning algorithms were employed to analyze the performance of the small gene panel in distinguishing and stratifying the TMB using 10 mutations/Mb as the cut-off.

Table 1 summarizes the statistical indicators for the performance of the small gene panel in estimating TMB. At a TMB cut-off of 10 mutations/Mb, the specificity and PPV were 83.6% and 62.4%, respectively. Both the specificity and PPV had an increasing trend with the increase in TMB achieving 100% when the TMB was at 21 mutations/Mb (*Figure S1*, *Table 1*). Meanwhile, the sensitivity had an opposite trend, with the highest sensitivity of 81.4% achieved at 10 mutations/Mb cut-off and concomitantly decreased with the increase in TMB (*Table 1*). The cut-off was confirmed as 10 mutations/Mb considering the highest proportion of true positives and the lowest number of false negatives.

Further analysis of the analytical performance with seven different learning algorithms using a TMB cut-off of 10 mutations/Mb achieved specificity between 92.1% and 96.4% (*Table 2*). Among the algorithms, BayesNet revealed the highest sensitivity of 70.6%, followed by the sensitivity of 67.6% from both Logistic and MultiClassClassifier. Meanwhile, Logistic revealed the highest sensitivity, PPV and MCC of the model achieving 96.4%, 85.9%, and 68.3%, respectively. Moreover, specificity, PPV and MCC of 95.1%, 82.1%, and 67.1% were respectively achieved by both Logistic and MultiClassClassifier (*Table 2*).

The inverse relationship between the sensitivity and TMB cut-off strongly supports the fact that small targeted gene panels are not dependable for the estimation of high TMB. In contrast, the increasing trend in both specificity and PPV with the corresponding increase in TMB demonstrated that TMB is likely to be overestimated by the small gene panel, but the likelihood of false-positive results is very low. Since TMB is overestimated, samples with TMB values above the cut-off (>10 mutations/Mb)

Table 2 Performance validation of TMB estimation using derived TMB cut-off value of 10 mutations/Mb for the small gene panel from 406 NSCLC patients

Method	TP	TN	FP	FN	Sensitivity	Specificity	PPV	MCC
NaiveBayes	59	291	13	43	57.8%	95.7%	81.9%	60.8%
BayesNet	72	280	24	30	70.6%#	92.1%	75.0%	64.0%
Logistic*	69	289	15	33	67.6%	95.1%	82.1%	67.1%
LogitBoost	65	289	15	37	63.7%	95.1%	81.3%	64.1%
RandomForest	63	284	20	39	61.8%	93.4%	75.9%	59.4%
SVM*	67	293	11	35	65.7%	96.4%#	85.9%#	68.3%#
MultiClassClassifier*	69	289	15	33	67.6%	95.1%	82.1%	67.1%

*, methods in bold represent the most suitable method for TMB estimation; #, values in bold indicate the highest value for sensitivity, specificity, PPV, or MCC. NSCLC, non-small cell lung cancer; TMB, tumor mutation burden; TP, true positive; FP, false positive; TN, true negative; FN, false negative; PPV, positive predictive value; MCC, Matthew's correlation coefficient.

requires revalidation with a larger gene panel, while samples with TMB values below the cut-off (<10 mutations/Mb) are more likely to be accurate and does not need further validation.

Identifying known genes and alterations associated with increased TMB

We further identified the specific genes associated with high TMB within the large 520-gene panel by performing statistical analysis on the sequencing data from 406 NSCLC patients with varied TMB status. This step aims to identify the genes that are associated with high TMB from the large gene panel and determine if they are present in the small gene panel.

At a TMB cut-off of 10 mutations/Mb, patients with high TMB had significantly more mutations in a total of 106 genes than patients with low TMB. Of these, 26 of these genes were part of the small gene panel and covered 46.4% (26/56) of the genes in the small panel (*Table S3*).

Interestingly, *TP53* mutations were the most predominant mutation among patients with high TMB (P<0.001, *Table S3*, *Figure S2*). Among all the other actionable mutations, *ALK* and *ROS1* fusions were also more likely to be detected among patients with low TMB (*ALK* fusion P=0.0095; *ROS1* fusion P=0.043).

Validation of TMB estimation with small gene panel using an independent cohort

After identifying the optimal cut-off and establishing the feasibility of TMB estimation with the small gene panel using simulated data from the training cohort, we next aimed to validate our findings with the use of an independent cohort consisting of an additional 30 NSCLC patients. This cohort was sequenced using both the small and the large gene panels to compare the TMB estimated from both panels. Furthermore, the statistical performance of the small gene panel was also evaluated with learning algorithms.

The mutation detection rate of *TP53* was 67%, with 91.7% (11/12) of the patients having TMB ≥20 mutations/Mb, 72.7% (8/11) having TMB between 10 to 20 mutations/Mb and 14.3% (1/7) having TMB <10 mutations/Mb (*Figure S3*).

Table 3 lists the TMB estimated with the small (LungCore) and further validated with the large (OncoScreen Plus) gene panel for each of the 30 patients. Most of the patients (66.7%, 20/30) had 5 or more mutations detected with an estimated TMB of ≥20 mutations/Mb. Four patients had TMB between 10–20 mutations/Mb, while the remaining 6 patients had TMB <10 mutations/Mb. In contrast, based on the TMB validated with the 520-gene panel, 40.0% (12/30) of the patients had TMB ≥20 mutations/Mb, 36.7% (11/30) had TMB between 10–20 mutations/Mb and the remaining 23.3% (7/30) had TMB <10 mutations/Mb.

The analytical performance of the small gene panel in TMB estimation using a cut-off value of 10 mutations/Mb achieved specificity of 71.4%, PPV of 91.7%, and sensitivity of 95.7% (*Table 4*, *Figure S4*). Consistent with the trend observed in the training dataset (*Table 1*), the specificity

Table 3 Estimated TMB of the 30 NSCLC patients from the small and large gene panels

Patient ID	MaxAF (LungCore)	Mutation count (LungCore)	TMB LungCore	TMB OncoScreen
F21510*	61.31%	19	77.6	90.5
F21511*	13.27%	12	49.0	39.7
F21501*	79.46%	11	44.9	68.3
F21492*	76.35%	11	44.9	51.6
F21508*	61.90%	10	40.8	49.2
F21514*	57.33%	9	36.7	20.6
F21516*	35.58%	9	36.7	20.6
F21503*	57.91%	9	36.7	16.7
F21498*	48.62%	8	32.7	11.1
F21496*	35.28%	7	28.6	34.9
F21515*	91.69%	7	28.6	26.2
F21502*	80.71%	6	24.5	29.4
F21504*	14.76%	6	24.5	20.6
F21495*	35.66%	6	24.5	17.5
F21505*	23.57%	6	24.5	15.9
F21493*	19.28%	6	24.5	11.1
F21499*	81.11%	5	20.4	23.0
F21513*	65.59%	5	20.4	12.7
F21506*	67.28%	5	20.4	11.9
F21497*	45.69%	5	20.4	11.1
F45309*	56.47%	4	16.3	12.7
F45303	55.60%	4	16.3	5.6
F45305*	86.63%	3	12.2	17.5
F45308	44.86%	3	12.2	4.8
F45304	54.78%	2	8.2	4.0
F45307	71.28%	2	8.2	5.6
F45298	38.30%	1	4.1	10.3
F45299	92.37%	1	4.1	4.8
F45300	31.24%	1	4.1	2.4
F45302	44.65%	0	0	5.6

Patient data with * indicate the patients whose actual TMB data from the small panel matches the actual TMB data from the 520-gene panel according to the cutoff of 10 mutations/Mb. NSCLC, non-small cell lung cancer; TMB, tumor mutation burden.

and PPV also had an increasing trend, with both specificity and PPV reaching 100% at a TMB of 20 mutations/Mb. Meanwhile, the sensitivity also had a decreasing trend with the concomitant increase in the TMB in the validation cohort (*Table 4*). Consistently, the analytical performance evaluated by the seven different learning algorithms revealed the sensitivity of 91.3%, specificity, and PPV of 100% and MCC of 84.3% in both LogitBoost and SVM

Table 4 Performance metrics for TMB estimation with the small gene panel from 30 NSCLC patients

Cutoff for the small panel	TP	FP	TN	FN	Sensitivity	Specificity	PPV
10	22	2	5	1	95.7%	71.4%	91.7%
15	21	1	6	2	91.3%	85.7%	95.5%
20	20	0	7	3	87.0%	100.0%	100.0%
21	16	0	7	7	69.6%	100.0%	100.0%
25	11	0	7	12	47.8%	100.0%	100.0%
30	9	0	7	14	39.1%	100.0%	100.0%
35	8	0	7	15	34.8%	100.0%	100.0%
40	5	0	7	18	21.7%	100.0%	100.0%
41	4	0	7	19	17.4%	100.0%	100.0%

NSCLC, non-small cell lung cancer; TMB, tumor mutation burden; TP, true positive; FP, false positive; TN, true negative; FN, false negative; PPV, positive predictive value.

Table 5 Performance metrics of TMB cut-off value of 10 mutations/Mb for the small gene panel using the data from 30 NSCLC patients

Method	TP	TN	FP	FN	Sensitivity	Specificity	PPV	MCC
NaiveBayes	21	6	1	2	91.3%	85.7%	95.5%	73.7%
BayesNet	21	6	1	2	91.3%	85.7%	95.5%	73.7%
Logistic	19	7	0	4	82.6%	100.0%	100.0%	72.5%
LogitBoost*	21	7	0	2	91.3%#	100.0%#	100.0%#	84.3%#
RandomForest	20	6	1	3	87.0%	85.7%	95.2%	67.1%
SVM*	21	7	0	2	91.3%#	100.0%#	100.0%#	84.3%#
MultiClassClassifier	19	7	0	4	82.6%	100.0%	100.0%	72.5%

*, methods in bold represent the most suitable method for TMB estimation; #, values in bold indicate the highest value for sensitivity, specificity, PPV, or MCC. NSCLC, non-small cell lung cancer; TMB, tumor mutation burden; TP, true positive; FP, false positive; TN, true negative; FN, false negative; PPV, positive predictive value; MCC, Matthew's correlation coefficient.

models (*Table 5*).

These data taken together indicate that the cut-off of 10 mutations/Mb estimated from the small 56-gene panel could reliably stratify patients with low (<10) and high (≥10) TMB. Furthermore, by being overestimated, the TMB of the patients having low TMB were accurately estimated by the small panel, while the patients with high TMB required further validation with larger targeted gene panel.

Discussion

To the best of our knowledge, this is the first study to evaluate the utility of a small 56-gene panel to derive TMB as a first screening method. Our results suggest that at a cut-off of 10 mutations/Mb, TMB derived from the small 56-gene panel can reliably identify the subset of patients with low TMB (<10 mutations/Mb). Also, the results were able to identify the individuals who would likely not benefit from TMB estimation using large targeted gene panel and the subset of patients with high TMB (≥10 mutations/Mb). Additionally, it was able to identify those who require further evaluation with large gene panel. Since smaller gene panels, such as the 56-gene panel analyzed in our study, are the more frequently used targeted panel in clinical practice, the inclusion of this TMB estimation into the clinical reports can provide a more meaningful contribution for making a timely treatment decision for lung cancer patients.

Previous studies have published that there is a highly imprecise estimate of TMB from small gene panels covering approximately 0.5 Mb of the human genome (10).

With the inclusion of about a third of the genes more likely associated with high TMB within the small LungCore panel (46.4%, 26/56), despite being limited, TMB estimation with the small panel still has the potential to be informative and clinically relevant. Using both simulated and actual TMB from the small panel, performance validation consistently revealed the same trend of increasing specificity and PPV, and decreasing sensitivity with a concomitant increase in TMB, strongly suggesting the likelihood of overestimating TMB with the small panel. Consistent with the results from the simulated data, the actual TMB estimated from both the small panel and the 520-gene panel in 30 patients in the independent validation cohort proved this concept. With a cut-off of 10 mutations/Mb, all of the patients with TMB ≥10 mutations/Mb (22/22) estimated by the 520-gene panel similarly had TMB ≥10 mutations/Mb from the small panel; however, only 91.7% (22/24) of the patients with TMB ≥10 mutations/Mb from the small panel had TMB ≥10 mutations/Mb estimated by the 520-gene panel. This data strongly shows the overestimation of TMB with the small gene panel, thus requiring further validation with a larger gene panel for the correct TMB estimation.

On the contrary, all the patients (6/6) with TMB <10 mutations/Mb estimated from the small panel consistently had TMB <10 mutations/Mb estimated by the 520-gene panel, indicating that estimation of low TMB is very accurate with no false-negative calls. These data further suggest that the TMB cut-off of 10 mutations/Mb from the small panel can accurately stratify the patients with low TMB who would not likely benefit from immunotherapy and does not require sequencing with a large gene panel. In addition, 5 of the patients who have more than 10 mutations detected and have estimated TMB of >40 mutations/Mb with the small gene panel consistently had TMB between 39.7 to 90.5 mutations/Mb from the 520-gene panel. These findings indicate that TMB estimation from the small gene panel, although not entirely accurate, can still serve as a valuable reference.

The increased use of immune checkpoint inhibitor therapy in advanced NSCLC patients whose tumors do not harbor actionable *EGFR* or *ALK* mutations have driven the need to establish a biomarker in predicting therapeutic benefit. TMB, although still controversial, has now been adopted as a predictive biomarker for immunotherapy response. Traditionally, TMB was assessed using WES until data simulation studies have demonstrated the feasibility of using targeted NGS with gene panels consisting of 300 to 500 genes (3,4,11). Several reports have since proven the utility of large targeted gene panels in accurately predicting TMB (3,4,8,10,11). Although large targeted gene panels providing a more comprehensive mutational profile of solid tumors, they are still substantially limited by their high cost and longer turnaround time. Recent reports have demonstrated that smaller targeted gene panels interrogating about 150 genes from blood samples were sufficient for estimating TMB.

Moreover, TMB estimated from the 150-gene panel were correlated with immune checkpoint inhibitor response in Chinese NSCLC patients, with patients having blood TMB (bTMB) of more than 6 mutations/Mb, considered as high bTMB, correlated with longer progression-free survival than those with low bTMB (P=0.001) (24). By providing a more concise but informative mutation profile, small targeted panels can serve as practical alternatives to large panels in clinical practice. Thus, the inclusion of TMB estimation into the analysis pipeline and clinical reports for small targeted gene panel could extend its utility as an initial TMB screening method, which can easily be integrated into the hospital-based diagnostic sequencing laboratories. Generally, NSCLC patients with actionable mutations benefit most from using small gene panels by revealing actionable mutations and allowing them prompt access to targeted drugs. On the other hand, patients with no actionable mutations revealed by testing with the small gene panel would either be treated with cytotoxic chemotherapy regimen or be recommended to undergo further molecular testing to explore other treatment options. By including TMB estimation in the small gene panel workflow, it extends the utility of small gene panels, particularly in patients with no actionable mutations, and ensures the timely treatment decisions for the patients that do not need further TMB validation with the larger gene panel.

Since our study is a proof of concept study in the utility of small gene panel in TMB prediction, we believe that the use of patient samples with various histologies and disease stage until mutations were no longer detected, and TMB could not be estimated, does not affect the study results.

Despite the limited number of patients and the inclusion of patients with various disease stage and histologies in both the training and validation cohorts, our findings demonstrate the reliability of the small targeted gene panel as an initial TMB screening method to distinguish the subset of patients with low TMB who will not benefit from sequencing with a large panel, thus providing a cost-effective and convenient screening method that provides meaningful application in clinical practice. Our study is also

limited by the lack of clinical outcomes from the patients. Prospective studies with a larger cohort are needed to validate the predictive value of data from small gene panel as well as the clinical outcome.

Conclusions

Our study demonstrated that a small targeted gene panel could provide additional information on the TMB of patients, albeit limited, indicating its potential as a cost-effective and convenient screening method and adding to its utility in the timely diagnosis and management of lung cancer patients.

Acknowledgments

The authors thank all the patients and their families. We also thank the investigators, study coordinators, operation staff, and the whole project team who worked on this study.
Funding: This work was supported by Sichuan Provincial Research Foundation for Basic Research (grant number 2018SZ0023 to M Huang).

Footnote

Conflicts of Interest: A Lizaso and T Hou are employees of Burning Rock Biotech. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was approved by the relevant Institutional Review Board of the West China Hospital. Prior written informed consent was obtained from each of the recruited patients.

Data Sharing Statement: No additional data available.

References

1. Snyder A, Makarov V, Merghoub T, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med 2014;371:2189-99.
2. Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 2015;348:124-8.
3. Goodman AM, Kato S, Bazhenova L, et al. Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. Mol Cancer Ther 2017;16:2598-608.
4. Tomasini P, Greillier L. Targeted next-generation sequencing to assess tumor mutation burden: ready for prime-time in non-small cell lung cancer? Transl Lung Cancer Res 2019;8:S323-6.
5. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. Science 2015;348:69-74.
6. Miller A, Asmann Y, Cattaneo L, et al. High somatic mutation and neoantigen burden are correlated with decreased progression-free survival in multiple myeloma. Blood Cancer J 2017;7:e612.
7. Zarogoulidis P, Papadopoulos V, Maragouli E, et al. Nivolumab as first-line treatment in non-small cell lung cancer patients-key factors: tumor mutation burden and PD-L1 ≥50. Transl Lung Cancer Res 2018;7:S28-30.
8. Hellmann MD, Nathanson T, Rizvi H, et al. Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer. Cancer Cell 2018;33:843-52.e4.
9. Frampton GM, Fichtenholtz A, Otto GA, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol 2013;31:1023-31.
10. Buchhalter I, Rempel E, Endris V, et al. Size matters: Dissecting key parameters for panel-based tumor mutational burden analysis. Int J Cancer 2019;144:848-58.
11. Campesato LF, Barroso-Sousa R, Jimenez L, et al. Comprehensive cancer-gene panels can be used to estimate mutational load and predict clinical benefit to PD-1 blockade in clinical practice. Oncotarget 2015;6:34221-7.
12. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med 2017;9:34.
13. Endris V, Penzel R, Warth A, et al. Molecular diagnostic

profiling of lung cancer specimens with a semiconductor-based massive parallel sequencing approach: feasibility, costs, and performance compared with conventional sequencing. J Mol Diagn 2013;15:765-75.
14. Goldstraw P, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. J Thorac Oncol 2016;11:39-51.
15. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.
16. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-303.
17. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012;22:568-76.
18. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research 2010;38:e164-e.
19. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 2012;6:80-92.
20. Newman AM, Bratman SV, Stehr H, et al. FACTERA: a practical method for the discovery of genomic rearrangements at breakpoint resolution. Bioinformatics 2014;30:3390-3.
21. Pirooznia M, Deng Y. SVM Classifier - a comprehensive java interface for support vector machine classification of microarray data. BMC Bioinformatics 2006;7 Suppl 4:S25.
22. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). The Annals of Statistics 2000;28:337-407.
23. Frank E, Hall M, Trigg L, et al. Data mining in bioinformatics using Weka. Bioinformatics 2004;20:2479-81.
24. Wang Z, Duan J, Cai S, et al. Assessment of Blood Tumor Mutational Burden as a Potential Biomarker for Immunotherapy in Patients With Non-Small Cell Lung Cancer With Use of a Next-Generation Sequencing Cancer Gene Panel. JAMA Oncol 2019;5:696-702.

Cite this article as: Tang Y, Li Y, Wang W, Lizaso A, Hou T, Jiang L, Huang M. Tumor mutation burden derived from small next generation sequencing targeted gene panel as an initial screening method. Transl Lung Cancer Res 2020;9(1):71-81. doi: 10.21037/tlcr.2019.12.27

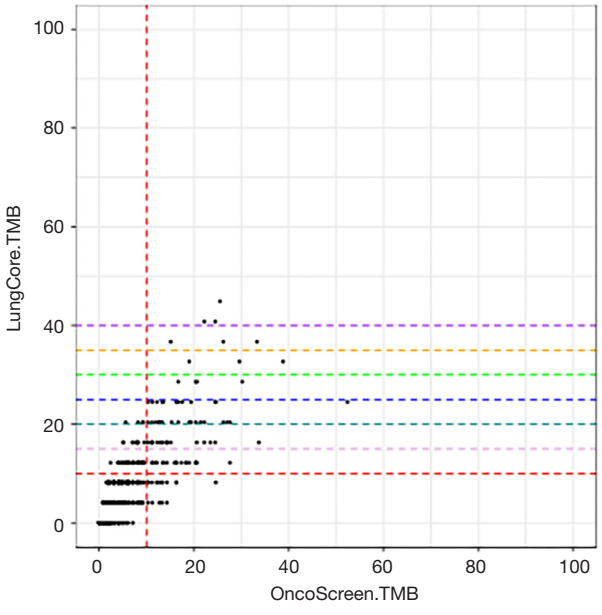

Figure S1 Scatter plots illustrating the derived TMB for the small panel and actual TMB from the 520-gene panel for 406 NSCLC patients using TMB cut-off of 10 mutations/Mb from the 520-gene panel. The X-axis denotes actual TMB derived from the large 520-gene panel. Y-axis denotes simulated TMB for the small panel. Dotted lines illustrate different cut-off points. Four quadrants clockwise from the upper left hand refer to false positives (FP), true positives (TP), false negatives (FN), and true negatives (TN). NSCLC, non-small cell lung cancer; TMB, tumor mutation burden.

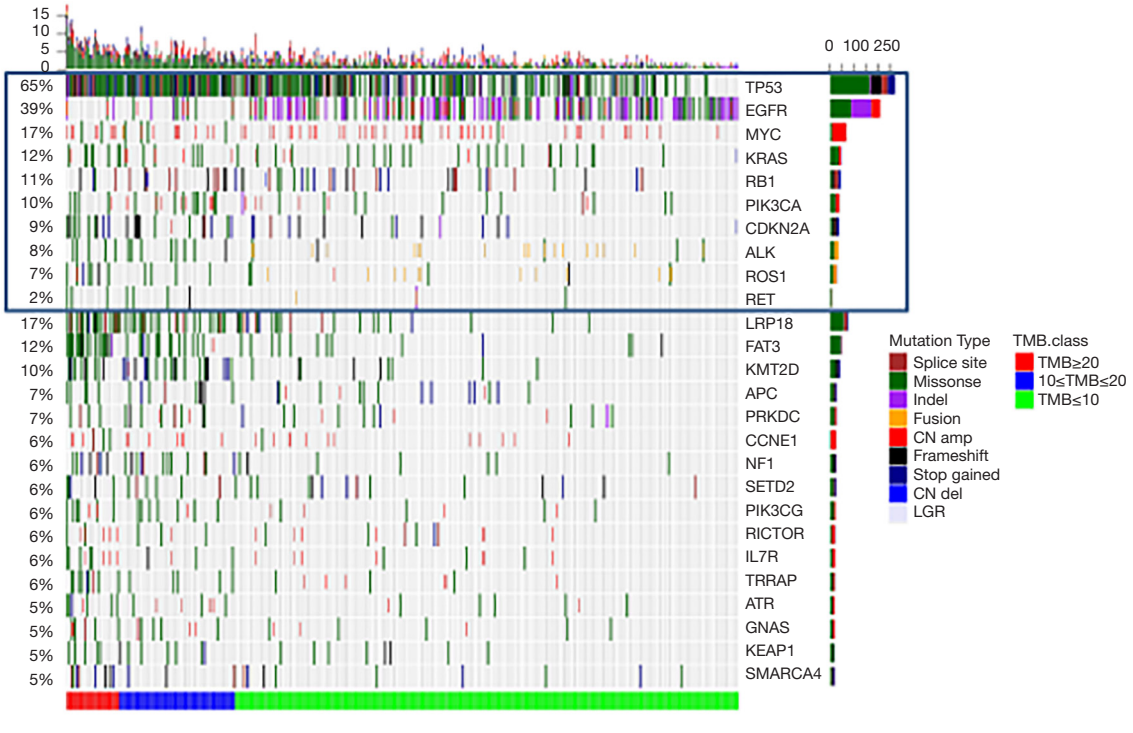

Figure S2 Mutational spectrum derived from a large 520-gene panel of the 406 NSCLC patients. The boxed area denotes the genes that are present in the small gene panel. Each column represents one patient. Each row represents a gene. The top bar denotes the number of mutations detected in each patient. Sidebar represents the number of patients with a mutation in a certain gene. Distinct colors represented mutation types. Patient data was arranged according to their TMB status, and are annotated at the bottom of the spectrum; wherein red denotes TMB ≥20 mutations/Mb (n=32), blue denotes TMB between 10–20 mutations/Mb (n=70) and green denotes TMB <10 mutations/Mb (n=306). NSCLC, non-small cell lung cancer; TMB, tumor mutation burden.

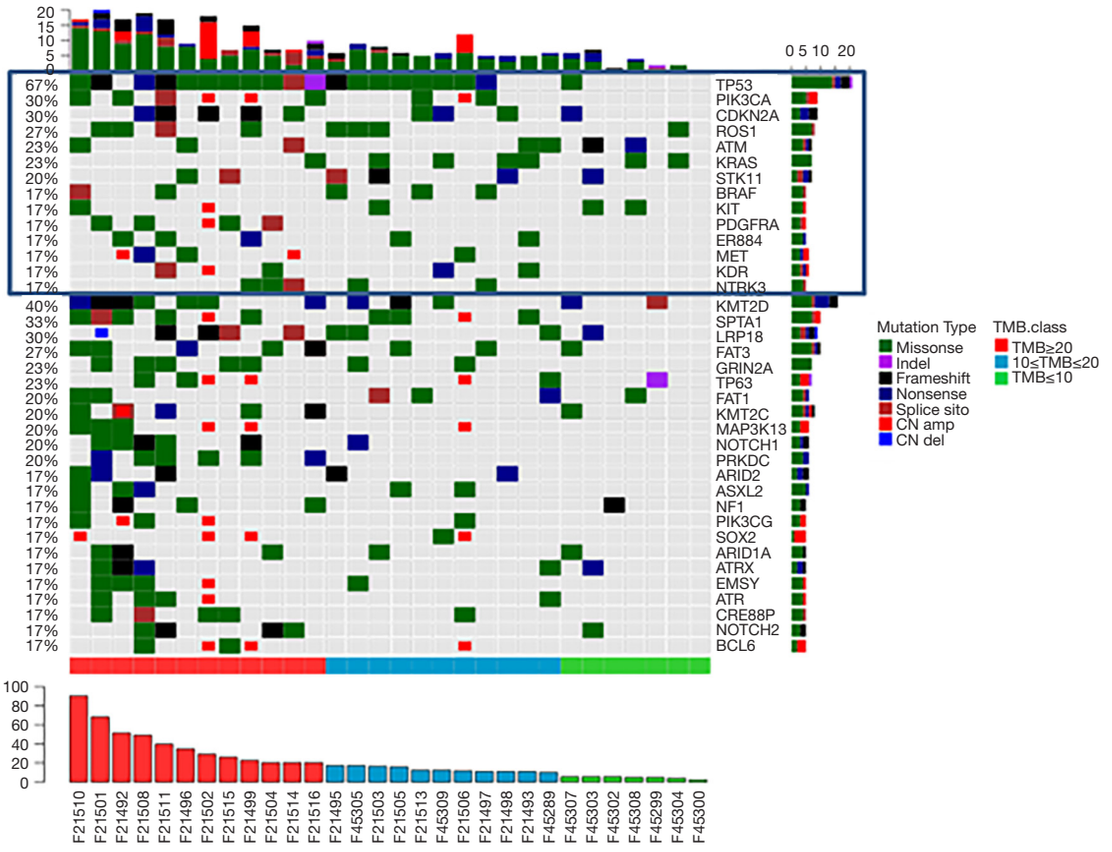

Figure S3 Mutational spectrum derived from the large 520-gene panel of the 30 NSCLC patients. The boxed area denotes the genes that are present in the small gene panel. Each column represents one patient. Each row represents a gene. The top bar denotes the number of mutations detected in each patient. Sidebar represents the number of patients with a mutation in a certain gene. Distinct colors represented mutation types. Patient data was arranged according to their TMB status, and are annotated at the bottom of the spectrum; wherein red denotes TMB ≥20 mutations/Mb (n=12), blue denotes TMB between 10–20 mutations/Mb (n=11) and green denotes TMB <10 mutations/Mb (n=7). The histogram below illustrates the actual TMB of each of the patients estimated with the 520-gene panel. NSCLC, non-small cell lung cancer; TMB, tumor mutation burden.

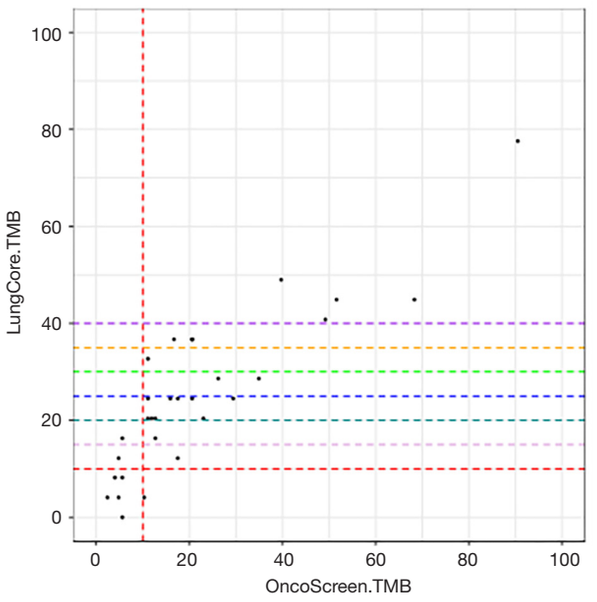

Figure S4 Scatter plots illustrating the actual TMB for the small panel and the 520-gene panel for 30 NSCLC patients using TMB cut-off of 10 mutations/Mb from the 520-gene panel. The X-axis denotes actual TMB derived from the large 520-gene panel. Y-axis denotes actual TMB for the small panel. Dotted lines illustrate different cut-off points. Four quadrants clockwise from the upper left hand refer to false positives (FP), true positives (TP), false negatives (FN), and true negatives (TN). NSCLC, non-small cell lung cancer; TMB, tumor mutation burden.

Table S1 The 520 cancer-related genes included in the OncoScreen Plus panel

ABL1, BCL2L1, CDKN2B, EGFL7, FAS, HIST1H3B, IRS1, MDM2, NOTCH4, PIK3R3, RAD54L, SMAD4, TMEM127ABL2, BCL2L11, CDKN2C, EGFR, FAT1, HIST1H3C, IRS2, MDM4, NPM1, PIM1, RAF1, SMARCA4, TMPR552ACVR1, BCL2L2, CEBPA, EIF1AX, FAT3, HIST1H3D, JAK1, MED12, NRAS, PLCG2, RANBP2, SMARCB1, TNFAIP3ACVR1B, BCL6, CENPA, EIF4A2, FBXW7, HIST1H3E, JAK2, MEF2B, NR4A3, PLK2, RARA, SMARCD1, TNFRSF14ADGRA2, BCOR, CHD1, EIF4E, FCGR2B, HIST1H3F, JAK3, MEN1, NRG1, PMAIP1, RASA1, SMO, TNFSF11AKT1, BCORL1, CHD2, ELOC, FGF10, HIST1H3G, JUN, MET, NSD1, PMS1, RB1, SNCAIP, TOP1AKT2, BCR, CHD4, EMSY, FGF12, HIST1H3H, KAT5A, MGA, NTHL1, PMS2, RBM10, SOCS1, TOP2AAKT3, BLM, CHEK1, EP300, FGF23, HIST1H3I, KDM5A, MITF, NTRK1, PNRC1, RECQL4, SOX2, TP53ALK, BMPR1A, CHEK2, EPCAM, FGF6, HIST1H3J, KDM5C, MLH1, NTRK2, POLD1, REL, SOX9, TRAF2ALOX12B, BRAF, CHUK, EPHA2, FGF7, HIST2H3C, KDM6A, MLH3, NTRK3, POLE, RET, SOX10, TRAF7AMER1, BRCA1, CIC, EPHA3, FGFR14, HIST2H3D, KDR, MPL, NUP93, POM121L12, RFWD2, SOX17, TRRAPANKRD11, BRCA2, CRBN, EPHA5, FGFR1, HIST3H3, KEAP1, MRE11A, PAK1, PPM1D, RHEB, SPEN, TSC1APC, BRD4, CREBBP, EPHA7, FGFR2, HLA-A, KEL, MSH2, PAK3, PPP2R1A, RHOA, SPOP, TSC2APCDD1, BRIP1, CRKL, EPHB1, FGFR3, HNF1A, KIT, MSH3, PAK7, PPP2R2A, RICTOR, SPTA1, TSHRAR, BTG1, CRLF2, ERBB2, FOXA1, HNF1B, KLF4, MSH6, PALB2, PPP6C, RIT1, SRC, U2AF1ARAF, BTK, CSF1R, ERBB3, FOXL2, HOXB13, KLHL6, MST1, PARK2, PRDM1, RNF43, SRSF2, VEGFAARFRP1, CALR, CSF3R, ERBB4, FRS2, HRAS, KMT2A, MST1R, PARP1, PREX2, ROS1, STAG2, VEGFBARID1A, CARD11, CTCF, ERBB5, FYN, HSD3B1, KMT2C, MTOR, PARP2, PRKAR1A, RPA1, STAT3, VEGFCARID1B, CASP8, CTLA4, ERCC1, GABRA6, HSP90AA1, KMT2D, MUTYH, PARP3, PRKC1, RPS6KA4, STAT4, VHLARID2, CBFB, CTNNA1, ERCC2, GATA4, ICOSLG, KRAS, MYC, PARP4, PRKDC, RPS6KB2, STAT5A, VTCN1ARID5B, CBL, CTNNB1, ERCC3, GATA6, ID3, LATS1, MYCL, PAX5, PRSS8, RPTOR, STAT5B, WISP3ASXL1, CCND1, CUL3, ERCC4, GID4, IDH1, LATS2, MYCN, PBRM1, PTCH1, RUNX1, STK11, WRNASXL2, CCND2, CUL4A, ERCC5, GNA13, IDH2, LMO1, MYD88, PDCD1, PTEN, RUNX1T1, STK40, WT1ATF1, CCND3, CUL4B, ERG, GPS2, IFNGR1, LRP1B, MYOD1, PDCD1LG2, PTK2, RYBP, SUFU, XIAPATM, CCNE1, CXCR4, ERRFI1, GREM1, IGF1, LYN, NBN, PDFRA, PTPN11, SDHA, SUZ12, XPO1ATR, CD274, CYCLD, ESR2, GRM3, IGF1R, LZTR1, NCOA3, PDGFRB, PTPRD, SDHAF2, SYK, XRCC2ATRX, CD276, CYP17A1, EWSR1, GSK3B, IGF2, MAG12, NCOR1, PDK1, PTPRS, SDHB, TACC3, XRCC3AURKA, CD79A, DAXX, EZH2, GSTM1, IKBKE, MALT1, NEB, PGR, PTPRT, SDHC, TAF1, YAP1AURKB, CD79B, DCUN1D1, FAM175A, GSTT1, IKZF1, MAP2K1, NEGR1, PHOX2B, QK1, SDHD, TBX3, YES1AXIN1, CDC73, DDR2, FAM46C, H3F3A, IL10, MAP2K2, NF1, PIK3CA, RAB35, SETD2, TCF3, ZBTB2AXIN2, CDH1, DICER1, FANCA, H3F3B, IL7R, MAP2K4, NF2, PIK3CB, RAC1, SF3B1, TCF7L2, ZFHX3AXL, CDK12, DIS3, FANCC, HDAC1, INHA, MAP3K1, NFE2L2, PIK3C2B, RAD21, SH2B3, TERC, ZNF217B2M, CDK4, DNAJB1, FANCD2, HDAC2, INHBA, MAP3K13, NFKB1A, PIK3C2G, RAD50, SH2D1A, TERT, ZNF703BACH1, CDK6, DNMT1, FANCE, HDAC4, INPP4A, MAP3K14, NKX2-1, PIK3C3, RAD51, SHQ1, TET1, ZNRF3BAP1, CDK8, DNMT3A, FANCF, HGF, INPP4B, MAP3K3, NKX3-1, PIK3CD, RAD51B, SLIT2, TET2, ZRSR2BARD1, CDKN1A, DNMT3B, FANCG, HIST1H1C, INSR, MAPK1, NOTCH1, PIK3CG, RAD51C, SLX4, TGFBR1, BBC3, CDKN1B, DOT1L, FANCI, HIST1H2BD, IRF2, MAX, NOTCH2, PIK3R1, RAD51D, SMAD2, TGFBR2, BCL2, CDKN1C, E2F3, FANCL, HIST1H3A, IRF4, MCL1, NOTCH3, PIK3R2, RAD52, SMAD3, TIPARP, BCL10, CDKN2A, EED, FANCM

Table S2 The 56 cancer-related genes included in the LungCore panel

AKT1, BRCA2, DDR2, FGF3, JAK1, MTOR, PDGFRA, ROS1, ALK, CCND1, DPYD, FGF4, JAK2, MYC, PIK3CA, SMO, ARAF, CDK4, EGFR, FGFR1, KDR, NRAS, PTCH1, STK11, ATM, CDK6, ERBB2, FGFR2, KIT, NRG1, PTEN, TP53, BCL2L11, CDKN2A, ERBB3, FGFR3, KRAS, NTRK1, RAF1, TSC1, BRAF, CTNNB1, ERBB4, FLT3, MAP2K1, NTRK2, RB1, TSC2, BRCA1, CYP2D6, FGF19, HRAS, MET, NTRK3, RET, UGT1A1

Table S3 Genes associated with high TMB using a cut-off of 10 mutations/Mb

Gene	TMB ≥10 count	TMB ≥10 frequency	TMB <10 count	TMB <10 frequency	P value	Inclusion in LungCore panel
*TP53**	97	0.951	165	0.543	2.65E−16	Yes
LRP1B	45	0.441	23	0.076	1.84E−15	No
FAT3	36	0.353	13	0.043	2.29E−14	No
KMT2D	30	0.294	9	0.03	8.59E−13	No
*CDKN2A**	23	0.225	15	0.049	1.10E−06	Yes
*PIK3CA**	22	0.216	18	0.059	2.36E−05	Yes
NF1	19	0.186	7	0.023	1.35E−07	No
*KRAS**	18	0.176	29	0.095	0.03222749	Yes
PIK3CG	15	0.147	8	0.026	3.30E−05	No
TRRAP	14	0.137	9	0.03	0.00019169	No
APC	14	0.137	16	0.053	0.00785903	No
PRKDC	14	0.137	14	0.046	0.00308381	No
ATR	14	0.137	7	0.023	4.38E−05	No
*EGFR**	13	0.127	146	0.48	4.33E−11	Yes
EPHA3	13	0.127	3	0.01	2.39E−06	No
NOTCH1	13	0.127	5	0.016	2.28E−05	No
GNAS	13	0.127	8	0.026	0.0002656	No
IL7R	13	0.127	10	0.033	0.00092812	No
CCNE1	13	0.127	13	0.043	0.00453923	No
MAP3K13	13	0.127	5	0.016	2.28E−05	No
NOTCH3	12	0.118	3	0.01	8.50E−06	No
KEAP1	12	0.118	8	0.026	0.00071711	No
GRIN2A	12	0.118	5	0.016	7.19E−05	No
KAT6A	12	0.118	7	0.023	0.00036383	No
SMARCA4	11	0.108	9	0.03	0.00327072	No
RUNX1T1	11	0.108	6	0.02	0.00049139	No
TET2	11	0.108	6	0.02	0.00049139	No
SETD2	11	0.108	14	0.046	0.03216951	No
BCL6	11	0.108	3	0.01	2.95E−05	No
ARID2	11	0.108	8	0.026	0.00186624	No
*PTEN**	11	0.108	8	0.026	0.00186624	Yes
EPHA5	11	0.108	4	0.013	8.72E−05	No
*ERBB4**	11	0.108	4	0.013	8.72E−05	Yes
RICTOR	10	0.098	13	0.043	0.04715735	No
*NTRK3**	10	0.098	6	0.02	0.00137068	Yes
KDM5A	10	0.098	5	0.016	0.00065146	No
KMT2A	10	0.098	8	0.026	0.00466477	No
EPHB1	10	0.098	3	0.01	9.97E−05	No
*BRCA2**	10	0.098	2	0.007	2.92E−05	Yes
AMER1	9	0.088	5	0.016	0.00186098	No
NFE2L2	9	0.088	10	0.033	0.03005436	No
FBXW7	9	0.088	3	0.01	0.00032797	No
*KDR**	9	0.088	4	0.013	0.00084159	Yes
CREBBP	8	0.078	7	0.023	0.02792485	No
NOTCH2	8	0.078	7	0.023	0.02792485	No
ZNF217	8	0.078	9	0.03	0.04450273	No
INHBA	8	0.078	9	0.03	0.04450273	No
*STK11**	8	0.078	9	0.03	0.04450273	Yes
PARP4	8	0.078	5	0.016	0.00510351	No
PDGFRB	8	0.078	5	0.016	0.00510351	No
*NTRK1**	8	0.078	6	0.02	0.00942689	Yes
*PTCH1**	8	0.078	6	0.02	0.00942689	Yes
KLHL6	8	0.078	9	0.03	0.04450273	No
PALB2	8	0.078	8	0.026	0.03417598	No
PIK3C2G	8	0.078	8	0.026	0.03417598	No
ASXL1	8	0.078	8	0.026	0.03417598	No
*ERBB3**	8	0.078	2	0.007	0.0003607	Yes
*FGF4**	8	0.078	6	0.02	0.00942689	Yes
GATA2	8	0.078	4	0.013	0.00248079	No
*KIT**	8	0.078	5	0.016	0.00510351	Yes
SOX2	8	0.078	6	0.02	0.00942689	No
*TSC2**	8	0.078	5	0.016	0.00510351	Yes
DOT1L	7	0.069	2	0.007	0.00121716	No
SRC	7	0.069	1	0.003	0.00034148	No
AKT2	7	0.069	4	0.013	0.00701304	No
*RET**	7	0.069	3	0.01	0.00321556	Yes
FANCM	7	0.069	4	0.013	0.00701304	No
FLT1	7	0.069	4	0.013	0.00701304	No
IGF1R	7	0.069	6	0.02	0.02298865	No
*FGF3**	7	0.069	4	0.013	0.00701304	Yes
PMS2	7	0.069	5	0.016	0.01335819	No
CDK12	7	0.069	3	0.01	0.00321556	No
FLT4	6	0.059	4	0.013	0.01888016	No
PPP2R1A	6	0.059	2	0.007	0.00397191	No
RAD50	6	0.059	5	0.016	0.0331207	No
MYCN	6	0.059	1	0.003	0.00124908	No
MSH2	6	0.059	3	0.01	0.0094814	No
KDM6A	6	0.059	1	0.003	0.00124908	No
*PDGFRA**	6	0.059	3	0.01	0.0094814	Yes
*BRCA1**	6	0.059	4	0.013	0.01888016	Yes
FANCI	6	0.059	4	0.013	0.01888016	No
*CDK6**	6	0.059	5	0.016	0.0331207	Yes
CTCF	6	0.059	1	0.003	0.00124908	No
FANCC	6	0.059	1	0.003	0.00124908	No
*DDR2**	5	0.049	3	0.01	0.02657788	Yes
FGF23	5	0.049	2	0.007	0.01244915	No
XPO1	5	0.049	3	0.01	0.02657788	No
FANCA	5	0.049	4	0.013	0.04794848	No
TOP1	5	0.049	1	0.003	0.0044491	No
*AKT1**	5	0.049	2	0.007	0.01244915	Yes
AXL	5	0.049	3	0.01	0.02657788	No
CSF1R	5	0.049	3	0.01	0.02657788	No
FANCL	5	0.049	1	0.003	0.0044491	No
NUP93	5	0.049	3	0.01	0.02657788	No
MUTYH	4	0.039	1	0.003	0.01533352	No
*TSC1**	4	0.039	2	0.007	0.03710234	Yes
BCOR	4	0.039	2	0.007	0.03710234	No
REL	4	0.039	1	0.003	0.01533352	No
CDC73	4	0.039	1	0.003	0.01533352	No
ERG	4	0.039	2	0.007	0.03710234	No
SYK	4	0.039	1	0.003	0.01533352	No
MEN1	4	0.039	2	0.007	0.03710234	No
*JAK1**	4	0.039	2	0.007	0.03710234	Yes
CEBPA	4	0.039	2	0.007	0.03710234	No
CHEK1	4	0.039	2	0.007	0.03710234	No
ESR1	4	0.039	2	0.007	0.03710234	No

*, cells are genes that are found in the small gene panel. TMB, tumor mutation burden.