



Integrating Histologic and Genomic Characteristics to Predict Tumor Mutation Burden of Early-Stage Non-Small-Cell Lung Cancer

Yuan Qiu^{1,2}, Liping Liu^{1,2,3}, Haihong Yang^{1,2}, Hanzhang Chen^{1,2}, Qihua Deng^{1,2}, Dakai Xiao^{1,2}, Yongping Lin^{1,2}, Changbin Zhu⁴, Weiwei Li⁴, Di Shao⁴, Wenxi Jiang⁴, Kui Wu^{4,5,6} and Jianxing He^{1,2*}

¹ National Clinical Research Center of Respiratory Disease, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, ² State Key Laboratory of Respiratory Diseases, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, ³ The Translational Medicine Laboratory, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, ⁴ BGI Genomics, BGI-Shenzhen, Shenzhen, China, ⁵ BGI-Shenzhen, Shenzhen, China, ⁶ China National GeneBank, BGI-Shenzhen, Shenzhen, China

OPEN ACCESS

Edited by:

Laura Mezquita,
Hospital Clínic de Barcelona, Spain

Reviewed by:

Alessandro Russo,
A.O. Papardo, Italy
Boris Duchemann,
Hôpital Avicenne, France

*Correspondence:

Jianxing He
hejx@vip.163.com

Specialty section:

This article was submitted to
Thoracic Oncology,
a section of the journal
Frontiers in Oncology

Received: 22 September 2020

Accepted: 29 December 2020

Published: 30 April 2021

Citation:

Qiu Y, Liu L, Yang H, Chen H, Deng Q, Xiao D, Lin Y, Zhu C, Li W, Shao D, Jiang W, Wu K and He J (2021) Integrating Histologic and Genomic Characteristics to Predict Tumor Mutation Burden of Early-Stage Non-Small-Cell Lung Cancer. *Front. Oncol.* 10:608989. doi: 10.3389/fonc.2020.608989

Tumor mutation burden (TMB) serves as an effective biomarker predicting efficacy of mono-immunotherapy for non-small cell lung cancer (NSCLC). Establishing a precise TMB predicting model is essential to select which populations are likely to respond to immunotherapy or prognosis and to maximize the benefits of treatment. In this study, available Formalin-fixed paraffin embedded tumor tissues were collected from 499 patients with NSCLC. Targeted sequencing of 636 cancer related genes was performed, and TMB was calculated. Distribution of TMB was significantly ($p < 0.001$) correlated with sex, clinical features (pathological/histological subtype, pathological stage, lymph node metastasis, and lympho-vascular invasion). It was also significantly ($p < 0.001$) associated with mutations in genes like *TP53*, *EGFR*, *PIK3CA*, *KRAS*, *EPHA3*, *TSHZ3*, *FAT3*, *NAV3*, *KEAP1*, *NFE2L2*, *PTPRD*, *LRRK2*, *STK11*, *NF1*, *KMT2D*, and *GRIN2A*. No significant correlations were found between TMB and age, neuro-invasion ($p = 0.125$), and tumor location ($p = 0.696$). Patients with *KRAS* p.G12 mutations and *FAT3* missense mutations were associated ($p < 0.001$) with TMB. *TP53* mutations also influence TMB distribution ($P < 0.001$). TMB was reversely related to *EGFR* mutations ($P < 0.001$) but did not differ by mutation types. According to multivariate logistic regression model, genomic parameters could effectively construct model predicting TMB, which may be improved by introducing clinical information. Our study demonstrates that genomic together with clinical features yielded a better reliable model predicting TMB-high status. A simplified model consisting of less than 20 genes and couples of clinical parameters were sought to be useful to provide TMB status with less cost and waiting time.

Keywords: early-stage non-small-cell lung cancer, tumor mutation burden (TMB), histology, genomics, model

INTRODUCTION

Immune checkpoint inhibitors (ICIs) targeting programmed death 1 (PD-1) and programmed death ligand 1 (PD-L1) achieved great success improving clinical outcomes of patients with advanced NSCLC. The efficacy of ICIs varies widely among individuals (1). Therefore, biomarkers stratifying patients who may benefit from ICI treatment are of great importance. Immunostaining of PD-L1 is considered as the first considered option. NSCLC patients with tumor proportion score (TPS) $\geq 1\%$ showed survival advantage from ICIs, especially, mono-immunotherapy. Tumor mutation burden (TMB) has been confirmed as a biomarker associating with efficacy of immunotherapy (2, 3). Meanwhile, in patients with resected NSCLC, TMB can help to evaluate long-term prognosis (4). Recent studies have shown that there are many factors affecting TMB distribution and ICIs (5, 6). TMB was negatively associated with clinical outcomes in metastatic *EGFR* mutant lung cancer patients treated with *EGFR*-TKI (7). *PIK3CA* amplification was significantly associated with TMB-H (8). Thus MSI-H/MMR-deficient tumors have much more somatic-mutations than MSS/MMR-proficient tumors (9), which have been demonstrated to have direct effects on TMB. Moreover, the molecular profile was associated with clinicopathological features and genetic ancestry markers of CRC patients (10). NSCLC tumors with elevated TMB and PD-L1 expression are associated with lympho-vascular invasion (11). It was also reported in patients with advanced gastric cancer that clinicopathological (lymph node metastasis) and molecular characteristics (*PIK3CA* mutations) are associated with responders to nivolumab (12).

TMB was precisely evaluated by whole-exon sequencing and could be predicted by a comprehensive genomic profiling (CGP) panel with a minimal size of 1 M. However, more turn-around time (TAT) would be taken when CGP is performed. Therefore, establishing a precise TMB predicting model is essential to monitor which populations are likely to respond to immunotherapy or prognosis and to maximize the benefits of treatment. In this article, we firstly aimed to select potential parameters by associating genetic and pathological characters with TMB distribution. An optimal TMB prediction model was constructed based on selected various clinical and genetic factors. Receiver operating curve analysis was applied to assess the performance of this prediction model.

MATERIALS AND METHODS

Patients

A total of 499 Formalin-Fixed, Paraffin Embedded tumor specimens of resected lung cancer were collected between March 2019 and September 2019. All patients signed the informed consent. Five hundred and eight cancer-related genes were sequenced.

Targeted Exome Capture Sequencing and Tumor Mutation Burden Assessment

Targeted exome capture sequencing data from 499 NSCLC samples were generated by MGI-500 platform. In detail,

genomic DNA (gDNA) was extracted from FFPE and peripheral blood samples using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) per protocol. DNA concentration and quality were assessed by Qubit (Life Technologies, Gaithersburg, MD, USA) and agarose gel electrophoresis. gDNA (250 ng) was used for sequencing library construction as previously described. The hybridization product was subsequently purified, amplified, and qualified. Finally, sequencing of 508 key cancer related genes was performed with a paired-end 100 bp and 8 bp barcode on a MGISEQ-2000 sequencer following the manufacturer's protocols.

Raw data was first filtered by SOAPnuke to exclude reads with low quality. The clean reads were then aligned to the reference human genome (UCSC hg19) using the BWA MEM algorithm. Single-nucleotide variants (SNVs) were detected by Genome Analysis Toolkit (GATK) Unified Genotyper. Small insertions and deletions (indels) were called using GATK Haplotype. Copy number variants (CNVs) were called using read-depth analysis. All the above variants were further filtered by quality depth, strand bias, mapping quality, and read position. Each variant was finally annotated with respect to gene location.

Targeted exome capture sequencing data of 499 NSCLC patients was analyzed in depth, and TMB was evaluated, which was defined as the total number of non-synonymous and indel somatic mutations present in a baseline tumor sample excluding known driver genes. TMB-high group was defined based on top 20% of TMB value.

Statistical Analysis

Chi-square test and Kruskal–Wallis test were used for comparing categorical and continuous variables, respectively. A P-value threshold of $p \leq 0.001$ (Chi-square test) and $p < 0.05$ (Kruskal–Wallis test) were used to define statistical significance. To determine the driver genes' differential between TMB-H and TMB-L, the Wilcoxon test was performed to figure out the significant driver genes. To determine the multivariable association of clinical and mutation characteristics with TMB-H, LASSO regression was used. In order to evaluate whether the histologic and genomic data could provide effective prediction of TMB, the receiver operating characteristic (ROC) curve analysis and area under the curve (AUC) were applied to evaluate the accuracy of TMB prediction model.

RESULTS

Sample Demographics and Clinical Characteristics

499 FFPE tissues were collected from patients with clinically diagnosed NSCLC, including 470 lung adenocarcinomas (LUADs) and 29 lung squamous cell carcinomas (LUSCs). Fifty-five percent of the samples were female ($n = 275$) (Table 1). The median age at diagnosis was 60 years (range, 29 to 85 years). More males than females were found in the TMB-high group ($P < 0.001$) (Table 1). More patients with LUSC were in the TMB-high group compared with those with LUAD ($P < 0.001$) (Table 1). 216 samples were located in the left lung, and

TABLE 1 | Baseline characteristics of included patients by TMB.

	TMB low (N = 400)	TMB high (N = 99)	Total (N = 499)	p value
Gender				<0.001
Female	251 (62.8%)	24 (24.2%)	275 (55.1%)	
Male	149 (37.2%)	75 (75.8%)	224 (44.9%)	
Age				0.003
<= 60	220 (55.0%)	38 (38.4%)	258 (51.7%)	
>60	180 (45.0%)	61 (61.6%)	241 (48.3%)	
Location				0.627
Left	171 (42.8%)	45 (45.5%)	216 (43.3%)	
Right	229 (57.2%)	54 (54.5%)	283 (56.7%)	
Histology				<0.001
N-Miss	2	0	2	
LUAD	391 (98.2%)	77 (77.8%)	468 (94.2%)	
LUSC	7 (1.8%)	22 (22.2%)	29 (5.8%)	
Pathological Subtype				<0.001
N-Miss	10	1	11	
<i>In situ</i>	18 (4.6%)	2 (2.0%)	20 (4.1%)	
invasive	279 (71.5%)	71 (72.4%)	350 (71.7%)	
Keratinizing	1 (0.3%)	13 (13.3%)	14 (2.9%)	
Micro-invasive	89 (22.8%)	4 (4.1%)	93 (19.1%)	
Non-keratinizing	3 (0.8%)	8 (8.2%)	11 (2.3%)	
Histological subtype				<0.001
N-Miss	115	31	146	
Leptic	87 (30.5%)	8 (11.8%)	95 (26.9%)	
Acinar	136 (47.7%)	33 (48.5%)	169 (47.9%)	
Papillary	36 (12.6%)	12 (17.6%)	48 (13.6%)	
Micropapillary	2 (0.7%)	3 (4.4%)	5 (1.4%)	
Solid	9 (3.2%)	9 (13.2%)	18 (5.1%)	
Mucinous	15 (5.3%)	3 (4.4%)	18 (5.1%)	
Lymph node				0.009
N-Miss	214	40	254	
Negative	171 (91.9%)	47 (79.7%)	218 (89.0%)	
Positive	15 (8.1%)	12 (20.3%)	27 (11.0%)	
Para-bronchial				<0.001
N-Miss	201	36	237	
Negative	188 (94.5%)	50 (79.4%)	238 (90.8%)	
Positive	11 (5.5%)	13 (20.6%)	24 (9.2%)	
LVI				0.001
N-Miss	100	15	115	
0	264 (88.0%)	62 (73.8%)	326 (84.9%)	
1	36 (12.0%)	22 (26.2%)	58 (15.1%)	
Stage				<0.001
IA1	165 (41.2%)	14 (14.1%)	179 (35.9%)	
IA2	112 (28.0%)	21 (21.2%)	133 (26.7%)	
IA3	50 (12.5%)	18 (18.2%)	68 (13.6%)	
IB	34 (8.5%)	11 (11.1%)	45 (9.0%)	
II	10 (2.5%)	19 (19.2%)	29 (5.8%)	
III	20 (5.0%)	13 (13.1%)	33 (6.6%)	
IV	9 (2.2%)	3 (3.0%)	12 (2.4%)	
T stage				<0.001
N-Miss	21	1	22	
T1	329 (86.8%)	59 (60.2%)	388 (81.3%)	
T2	44 (11.6%)	28 (28.6%)	72 (15.1%)	
T3	3 (0.8%)	7 (7.1%)	10 (2.1%)	
T4	3 (0.8%)	4 (4.1%)	7 (1.5%)	
N stage				<0.001
N-Miss	7	2	9	
N0	365 (92.9%)	75 (77.3%)	440 (89.8%)	
N1	8 (2.0%)	11 (11.3%)	19 (3.9%)	
N2	20 (5.1%)	11 (11.3%)	31 (6.3%)	
M stage				0.783
N-Miss	3	1	4	
M0	387 (97.5%)	96 (98.0%)	483 (97.6%)	
M1	10 (2.5%)	2 (2.0%)	12 (2.4%)	

the other 283 samples were in the right lung. In the left lung, 69 (13.9%) cases were located in the lower left lobe, while 29.1% of NSCLC were located in the upper lobe. The distribution of TMB is not significantly affected by tumor location (**Table 1**). In this study, most patients were stage IA (n = 360; 72.0%) and IB (n = 44; 8.9%). Seventy-three (14.4%) patients belonged to stage II/III/IV (7.6%). Distribution of T stage was as follows: T1 (n = 388, 81.3%), T2 (n = 72, 15.1%), T3 (n = 10, 2.1%), and T4 (n = 7, 1.5%). Most of them were N0 (n = 440, 89.8%) and M0 stage (n = 483, 97.6%) (**Table 1**). Clinical stage (IA), T1, and N0 are significantly related to higher level of TMB (**Table 1** and **Figures 1A–D**). Here, we also find a significant association between TMB distribution and clinicopathological features such as pathological subtype (P < 0.001) (**Figure 1E**), histological subtype (P < 0.001) (**Figure 1F**), para-bronchial lymph nodes (P < 0.001), lymph node metastasis (P = 0.009), and LVI (lympho-vascular invasion) (P < 0.001) (**Table 1**). But, TMB distribution is not significantly affected by neuro-invasive (p = 0.125) (**Table 1**).

Mutation Burden and Frequently Mutated Genes

Samples were divided into high (99) and low (400) TMB groups (**Table 1**) according to TMB top 20% in all histology (n = 6.15/Mb) (**Supplementary Figure 1**). In two groups, canonical driver mutations were found in *EGFR* (TMB-L: n = 246, TMB-H: n = 42), *KRAS* (TMB-L: n = 32, TMB-H: n = 18), *PIK3CA* (TMB-L: n = 13, TMB-H: n = 12), *BRAF* (TMB-L: n = 23, TMB-H: n = 5) and *TP53* (TMB-L: n = 80, TMB-H: n = 55) (**Supplementary Table 1** and **Figure 1**). There is no association between TMB distribution and driver gene mutational status (P = 0.27) (**Supplementary Figure 2**). Genes, differentially mutated between TMB-L and TMB-H patients (TMB-Low vs. TMB-High) were *EGFR* (62 vs 42%, P < 0.001), *EPHA3* (2 vs 13%, P < 0.001), *FAT3* (4 vs 20%, P < 0.001), *KEAP1* (1 vs 7%, P = 0.001), *KMT2D* (2 vs 10%, P < 0.001), *LRRK2* (1 vs 10%, P < 0.001), *NAV3* (1 vs 12%, P < 0.001), *NFI* (2 vs 10%, P < 0.001), *NFE2L2* (1 vs 11%, P < 0.001), *PIK3CA* (3.2 vs 12.1%, P < 0.001), *PTPRD* (1 vs 10%, P < 0.001), *STK11* (1 vs 8%, P < 0.001), *TP53* (20 vs 56%, P < 0.001), and *TSHZ3* (1 vs 13%, P < 0.001) (**Figure 2**, **Supplementary Table 1** and **Supplementary Figure 3**). At the same time, some gene mutations associated with immunotherapy resistance are not related to genetic mutations (**Supplementary Figure 4**).

Associations of four mutated genes with high frequency (*TP53*, *EGFR*, *KRAS*, and *FAT3*) and TMB were further investigated. 124 patients (24.85%) harbored *EGFR* L858R mutation, and 93 had *EGFR* exon 19 deletion (**Figure 3A**). No correlation of *EGFR* mutation status with TMB distribution was observed (P = 0.29) (**Figure 3A**). *TP53* mutations (missense, nonsense, and frameshift mutations) were significantly associated with TMB distribution (P < 0.001) (**Figure 3B**). TMB was also significantly affected by three mutation types in *PIK3CA* genes, including p.E542X, p.E545X, and p.Q546K (P < 0.05) (**Figure 3C**). There was a significant correlation between TMB distribution and *KRAS* p.G12X (P < 0.05) (**Figure 3D**).

Besides, missense and truncated *FAT3* mutations were significantly related to TMB (P < 0.05) (**Figure 3E**).

Constructing Tumor Mutation Burden Prediction Model

Based on the above clinical and genetic results, we hypothesized whether combination of clinical and genetic features could predict TMB status. Therefore, we trained a multivariable logistic regression model that included clinical parameters, age, histology, clinical stage (TNM) as well as genetic factors (*TP53*, *FAT3*, *APC*, *EPHA3*, *TERT*, *LRRK2*, *RB1*, *PTPRD*, *STK11*, and *NFI*).

Three factors (histology, Stage and *TP53*) were extremely powerful predictors for TMB through multivariate analysis (p < 0.001) (**Supplementary Table 1**). Other factors like *FAT3*, *APC*, *PTPRD* (P = 0.01), lymph-node metastasis, *EPHA3*, *TERT*, and *STK11* (P = 0.05) that have been found to be related to TMB distribution (**Supplementary Table 2**). Using TMB = 6.15 muts/Mb, the prediction model achieved a sensitivity of 73.8% and a specificity of 90.3%; the AUC (area under the ROC curve) was 0.899 (95% confidence interval, 0.861–0.938) indicating its potential for reliably identifying patients with greater TMB. After removing histological parameters, the AUC (area under the ROC curve) of these factors was 0.863 with a sensitivity of 76.3% and a specificity of 87.1% (**Figure 4**).

DISCUSSION

TMB, PD-1/L-1 expression are used to select patients who may benefit from immunotherapy (13). TMB is an emerging predictive marker of immune checkpoint blockade response (4) as well as prognosis for patients with NSCLC (14). In particular, it is important to accurately predict the benefit of immunotherapy based on TMB status. Also, it is reported that high TMB was associated with a better prognosis in patients with resected NSCLC (15). Our integrated histologic and genomic model is an important step toward addressing this unmet need.

In this analysis, we evaluated the association between TMB and clinical characteristics in patients with early-stage NSCLC. In the univariate analysis, despite being significantly associated with sex, our results further found that TMB was correlated with several clinicopathological features, like histological subtype, LVI, pathological subtype, para-bronchial lymph nodes, lymph node metastasis as well as tumor size. A recent study also explored the correlation of TMB and clinical characteristics in early-stage squamous cell lung carcinoma; however, no significant association was observed between TMB and age, gender, smoking history, stage (16). Another study assessed associations between clinical and TMB in resected NSCLC and identified that histological type, gender, and smoking status were associated with higher TMB (17). These inconsistent findings may be due to differences in ethnicity and pathologic types of the cohorts. In addition, the differences in panels used for TMB evaluation also have a significant impact on the results. TMB was initially detected using whole exon sequencing,

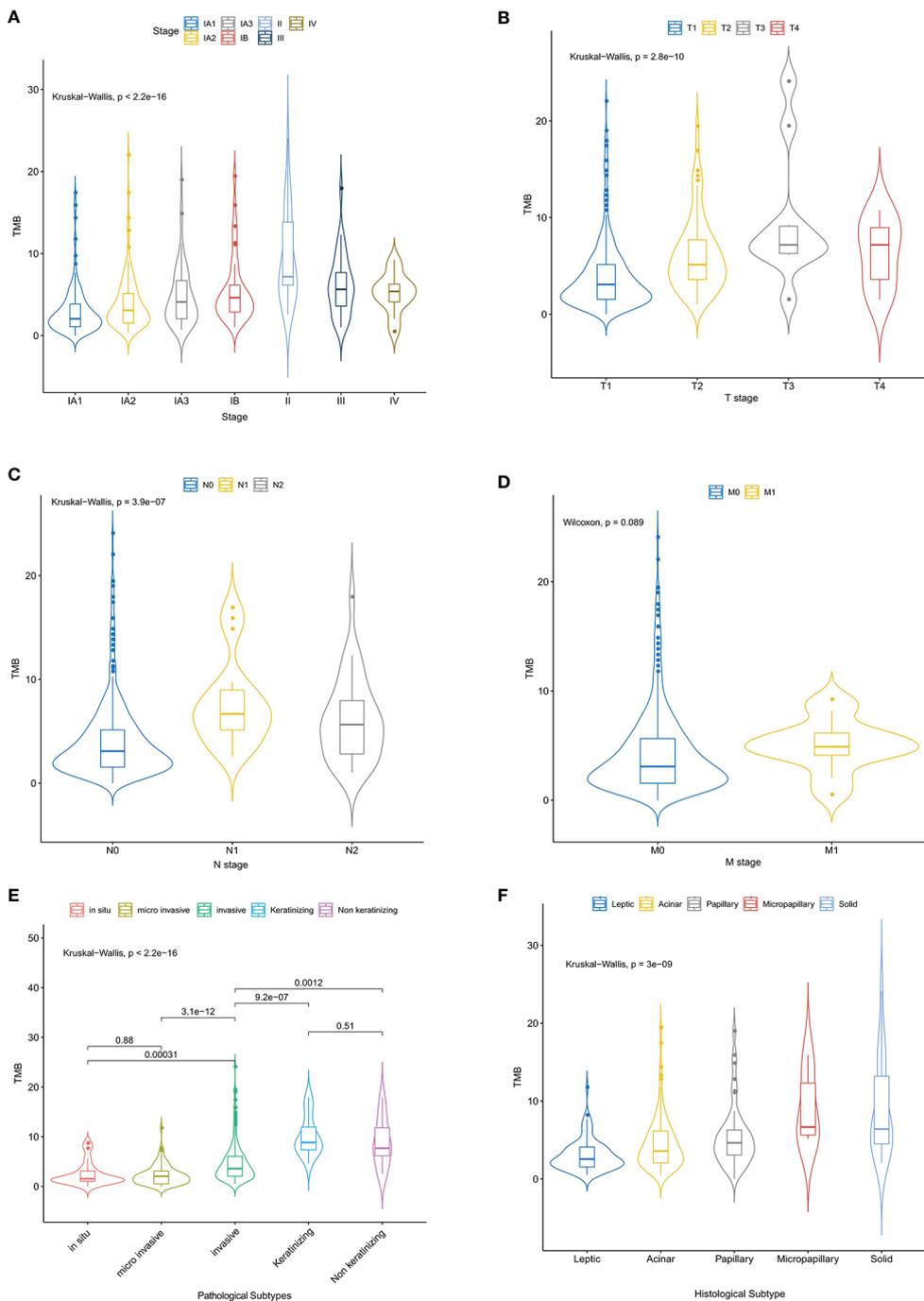
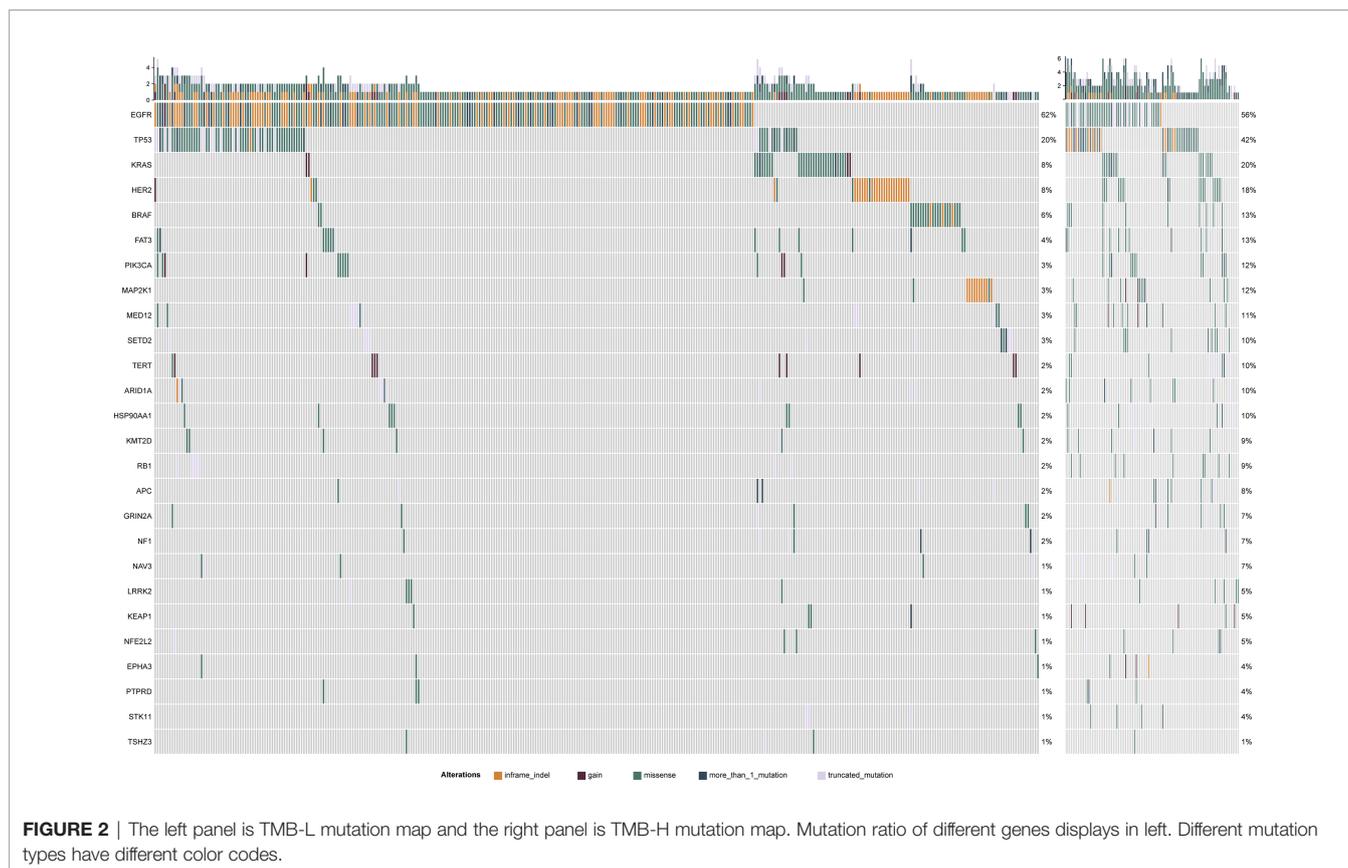


FIGURE 1 | The relationship between the TMB distribution and the tumor stages of NSCLC patients. Correlation analysis of TMB and TNM stage (A–D), pathological (E) and histological subtypes (F).

but a growing number of clinical trials are now using commercial panel sequencing to detect TMB. There is no uniform standard for TMB calculation method and threshold determination. Moreover, TMB varies greatly among different cancers and even different pathological subtypes. These are challenges that need to be overcome before further application of TMB (18–20).

In LUAD, carcinoma *in situ*, invasive and microinvasive cancers have different cell growth patterns and stages, which in turn affect the patient’s treatment and prognosis (21). In stage I LUADs, the micropapillary component was significantly associated with nodal micro-metastasis of tumor cells and may be a manifestation of aggressive behavior (22). TMB discrepancy



was observed among LUAD with various components. Remarkably, it is critical to determine the heterogeneity of LUAD components (histological subtype) by genetic profile. Solid predominant LUADs were more likely to harbor *KRAS* mutations than are other predominant subtypes (23). The solid predominant subtype of tumor has been found to correlate remarkably with an inflamed phenotype characterized by a high proportion of PD-L1/CD8+TILs and active cytotoxic immune profiling and that increased tumor immunogenicity from a high TMB (24). The alterations of *EGFR*, *KRAS*, and *BRAF* genes proved to be more frequent in micropapillary LUAD (24, 25). The studies have suggested that the molecular pathogenesis of micropapillary component may differ from other types of LUAD (26).

Pre-invasive LUAD displayed distinct mutation profiles. *In situ* and micro-invasive LUAD showed higher prevalence of driver mutations, for example, *EGFR* mutations and *ALK* fusion. Thus, compared with invasive LUAD, *in situ* and micro-invasive LUAD had lower TMB, which are concordant with variant distribution of driver gene mutations in these two histologic subtypes. LUSC (27) and LVI (28) were previously found to have higher TMB, which was similar to the results of our study. Beside, LVI has been linked to an increase in immune cell infiltration (28). Our result, together with other reported data may provide a TMB related immune activation hypothesis. Patients with different tumor stages exhibited distinct clinical behaviors (29). Para-bronchial lymph nodes is associated with a

poorer prognosis (21, 30). Thus, we inferred that these factors may affect immunogenicity through immune microenvironment and molecular profile. Neuro-invasive and stage-M did not affect TMB distribution, contrary to our expectation. The specific mechanism needs to be elucidated.

On the genomic level, the results showed that fifteen genes (*TP53*, *PIK3CA*, *KRAS*, *EPHA3*, *TSHZ3*, *FAT3*, *NAV3*, *KEAP1*, *NFE2L2*, *PTPRD*, *LRRK2*, *STK11*, *NF1*, *KMT2D*, *GRIN2A*) were significantly associated with TMB-H; *EGFR* was associated with TMB-L. Among these high-TMB-related genes, recent studies have shown that *TP53*-mutated tumors showed prominently increased somatic mutation burden compared with other mutant groups (*KRAS*, *EGFR*, *STK11*); and patients with *TP53* or *KRAS* mutations showed remarkable clinical benefit to PD-1 inhibitors (31, 32). *PIK3CA* and *KRAS* are mainly involved in the PI3K signaling pathway, which is one of the most important signal transduction pathways in the development of LUADs (33). In particular, four activated mutations of *PIK3CA* (p.E542X, p.E545X and p.Q546K) were found to have significant effect on TMB-H. *PIK3CA* gene mutations in the helical domain were correlated with TMB-H and poor prognosis in metastatic breast carcinomas with late-line therapies (34). Studies in lung cancer suggested that *PIK3CA* amplification was associated with higher TMB (8). Moreover, *KRAS* G12 mutations also correlated with high TMB group. As reported, NSCLC patients with *KRAS* G12 mutations showed an increased proportion of PD-L1+/CD8+ TILs (35). These genes stimulated PTEN/*PIK3CA*/AKT pathway,

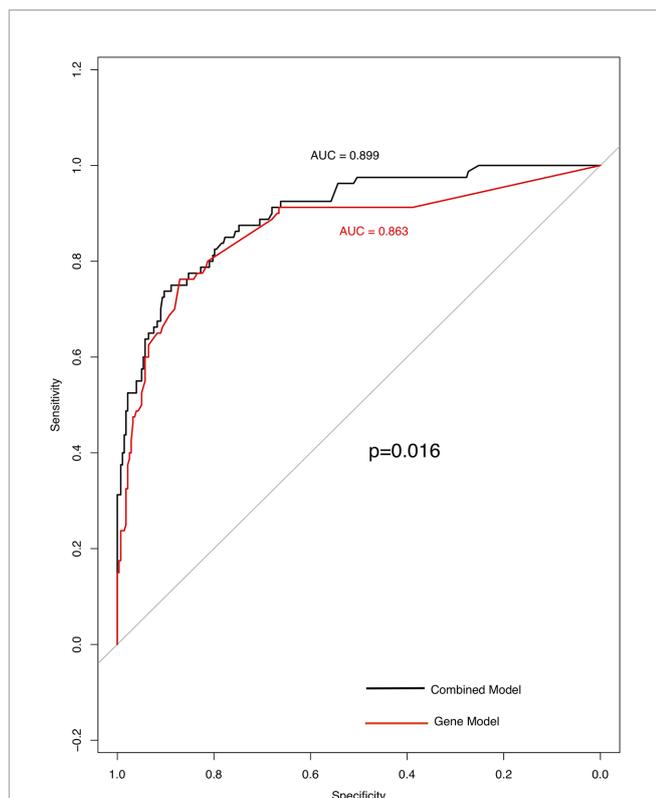


FIGURE 4 | High specificity of genes and clinical model for predicting TMB status. ROC curve analysis was used to determine the sensitivity and specificity of the two models. The black curve is the combined model; the area under the ROC curve is 0.899(95% confidence interval: 0.861–0.938). Curve in red is gene model; the area under the ROC curve is 0.863 (95% confidence interval: 0.811–0.916).

>Thus, we trained a multivariable logistic regression model predicting TMB category with 6.15 mutations/Mb as the cut-off value utilizing five clinical features (age, histology, T, N, M) and 10 genes (*TP53*, *FAT3*, *APC*, *EPHA3*, *TERT*, *LRRK2*, *RB1*, *PTPRD*, *STK11* and *NF1*). By comparing sensitivity and specificity, the results of two predictive model for TMB (histologic + genomics, genomics) confirmed that histologic features made a strong contribution to the integrated model for TMB prediction. There is also a small sample study which found that integrating multiple factors helps accurate prediction of TMB (37) although the ROC curves of the two studied models are close (0.89). Our study selected fewer histologic parameters without involving radiologic parameters. Another study showed that 56-gene panel could be used as a screening method for patients with low TMB. Compared with the panel, our prediction model has fewer genetic parameters, but achieved comparative efficiency (41). Therefore, this model may be better used to screen TMB-L or TMB-H status in early-stage NSCLC patients.

CONCLUSION

Overall, comprehensive clinical and genomic information can effectively evaluate TMB-high or low status. Our results showed

that an integrated prediction model combining histology and genomic parameters significantly improved the accuracy of TMB prediction. However, whether this integrated model plays a key role in predicting the clinical outcome to immunotherapy and prognosis, still needs further investigation.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number can be found below: China National GeneBank DataBase (CNGBdb) <https://db.cngb.org/>. Accession number CNP0001479.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of the First Affiliated Hospital of Guangzhou Medical University. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by Institutional Review Board of the First Affiliated Hospital of Guangzhou Medical University.

AUTHOR CONTRIBUTIONS

JH, YQ, CZ, and LL conceptualized and designed the study. KW provided administrative support. YY and HC provided the study materials or patients. QD, YL, and WJ collected and assembled the data. WL and DS analyzed and interpreted the data. All authors contributed to the article and approved the submitted version.

FUNDING

This study is funded by the Foundation and Applied Basic Research Fund of Guangdong Province (02020A1515011293) and the National Natural Science Foundation of China (Grant No. 81772486).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2020.608989/full#supplementary-material>

Supplementary Figure 1 | The distribution of TMB.

Supplementary Figure 2 | Driver mutation status and association with TMB.

Supplementary Figure 3 | The distribution of high-frequency genes in the TMB-H and TMB-L groups.

Supplementary Figure 4 | ICI related genes and association with TMB.

Supplementary Table 1 | Correlation between TMB and high frequently mutated genes.

Supplementary Table 2 | Gene and combined model for predicting TMB in NSCLC.

REFERENCES

- Yu Y, Fan X, Wei Y, Liu Z, Lian Z, Han H, et al. Correlation analysis between PD-L1 expression, TMB and clinical characteristics in Chinese non-small cell lung cancer. *AAO (2019)* 79(13 Suppl):Abstract nr 4056. doi: 10.1158/1538-7445.AM2019-4056
- Wang Z, Duan J, Cai S, Han M, Dong H, Zhao J, 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(5):696–702. doi: 10.1001/jamaoncol.2018.7098
- Marabelle A, Fakih M, Lopez J, Shah M, Shapira-Frommer R, Nakagawa K, et al. Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. *Lancet Oncol* (2020) 21(10):1353–65. doi: 10.1016/S1470-2045(20)30445-9
- Devarakonda S, Rotolo F, Tsao M-S, Lanc I, Brambilla E, Masood A, et al. Tumor mutation burden as a biomarker in resected non-small-cell lung cancer. *J Clin Oncol* (2018) 36(30):2995. doi: 10.1200/JCO.2018.78.1963
- Kim JH, Kim HS, Kim BJ. Prognostic value of KRAS mutation in advanced non-small-cell lung cancer treated with immune checkpoint inhibitors: a meta-analysis and review. *Oncotarget* (2017) 8(29):48248. doi: 10.18632/oncotarget.17594
- Cheng ML, Oxnard GR. Does TMB impact the effectiveness of TKIs in EGFR-mutant NSCLC? *Clin Cancer Res* (2019) 25(3):899–900. doi: 10.1158/1078-0432.CCR-18-2368
- Offin M, Rizvi H, Tenet M, Ni A, Sanchez-Vega F, Li BT, et al. Tumor Mutation Burden and Efficacy of EGFR-Tyrosine Kinase Inhibitors in Patients with EGFR-Mutant Lung Cancers. *Clin Cancer Res* (2019) 25(3):1063–9. doi: 10.1158/1078-0432.CCR-18-1102
- Jiang T, Shi J, Dong Z, Hou L, Zhao C, Li X, et al. Genomic landscape and its correlations with tumor mutational burden, PD-L1 expression, and immune cells infiltration in Chinese lung squamous cell carcinoma. *J Hematol Oncol* (2019) 12(1):75. doi: 10.1186/s13045-019-0762-1
- Song W, Shen L, Wang Y, Liu Q, Goodwin TJ, Li J, et al. Synergistic and low adverse effect cancer immunotherapy by immunogenic chemotherapy and locally expressed PD-L1 trap. *Nat Commun* (2018) 9(1):2237. doi: 10.1038/s41467-018-04605-x
- Dos Santos W, Sobanski T, de Carvalho AC, Evangelista AF, Matsushita M, Berardinelli GN, et al. Mutation profiling of cancer drivers in Brazilian colorectal cancer. *Sci Rep* (2019) 9(1):13687. doi: 10.1038/s41598-019-49611-1
- Mitchell KG, Negrao MV, Parra ER, Li J, Zhang J, Dejima H, et al. Lymphovascular Invasion Is Associated With Mutational Burden and PD-L1 in Resected Lung Cancer. *Ann Thorac Surg* (2020) 109(2):358–66. doi: 10.1016/j.athoracsur.2019.08.029
- Mishima S, Kawazoe A, Nakamura Y, Sasaki A, Kotani D, Kuboki Y, et al. Clinicopathological and molecular features of responders to nivolumab for patients with advanced gastric cancer. *J Immunother Cancer* (2019) 7(1):24. doi: 10.1186/s40425-019-0514-3
- Cao D, Xu H, Xu X, Guo T, Ge W. High tumor mutation burden predicts better efficacy of immunotherapy: a pooled analysis of 103078 cancer patients. *Oncimmunology* (2019) 8(9):e1629258. doi: 10.1080/2162402X.2019.1629258
- Owada-Ozaki Y, Muto S, Takagi H, Inoue T, Watanabe Y, Fukuhara M, et al. Prognostic impact of tumor mutation burden in patients with completely resected non-small cell lung cancer: brief report. *J Thoracic Oncol* (2018) 13(8):1217–21. doi: 10.1016/j.jtho.2018.04.003
- Devarakonda S, Rotolo F, Tsao MS, Lanc I, Brambilla E, Masood A, et al. Tumor Mutation Burden as a Biomarker in Resected Non-Small-Cell Lung Cancer. *J Clin Oncol* (2018) 36(30):2995–3006. doi: 10.1200/JCO.2018.78.1963
- Yu H, Chen Z, Ballman KV, Watson MA, Govindan R, Lanc I, et al. Correlation of PD-L1 Expression with Tumor Mutation Burden and Gene Signatures for Prognosis in Early-Stage Squamous Cell Lung Carcinoma. *J Thorac Oncol* (2019) 14(1):25–36. doi: 10.1016/j.jtho.2018.09.006
- Ono A, Terada Y, Kawata T, Serizawa M, Isaka M, Kawabata T, et al. Assessment of associations between clinical and immune microenvironmental factors and tumor mutation burden in resected nonsmall cell lung cancer by applying machine learning to whole-slide images. *Cancer Med* (2020) 9(13):4864–75. doi: 10.1002/cam4.3107
- Sholl LM, Hirsch FR, Hwang D, Botling J, Lopez-Rios F, Bubendorf L, et al. The Promises and Challenges of Tumor Mutation Burden as an Immunotherapy Biomarker: A Perspective from the International Association for the Study of Lung Cancer Pathology Committee. *J Thorac Oncol* (2020) 15(9):1409–24. doi: 10.1016/j.jtho.2020.05.019
- Heeke S, Benzaquen J, Hofman V, Long-Mira E, Lespinet V, Bordone O, et al. Comparison of Three Sequencing Panels Used for the Assessment of Tumor Mutational Burden in NSCLC Reveals Low Comparability. *J Thorac Oncol* (2020) 15(9):1535–40. doi: 10.1016/j.jtho.2020.05.013
- Budczies J, Allgauer M, Litchfield K, Rempel E, Christopoulos P, Kazdal D, et al. Optimizing panel-based tumor mutational burden (TMB) measurement. *Ann Oncol* (2019) 30(9):1496–506. doi: 10.1093/annonc/mdz205
- Kay FU, Kandathil A, Batra K, Saboo SS, Abbara S, Rajiah P. Revisions to the Tumor, Node, Metastasis staging of lung cancer: Rationale, radiologic findings and clinical implications. *World J Radiol* (2017) 9(6):269. doi: 10.4329/wjr.v9.i6.269
- Roh MS, Lee JI, Choi PJ, Hong YS. Relationship between micropapillary component and micrometastasis in the regional lymph nodes of patients with stage I lung adenocarcinoma. *Histopathology* (2004) 45(6):580–6. doi: 10.1111/j.1365-2559.2004.01953.x
- Kadota K, Yeh Y-C, D'Angelo SP, Moreira AL, Kuk D, Sima CS, et al. Associations Between Mutations and Histologic Patterns of Mucin in Lung Adenocarcinoma: Invasive Mucinous Pattern and Extracellular Mucin Are Associated With KRAS Mutation. *Am J Surg Pathol* (2014) 38(8):1118–27. doi: 10.1097/PAS.0000000000000246
- De Oliveira Duarte Achcar R, Nikiforova MN, Yousem SA. Micropapillary lung adenocarcinoma: EGFR, K-ras, and BRAF mutational profile. *Am J Clin Pathol* (2009) 131(5):694–700. doi: 10.1309/AJCPBS85VJE0BPDO
- Cai YR, Dong YJ, Wu HB, Liu ZC, Zhou LJ, Su D, et al. Micropapillary: A component more likely to harbour heterogeneous EGFR mutations in lung adenocarcinomas. *Sci Rep* (2016) 6:23755. doi: 10.1038/srep23755
- Zang Y-S, Jiao X-D, Zhang X-C, Qin B, Liu D, Liu L, et al. Tumour mutation burden analysis in a 5660-cancer-patient cohort reveals cancer type-specific mechanisms for high mutation burden. *Ann Oncol* (2019) 30:ix125. doi: 10.1093/annonc/mdz431.007
- Mitchell KG, Negrao MV, Parra ER, Li J, Zhang J, et al. Lymphovascular Invasion Is Associated With Mutational Burden and PD-L1 in Resected Lung Cancer. *Ann Thoracic Surgery* (2020) 109(2):358–66. doi: 10.1016/j.athoracsur.2019.08.029
- Ettinger DS, Wood DE, Aggarwal C, Aisner DL, Akerley W, Bauman JR, et al. NCCN Guidelines Insights: Non-Small Cell Lung Cancer, Version 1.2020: Featured Updates to the NCCN Guidelines. *J Natl Compr Cancer Netw* (2019) 17(12):1464–72. doi: 10.6004/jnccn.2019.0059
- Zhang Y-K, Tan L-L, Wang Z-y, Chen Z-j, Le H-B, Zhu W-Y. Association of lymph node involvement with the prognosis of pathological T1 invasive non-small cell lung cancer. *World J Surg Oncol* (2017) 15(1):64. doi: 10.1186/s12957-017-1098-3
- Tang Y, Li J, Xie N, Yang X, Liu L, Wu H, et al. PIK3CA gene mutations in the helical domain correlate with high tumor mutation burden and poor prognosis in metastatic breast carcinomas with late-line therapies. *Aging (Albany NY)* (2020) 12(2):1577–90. doi: 10.18632/aging.102701
- Liu C, Zheng S, Jin R, Wang X, Wang F, Zang R, et al. The superior efficacy of anti-PD-1/PD-L1 immunotherapy in KRAS-mutant non-small cell lung cancer that correlates with an inflammatory phenotype and increased immunogenicity. *Cancer Lett* (2020) 470:95–105. doi: 10.1016/j.canlet.2019.10.027
- Schrock AB, Li SD, Frampton GM, Suh J, Braun E, Mehra R, et al. Pulmonary Sarcomatoid Carcinomas Commonly Harbor Either Potentially Targetable Genomic Alterations or High Tumor Mutational Burden as Observed by Comprehensive Genomic Profiling. *J Thorac Oncol* (2017) 12(6):932–42. doi: 10.1016/j.jtho.2017.03.005
- Ahmed SM, Luo L, Namani A, Wang XJ, Tang X. Nrf2 signaling pathway: Pivotal roles in inflammation. *Biochim Biophys Acta Mol Basis Dis* (2017) 1863(2):585–97. doi: 10.1016/j.bbdis.2016.11.005
- Herroon MK, Rajagurubandara E, Diedrich JD, Heath EI, Podgorski I. Adipocyte-activated oxidative and ER stress pathways promote tumor

- survival in bone via upregulation of Heme Oxygenase 1 and Survivin. *Sci Rep* (2018) 8(1):40. doi: 10.1038/s41598-017-17800-5
35. Hajdu T, Juhasz T, Szucs-Somogyi C, Racz K, Zakany R. NR1 and NR3B Composed Intranuclear N-methyl-d-aspartate Receptor Complexes in Human Melanoma Cells. *Int J Mol Sci* (2018) 19(7):1929. doi: 10.3390/ijms19071929
36. Hastings K, Yu HA, Wei W, Sanchez-Vega F, DeVaux M, Choi J, et al. EGFR mutation subtypes and response to immune checkpoint blockade treatment in non-small-cell lung cancer. *Ann Oncol* (2019) 30(8):1311–20. doi: 10.1093/annonc/mdz141
37. Zhang N, Wu J, Yu J, Zhu H, Yang M, Li R, et al. Histologic, and Genetic Features to Predict Tumor Mutation Burden of Non-Small-Cell Lung Cancer. *Clin Lung Cancer* (2019) 105(1 Suppl):Abstract 3243. doi: 10.1016/j.ijrobp.2019.06.2460
38. Marinelli D, Mazzotta M, Scalerà S, Terrenato I, Sperati F, D'Ambrosio L, et al. KEAP1-driven co-mutations in lung adenocarcinoma unresponsive to immunotherapy despite high tumor mutational burden. *Ann Oncol* (2020) 31(12):1746–54. doi: 10.1016/j.annonc.2020.08.2105
39. Owada Y, Yamaura T, Okabe N, Matsumura Y, Higuchi M, Tanaka D, et al. Comprehensive gene analysis including next generation sequencing (NGS) study with immunological parameters to obtain the biomarkers for immune checkpoint inhibitors (ICIs) in non-small cell lung cancer (NSCLC). *J Clin Oncol* (2016) 34(15_suppl):e20040–e. doi: 10.1200/JCO.2016.34.15_suppl.e20040
40. Lamberti G, Spurr LF, Li Y, Ricciuti B, Recondo G, Umeton R, et al. Clinicopathological and genomic correlates of programmed cell death ligand 1 (PD-L1) expression in nonsquamous non-small-cell lung cancer. *Ann Oncol* (2020) 31(6):807–14. doi: 10.1016/j.annonc.2020.02.017
41. Tang Y, Li Y, Wang W, Lizaso A, Hou T, Jiang L, et al. 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

Conflict of Interest: CZ, WL, DS, KW and WJ are employees of BGI Genomics that produces the panel test used in this study.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Qiu, Liu, Yang, Chen, Deng, Xiao, Lin, Zhu, Li, Shao, Jiang, Wu and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.