Research Article

Integrative Modeling of Multiomics Data for Predicting Tumor Mutation Burden in Patients with Lung Cancer

Jun Wang [ID],[1] **Peng Chen,**[1] **Mingyang Su,**[1] **Guocheng Zhong,**[2] **Shasha Zhang,**[1] **and Deming Gou** [ID][1]

[1]*Shenzhen Key Laboratory of Microbial Genetic Engineering, Vascular Disease Research Center, College of Life Sciences and Oceanography, Guangdong Provincial Key Laboratory of Regional Immunity and Diseases, Carson International Cancer Center, Shenzhen University, Nanhai Ave 3688, Shenzhen, 518060 Guangdong, China*
[2]*Department of Hematology and Oncology, Shenzhen University General Hospital, Shenzhen University Clinical Medical Academy, Xueyuan Ave 1098, Shenzhen, 518055 Guangdong, China*

Correspondence should be addressed to Deming Gou; dmgou@szu.edu.cn

Immunotherapy has been widely used in the treatment of lung cancer, and one of the most effective biomarkers for the prognosis of immunotherapy currently is tumor mutation burden (TMB). Although whole-exome sequencing (WES) could be utilized to assess TMB, several problems prevent its routine clinical application. To develop a simplified TMB prediction model, patients with lung adenocarcinoma (LUAD) in The Cancer Genome Atlas (TCGA) were randomly split into training and validation cohorts and categorized into the TMB-high (TMB-H) and TMB-low (TMB-L) groups, respectively. Based on the 610 differentially expressed genes, 50 differentially expressed miRNAs and 58 differentially methylated CpG sites between TMB-H and TMB-L patients, we constructed 4 predictive signatures and established TMB prediction model through machine learning methods that integrating the expression or methylation profiles of 7 genes, 7 miRNAs, and 6 CpG sites. The multiomics model exhibited excellent performance in predicting TMB with the area under curve (AUC) of 0.911 in the training cohort and 0.859 in the validation cohort. Besides, the significant correlation between the multiomics model score and TMB was observed. In summary, we developed a prognostic TMB prediction model by integrating multiomics data in patients with LUAD, which might facilitate the further development of quantitative real time-polymerase chain reaction- (qRT-PCR-) based TMB prediction assay.

1. Introduction

Lung cancer is one of the most common malignancies worldwide, and it is the first leading cause of tumor-related mortality with an increasing incidence in recent years [1]. It was reported that 2.1 million new cases of lung cancer were diagnosed around the world in 2018, which accounted for 11.6% of all new cancer patients [2, 3]. Despite the improvements in chemotherapy and targeted therapy, the 5-year overall survival (OS) for patients with lung cancer remained poor [1, 4]. Nevertheless, immunotherapy, especially the application of immune checkpoint inhibitors (ICIs), had made a great breakthrough in the treatment of cancer and dramatically increased survival rate and quality of life for patients with lung cancer [5–9].

As the most successful representative of immunotherapy, programmed cell death-1/programmed cell death ligand-1 (PD-1/PD-L1) inhibitor had shown better performance over conventional chemotherapy in terms of OS, response rate, and progression-free survival (PFS) for the treatment of lung cancer [10, 11]. Furthermore, a large amount of clinical research had also demonstrated that immunotherapy alone or in combination with chemotherapy could be used for the first-line treatment of patients with metastatic lung cancer [12–15]. It was reported that patients with higher PD-L1 expression had better outcomes compared to patients with

lower or no PD-L1 expression using anti-PD-L1 antibody clone 22C3 [16]. Unfortunately, only 10%-20% of non-small-cell lung cancer (NSCLC) patients have considerable curative effects, and most patients cannot benefit from immunotherapy [17–19]; therefore, biomarkers are urgently needed to rationalize the utilization of immunotherapy for patients.

Tumor mutation burden (TMB) emerged recently as a reliable biomarker that significantly correlated with immunotherapy efficacy across a wide spectrum of tumor types. TMB is defined as the number of somatic mutations per megabase (Mb) of the genome examined. Previous studies found that higher TMB was associated with improved objective response, durable clinical benefit, and PFS in NSCLC patients under immunotherapy [20]. It had been reported that PFS among stage IV patients with high TMB was significantly longer with PD-1/PD-L1 plus cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) treatment than with chemotherapy [21]. Moreover, through analyzing 7,033 patients with different types of cancer, TMB was found to be a useful biomarker for predicting response of ICIs across different types of cancer, and higher TMB (highest 20% in each histology) was associated with better OS [22].

Whole-exome sequencing (WES) is considered the gold standard for evaluating TMB, but it is time-consuming and carries high cost [23]. Thus, targeted next generation sequencing (NGS) has been adopted as an alternative approach for predicting TMB. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) (468 genes) and FoundationOne companion diagnostic (CDx) (324 genes) are two extensively utilized targeted NGS methods, and both of them have been approved by the U.S. Food and Drug Administration (FDA) for clinical application. Despite the fact that targeted NGS is effective in predicting TMB, various problems arise for its routine clinical application, such as the limit of detection, germline mutation exclusion, and standard cutoff threshold determination [24]. Moreover, targeted NGS is still time-consuming and expensive compared with other clinical molecular tests [24]. In an effort to establish a simplified, cost-effective approach to predict TMB in patients with lung adenocarcinoma (LUAD), we intended to integrate a multiomics data to develop a predicting model for TMB.

In this study (Figure 1), patients with lung adenocarcinoma (LUAD) in The Cancer Genome Atlas (TCGA) were first split into training and validation cohorts. Then patients in training cohort were divided as TMB-high (TMB-H) and TMB-low (TMB-L), and the differentially expressed genes, miRNAs, and differentially methylated CpG sites were identified. Subsequently, a multiomics TMB prediction model (TPM) involving expression profiles of selected genes, miRNAs, and methylation profiles of CpG sites was established. Finally, patients from the validation cohort were used to verify the performance of TPM.

2. Materials and Methods

2.1. Multiomics Dataset Acquisition from TCGA.
Somatic mutation profiles of 567 samples, gene expression profiles of 594 samples, DNA methylation profiles of 507 samples, and miRNA expression profiles of 495 samples were obtained from TCGA database using either GDC tool (https://portal.gdc.cancer.gov/) or TCGAbiolinks R package (Supplementary Table S1) [25, 26]. The somatic mutation profiles (mutation annotation format, MAF) were processed by Mutect software. Missense mutations, nonsense mutations, splice-site mutations, frameshift insertions, frameshift deletions, in-frame insertions, or in-frame deletions identified in the samples were regarded as positive mutations. Gene expression profiles of 594 samples were annotated through g:Profiler website [27] and normalized using the scale method in limma package [28]. The low abundance profiles were eliminated. DNA methylation profiles were annotated using IlluminaHumanMethylation450kanno.ilmnl12.hg19 R package. Quality control for DNA methylation profiles was conducted through minfi R package to eliminate certain CpG sites [29], in which the single-nucleotide polymorphisms (SNPs) existed [30], multiple mapping to human reference genome was found [31], and the methylation information of any samples was not available. In addition, CpG sites located in sex chromosomes were excluded for analysis [32]. miRNA expression profiles of 495 samples including 450 samples from LUAD tissue and 45 samples from matched normal lung tissue in TCGA database were downloaded from the University of California Santa Cruz (UCSC) Xena database (https://xena.ucsc.edu/public). Then, the miRNA expression profiles were transformed into reads per million (RPM), and miRNAs expressed in more than 10% of patients with LUAD were extracted. The clinical information of 522 patients with LUAD from TCGA database was obtained using TCGAbiolinks R package [25], which covered id, age, gender, tumor stage, state, weight, Body Mass Index (BMI), alcohol history, height, days to last follow-up, years smoked, and race (Table 1).

2.2. TMB Calculation and Classification of Patients.
Somatic mutation profiles were processed by Mutect software, and the identified somatic mutations, including base substitution, deletions, and insertions, were filtered according to the following criteria: (1) the minimum sequencing coverage for mutations should be greater than or equal to 10; (2) the variant allelic fraction should be greater than or equal to 5%. TMB was calculated as the total count of somatic mutations identified divided by 38 Mb, which is the length of exons in human genome. According to previously reported cutoff threshold of 10 in patients with LUAD [21, 33], patients were divided into TMB-H (TMB $\geq$ 10) and TMB-L (TMB < 10). Density plot of TMB-distribution for all patients with LUAD and boxplot of correlation between tumor stage and TMB was drawn by ggplot2 R package.

2.3. Tumor-Infiltrating Immune Analysis.
Tumor-infiltrating immune analysis was performed through Tumor Immune Estimation Resource (TIMER) tool [34]. The estimated abundances of six immune infiltrates (B cells, CD4(+) T cells, CD8(+) T cells, neutrophils, macrophages, and dendritic cells) were compared between TMB-H and TMB-L patients.
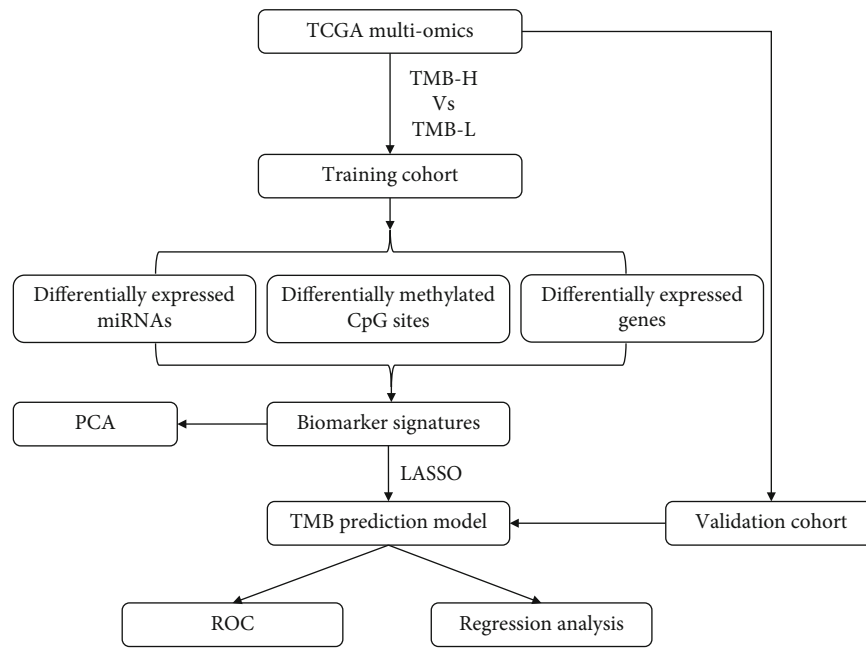
FIGURE 1: Flowchart of the analysis process in this study. TMB: tumor mutation burden; TMB-H: TMB-high; TMB-L: TMB-low; PCA: principal component analysis; LASSO: least absolute shrinkage and selection operator; ROC: receiver operating characteristic.

TABLE 1: Clinical information of 522 TCGA-LUAD patients.

Variables	Statistics
Gender	
Male (%)	242 (46.4%)
Female (%)	280 (53.6%)
Age	
80~89 (%)	30 (5.8%)
70~79 (%)	150 (28.7%)
60~69 (%)	146 (28.0%)
50~59 (%)	83 (16.0%)
40~49 (%)	25 (4.8%)
30~39 (%)	2 (0.4%)
Not reported (%)	86 (16.3%)
Race	
White (%)	393 (75.3%)
Black or African American (%)	53 (10.2%)
Asian (%)	8 (1.5%)
American Indian or Alaska native (%)	1 (0.2%)
Not reported (%)	67 (12.9%)
Status	
Alive (%)	334 (64.0%)
Dead (%)	188 (36.0%)
Tumor stage	
I (%)	279 (53.4%)
II (%)	124 (23.8%)
III (%)	85 (16.3%)
IV (%)	26 (5.0%)
Not reported (%)	8 (1.5%)

LUAD: lung adenocarcinoma.

2.4. Multiomics Analysis between TMB-H and TMB-L Patients. Differentially expressed genes between TMB-H and TMB-L patients in training cohort were identified through limma R package with -log10 adj.p value > 2 (adj.p value < 0.01) and log2 FoldChange > 3 [28] and then illustrated in volcano plot and heatmap by ggplot2 and pheatmap R package, respectively. Differentially expressed miRNAs between TMB-H and TMB-L patients in training cohort were identified through limma R package with -log10 adj.p-value > 1.30 (adj.p-value < 0.05) and log2 FoldChange > 0.35 [28] and then illustrated in volcano plot and heatmap by ggplot2 and pheatmap R package, respectively. In addition, target genes of the differentially expressed miRNAs were searched and analyzed through miRWalk website tool (http://mirwalk.umm.uni-heidelberg.de/) [35]. Differentially methylated CpG sites between TMB-H and TMB-L patients in training cohort were identified through limma R package with -log10 adj.p value > 1.30 (adj.p value < 0.05) and log2 FoldChange > 0.15 [28] and then illustrated in volcano plot and heatmap by ggplot2 and pheatmap R package, respectively.

2.5. Functional Enrichment Analysis. We first converted gene symbols into ENTREZ ID via org.Hs.eg.db R package, and then Gene Ontology (GO) and Kyoto Encyclopedia of Genes (KEGG) analysis of differentially expressed genes were implemented using ggplot2, enrichplot, and clusterProfiler R packages [36]. Meanwhile, GO and KEGG enrichment analysis were conducted for target genes of differentially expressed miRNAs using the same method described above.

2.6. Construction of TPM. We constructed 4 possible prediction biomarker signatures: gene signature (45 genes), miRNA signature (45 miRNAs), CpG site signature (45 CpG sites),
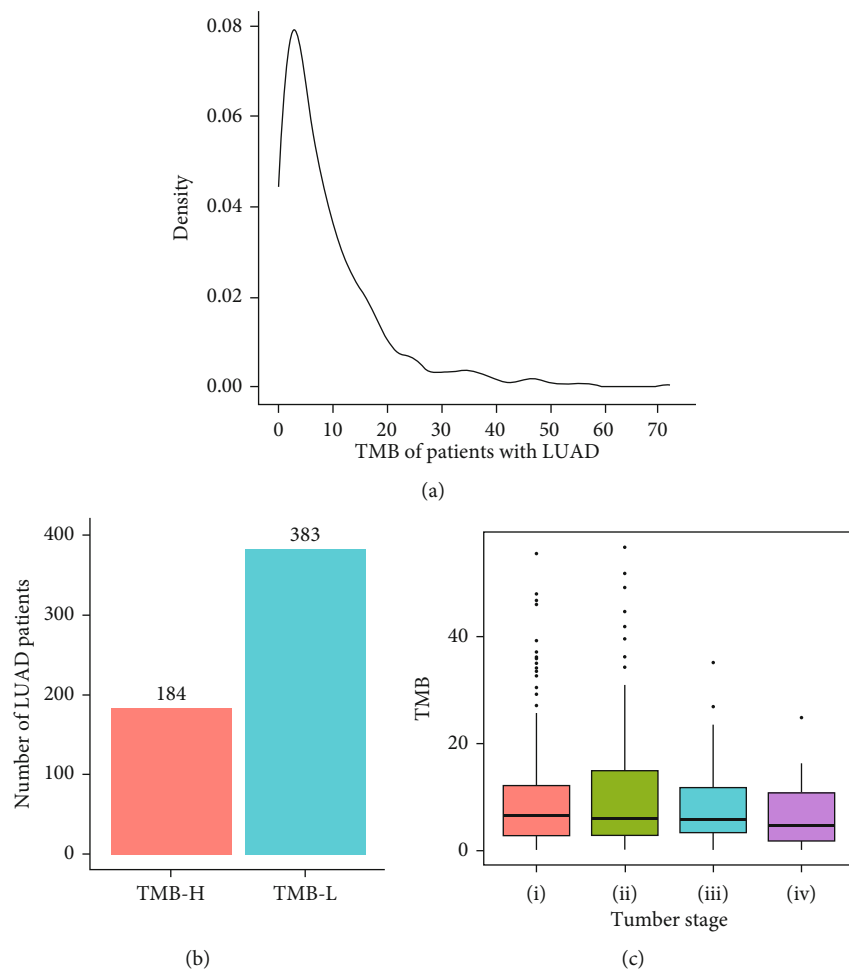
(a)

(b)

(c)

Figure 2: Division of patients with LUAD into TMB-H and TMB-L subgroups. (a) The distribution of TMB in patients with LUAD; (b) the number of TMB-H and TMB-L patients with LUAD; (c) the distribution of TMB across different tumor stages. TMB: tumor mutation burden; LUAD: lung adenocarcinoma; TMB-H: TMB-high; TMB-L: TMB-low; OS: overall survival.
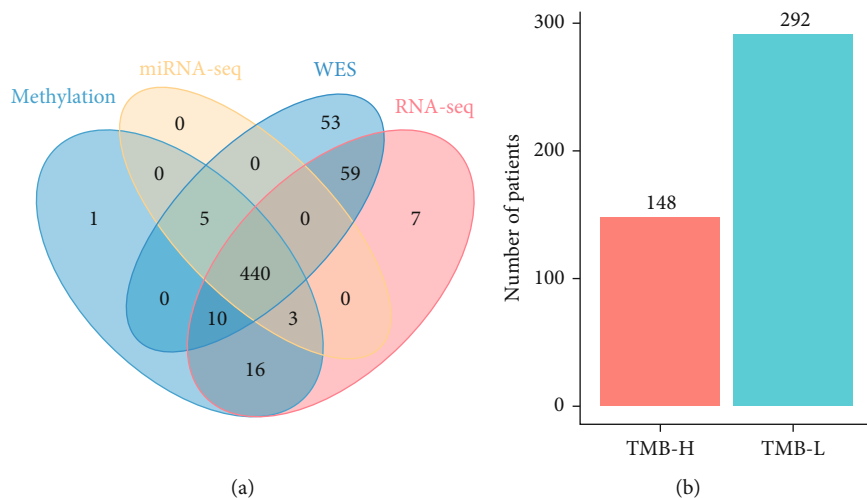

(a)

(b)

Figure 3: Multiomics data obtained from TCGA for patients with LUAD. (a) 440 patients with LUAD were found having coupled WES, DNA methylation, RNA-seq and miRNA-seq data; (b) 148 patients were classified as TMB-H and 292 patients were classified as TMB-L. WES: whole-exome sequencing; TMB-H: TMB-high; TMB-L: TMB-low.
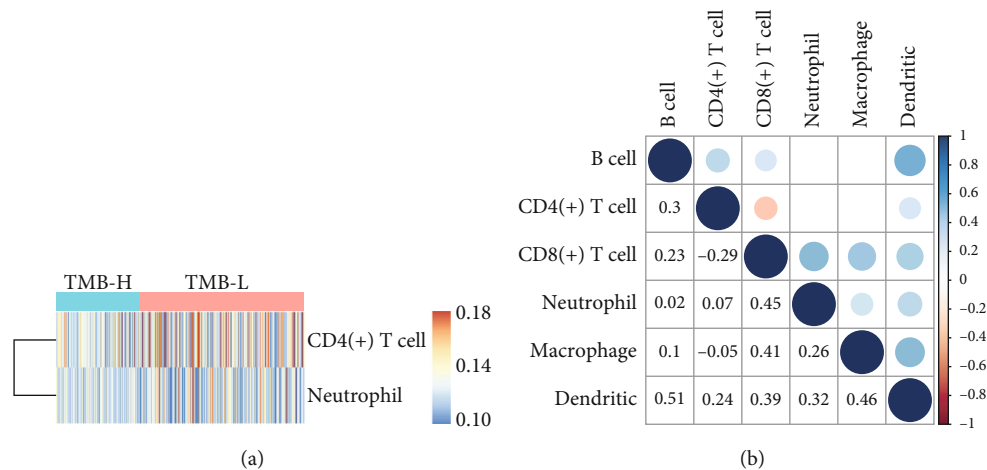
(a)

(b)

FIGURE 4: The landscape of tumor-infiltrating immune cells in TMB-H patients and TMB-L patients. (a) Relative proportions of infiltrating immune cells in TMB-H patients and TMB-L patients; (b) correlation matrix of all the proportions of 6 detected immune cell types. TMB-H: TMB-high; TMB-L: TMB-low.

and multiomics signature (15 genes + 15 miRNAs + 15 CpG sites) using differentially expressed genes, miRNAs, and differentially methylated CpG sites between TMB-H and TMB-L patients in the training cohort. Next, the least absolute shrinkage and selection operator (LASSO) logistic regression model analysis was performed to select the optimal biomarker signature for predicting TMB through glmnet R package [37]. The predictive performance for each biomarker signature was evaluated by lambda.min and matched area under curve (AUC). Finally, differentially expressed or methylated genes, miRNAs, and CpG sites identified with nonzero regression coefficients were used to construct the TPM. The TPM score was calculated using the regression coefficients from LASSO analysis to weigh the expression or methylation of the chosen biomarkers. The validation cohort was used to evaluate the performance of the TPM through assessing the predicting sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and AUC.

2.7. Principal Component Analysis (PCA).
Differentially expressed genes, miRNAs, and differentially methylated CpG sites identified through LASSO analysis were used to perform PCA. Expression or methylation profiles of the genes, miRNAs, and CpG sites were extracted from each patient, and ggfortify R package was utilized to conduct the PCA.

2.8. ROC Analysis.
ROC curve analysis was conducted using pROC R package to investigate the performance of TPM in predicting TMB [38].

2.9. Correlation Analysis and Regression Analysis.
Correlation between TPM score and TMB was analyzed by cor.test R function with the two-side Pearson's method. Samples were plotted in two-dimensional plots with the TPM score and TMB value. Regression analysis between TPM score and TMB was performed using lm R function.

3. Results

3.1. TMB-Based Division of Patients with LUAD.
WES data of tumor tissue from a total of 567 patients were acquired from TCGA-LUAD database, and clinical characteristics of the patients were summarized (Table 1); the mean age of the patients was 65.8, among which 242 individuals were males and 280 individuals were females (Table 1). TMB was calculated as the number of somatic mutations identified per megabase (Mb) in tumor tissue of each patient. It was found that most of patients with LUAD had a TMB ranging from 0 to 40 (Figure 2(a)). According to cutoff threshold of TMB = 10, 184 patients were classified as TMB-H and 383 patients were classified as TMB-L (Figure 2(b)). The TMB-H and TMB-L patients were found evenly distributed in different tumor stages as expected (Figure 2(c)). Furthermore, tumor samples from 440 patients were found also having RNA-seq, miRNA-seq, and DNA methylation datasets (Figure 3(a), Supplementary Table S1), among which 148 patients belong to the TMB-H group and 292 patients belong to the TMB-L group (Figure 3(b)). Patients were then randomly split into a training cohort (70%, 103 TMB-H patients vs. 204 TMB-L patients) and a validation cohort (30%, 45 TMB-H patients vs. 88 TMB-L patients) without overlap for developing a multiomics model to predict TMB.

3.2. Landscape of Tumor-Infiltrating Immune Cells in Patients with LUAD.
Proportions of different tumor-infiltrating immune cells between TMB-H patients and TMB-L patients were calculated and summarized (Figure 4(a), Supplementary Figure S1), in which the abundance of CD4 (+) T cells ($p = 0.030$) showed more abundant density in TMB-L patients compared with TMB-H patients, whereas the B cell, CD8 (+) T cell, dendritic cell, macrophage cell, and neutrophil cell had similar density between TMB-H patients and TMB-L patients (Supplementary Figure S1, Figure 4(a)). Meanwhile, correlations among different tumor-infiltrating immune cell types were moderate or weak (Figure 4(b)).
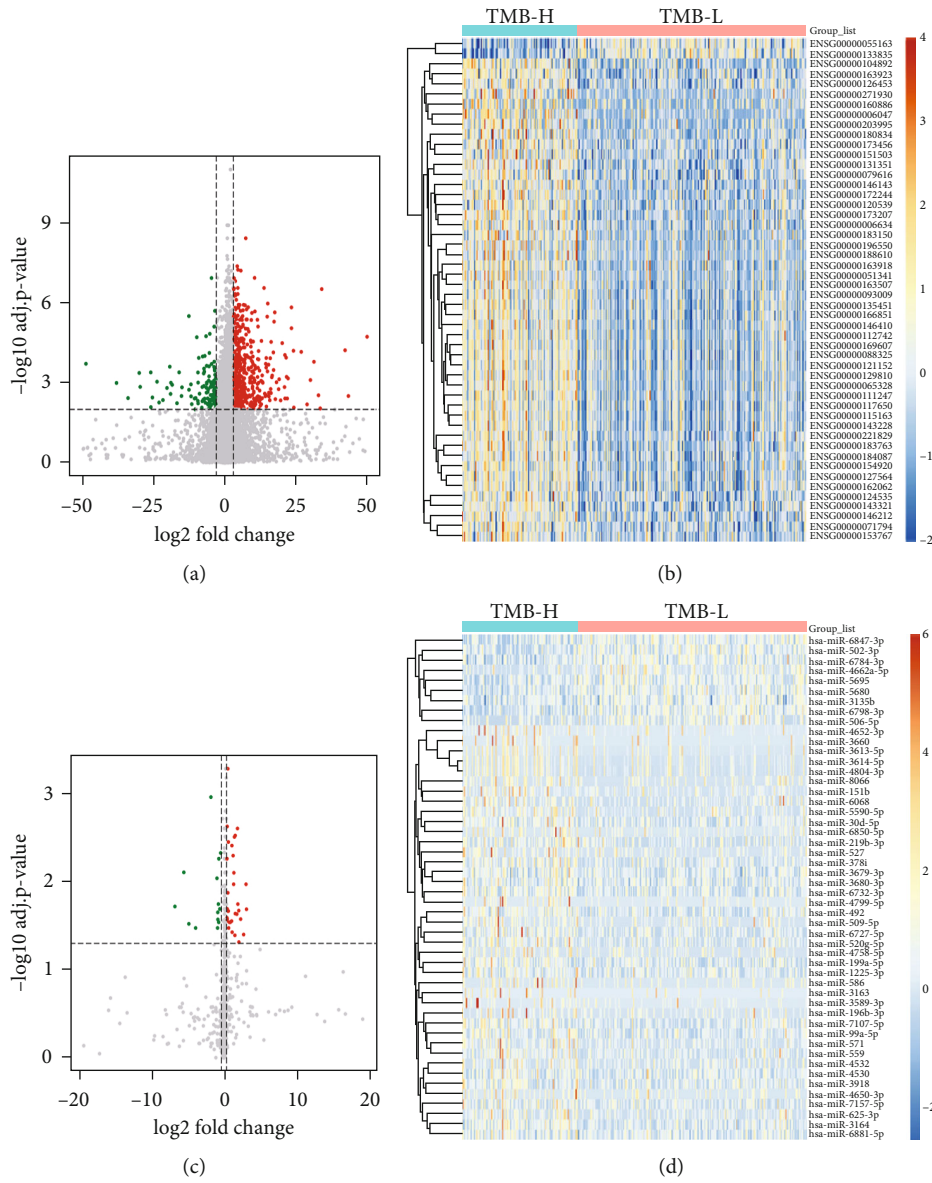
(a)

(b)
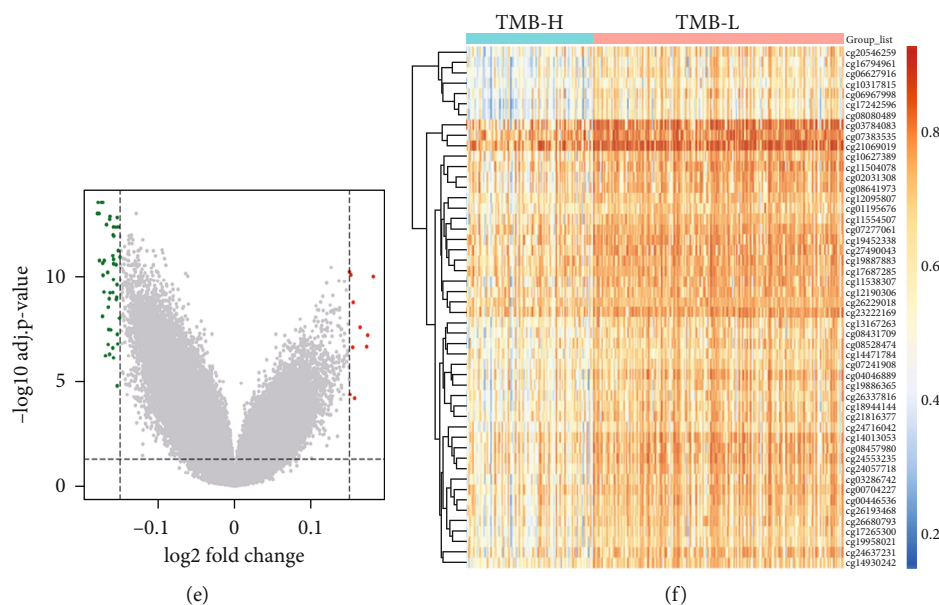

(c)

(d)

FIGURE 5: Continued.

FIGURE 5: Characterization of the top50 differential expressed genes, miRNAs, and differential methylated CpG sites between TMB-H and TMB-L patients. Volcano plot showed the differentially expressed genes (a) and miRNAs (c) or differentially methylated CpG sites (e) between TMB-H and TMB-L patients. The red dots represent upregulated genes, miRNAs, or hypermethylated CpG sites; the blue dots represent downregulated genes, miRNAs, or hypomethylated CpG sites; the black dots represent genes, miRNAs, or CpG sites with no significantly differential expression or methylation. Hierarchical clustering heatmap of differentially expressed genes (b) and miRNAs (d) or differentially methylated CpG sites (f) between TMB-H and TMB-L patients. Orange indicates the upregulated genes, miRNAs, or hypermethylated CpG sites; blue indicates the downregulated genes, miRNAs, or hypomethylated CpG sites. TMB-H: TMB-high; TMB-L: TMB-low.

These results suggested that TMB might be associated with the abundance of CD4 (+) T cell in patients with lung cancer.

3.3. Multiomics Analysis of Transcriptome, miRNAome, and Methylome between TMB-H and TMB-L Patients. To explore differences in a tumor microenvironment between TMB-H and TMB-L patients, differentially expressed genes, miRNAs, and differentially methylated CpG sites between TMB-H patients and TMB-L patients in the training cohort were identified. In summary, 480 genes and 36 miRNAs were upregulated in TMB-H patients, whereas 130 genes and 14 miRNAs were downregulated in TMB-H patients (Figures 5(a)–5(d)). Moreover, 10 CpG sites were hypermethylated and 48 CpG sites were hypomethylated in TMB-H patients (Figures 5(e) and 5(f)). GO-enrichment analysis suggested that the differentially expressed genes were mainly involved in the biological processes including nuclear division, chromosome segregation, and organelle fission (Supplementary Figure S2A). KEGG pathway enrichment analysis suggested that the differentially expressed genes were mainly related to pyrimidine metabolism (Supplementary Figure S2B). These results demonstrated that differentially expressed genes might be correlated with carcinogenesis-related processes [39]. In addition, GO and KEGG enrichment analysis were also performed for the target genes of differentially expressed miRNAs, which were found to be mainly enriched in netrin-activated signaling pathway, DNA-binding transcription activator, and single-stranded RNA binding (Supplementary Figure S3A,

Figure S3B). In addition, most of the differentially methylated CpG sites were found to locate in gene body regions (Supplementary Figure S4), and 5 CpG sites were found to locate in the TSS1500 (sequence region from -200 to -1500 nt upstream of the transcription start site) and TSS200 (sequence region -200 nt upstream of the transcription start site) region of genes (Supplementary Figure S4, Supplementary Table S2).

3.4. Machine Learning-Based Construction of TMB Prediction Model. To develop TPM based on the differences identified through multiomics analysis, we firstly generated 4 possible prediction biomarker signatures including gene signature, miRNA signature, CpG site signature, and multiomics signature, which were composed of expression profiles of top 45 differentially expressed genes, top 45 differentially expressed miRNAs and top 45 differentially methylated CpG sites between TMB-H and TMB-L patients, respectively (Supplementary Table S3). To further compare predicting efficacy of the four signature, we implemented LASSO logistic analysis to select the optimal signature from training cohort. The optimal biomarkers for the 4 prediction signature were obtained with nonzero regression coefficients (Figure 6, Table 2), and as a result, the multiomics signature with maximum measure (0.868) was selected as the optimal biomarker signature for predicting TMB (Figure 6(d), Table 2). PCA using the shrunk multiomics signature suggested that TMB-H patients and TMB-L patients could be separated obviously (Supplementary Figure S5). Based on
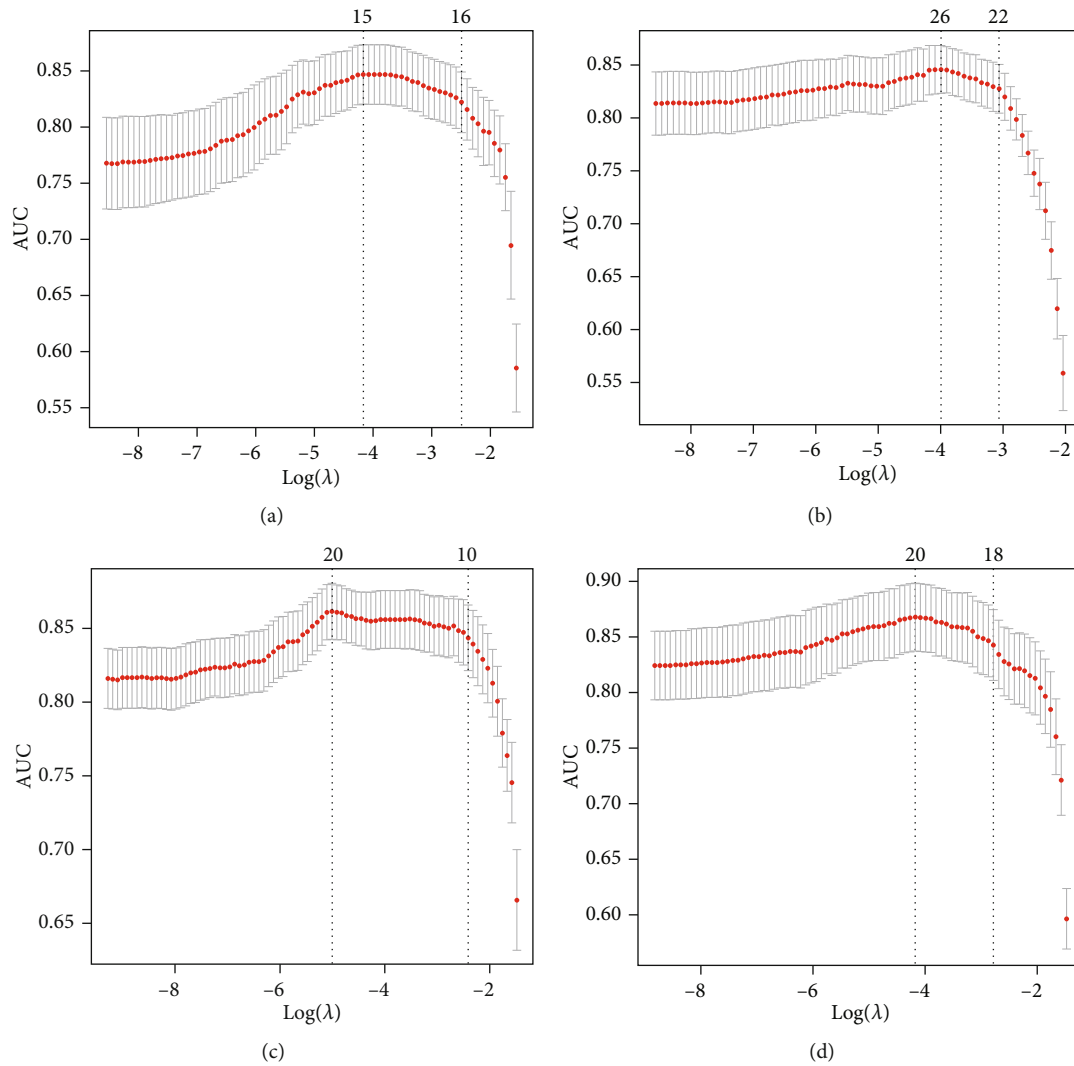
(a)

(b)

(c)

(d)

FIGURE 6: LASSO regression analysis for 4 possible prediction biomarker signatures. 10-fold cross-validation in LASSO regression analysis for gene signature (a), miRNA signature (b), CpG site signature (c), and multiomics signature (d). LASSO: least absolute shrinkage and selection operator; AUC: area under curve.

the multiomics signature, we finally constructed TPM by weighing expression or methylation of the genes, miRNAs, and CpG sites through regression coefficients from the LASSO analysis (Table 3). The TPM was showed as the following math formula: TPM score $= -1.555696454 * \mathrm{cg}$ $02031308 - 0.939485314 * \mathrm{cg}03286742 - 0.532855695 * \mathrm{cg}$ $04046889 - 1.603385472 * \mathrm{cg}12095807 - 1.171295176 * \mathrm{cg}$ $16794961 - 1.341848062 * \mathrm{cg}24553235 + 0.203290638 *$ $\mathrm{YBX2} + 0.000323171 * \mathrm{HLTF} + 0.355814358 * \mathrm{KLC3} +$ $0.017454209 * \mathrm{WRNIP1} + 0.010739241 * \mathrm{CKS1B} +$ $0.013056543 * \mathrm{RNF26} + 0.039397451 * \mathrm{ZYG11A} +$ $0.582628142 * \mathrm{hsa-miR-571} + 3.954182602 * \mathrm{hsa-miR}$ $-586 + 0.068239671 * \mathrm{hsa-miR-151b} + 0.000724033 *$ $\mathrm{hsa-miR-378i} + 0.25824073 * \mathrm{hsa-miR-6727-5p} -$ $0.731679875 * \mathrm{hsa-miR-502-3p} - 0.007119299 * \mathrm{hsa-}$ $\mathrm{miR-6798-3p}$. The AUC of the constructed TPM in the training cohort was 0.911 showing its superior predictive accuracy (Figure 7(a)). Besides, the p value of a two-side t-test was $3.40e-48$ between TPM score and TMB (Figure 7(b)),

which suggested that TPM score was highly correlated with TMB in patients with LUAD.

3.5. Evaluation of the Predicting Accuracy of TPM in the Validation Cohort. To evaluate the predicting efficacy of TPM constructed in training cohort, expression or methylation profiles of genes, miRNAs, and CpG sites in patients from validation cohort were used as input parameters for calculating TPM score. According to the threshold of -3.366, 41 patients from the TMB-H group were predicted as TMB-H, and 66 patients from the TMB-L group were predicted as TMB-L. In summary, the TPM has a sensitivity of 0.911, a specificity of 0.750, and an accuracy of 0.805 in predicting TMB in the validation cohort, and the PPV was 0.651 and NPV was 0.943 (Supplementary Table S4). ROC analysis revealed the AUC of the TPM in validation cohort was 0.859 (Figure 8(a)), and p value of the two-side t-test was $1.19e-14$ between the TPM score and TMB (Figure 8(b)). These results suggested that the TPM

TABLE 2: The performance of 4 optimal biomarker signatures obtained by LASSO regression analysis.

Biomarker signature	Optimal biomarkers	lambda.min	Measure
Gene	GTF2IRD1, FTSJ1, CHMP4B, KLC3, DMAC2, GIT1, SOHLH2, SYNGR3, SAP130, LRRC1, FN3KRP, POU4F1, ZNF526, KRT80, UBE2C, FOXE1, MEX3D, CIDECP1, PRR19, DHX16, FANCG, AC010632.1, AC019171.1	0.018	0.884
miRNA	hsa-miR-22-5p, hsa-miR-486-5p, hsa-miR-492, hsa-miR-561-5p, hsa-miR-151b, hsa-miR-3677-5p, hsa-miR-3923, hsa-miR-4425, hsa-miR-4434, hsa-miR-4536-5p, hsa-miR-4679, hsa-miR-5702, hsa-miR-6727-5p, hsa-miR-6858-5p, hsa-miR-7107-5p, hsa-let-7 g-3p, hsa-miR-136-3p, hsa-miR-155-3p, hsa-miR-371a-3p, hsa-miR-491-3p, hsa-miR-432-3p, hsa-miR-574-3p, hsa-miR-3074-3p, hsa-miR-3622b-3p, hsa-miR-3679-3p, hsa-miR-3150b-3p, hsa-miR-4639-3p, hsa-miR-4655-3p, hsa-miR-6798-3p, hsa-miR-6847-3p	0.017	0.734
CpG site	cg01862650, cg02031308, cg02916472, cg07184316, cg07729440, cg10120778, cg10488199, cg11002952, cg20151576, cg20297017, cg20671274, cg21827634, cg22773522, cg23049130, cg25841348	0.015	0.845
Multiomics	cg01862650, cg07729440, cg20671274, cg21827634, cg22773522, GTF2IRD1, FTSJ1, TTI1, CHMP4B, KLC3, HNRNPUL1, UBE2S, BCL2L12, SYNGR3, KRT80, FOXE1, AC006213.3, hsa-miR-22-5p, hsa-miR-492, hsa-miR-4536-5p, hsa-miR-6727-5p, hsa-miR-7107-5p, hsa-miR-136-3p, hsa-miR-3679-3p, hsa-miR-6816-3p	0.010	0.938

LASSO: least absolute shrinkage and selection operator: TMB: tumor mutation burden.

TABLE 3: Coefficient of each biomarker of multiomics signature in LASSO model analysis.

	Biomarkers	Coefficient
Multiomics	cg01862650	-1.818730262
	cg07729440	-6.256940647
	cg20671274	-0.911730577
	cg21827634	-3.867697524
	cg22773522	-2.460934084
	GTF2IRD1	0.039415926
	FTSJ1	0.012286235
	TTI1	0.042479337
	CHMP4B	0.009500278
	KLC3	0.454704618
	HNRNPUL1	0.015834511
	UBE2S	0.014978452
	BCL2L12	0.078730525
	SYNGR3	0.192724739
	KRT80	0.017563239
	FOXE1	0.011660062
	AC006213.3	0.175314579
	hsa-miR-22-5p	1.230322969
	hsa-miR-492	-0.185026622
	hsa-miR-4536-5p	7.457452192
	hsa-miR-6727-5p	0.560095048
	hsa-miR-7107-5p	-0.131067725
	hsa-miR-136-3p	0.93577811
	hsa-miR-3679-3p	-0.007961529
	hsa-miR-6816-3p	1.359432308

LASSO: least absolute shrinkage and selection operator.

performed relatively high TMB-predicting accuracy in an independent validation cohort.

4. Discussion

Immunotherapy has been demonstrated particularly successful in NSCLC treatment and is being adopted as a first-line treatment option worldwide [12, 21, 40]. Nevertheless, only a small portion of unselected patients can benefit from immunotherapy [24, 41]. Therefore, biomarkers for patient selection become important to achieve effective therapy. TMB has been recognized as the effective prognostic biomarker in NSCLC patients according to the results from numerous clinical trials [21, 22, 42, 43]. Although targeted NGS has been proved to be an alternative approach of WES for the prediction of TMB, the accuracy and simplicity of targeted NGS remain challenging as various parameters should be taken into consideration [44]. In this study, we developed a mathematic multiomics model that could precisely predict TMB in patients with LUAD, and the prediction accuracy of the model was validated in an independent cohort with high sensitivity and specificity. Furthermore, as the input parameter in this model includes expression profiles of 7 genes, 7 miRNAs, and the methylation profiles of 6 CpG sites, which could be obtained through quantitative real time-polymerase chain reaction (qRT-PCR). This model paved the way for further development of the simplified qRT-PCR-based clinical assay for TMB prediction.

The tumor microenvironment refers to the network of cells and structures surrounding a tumor cell, and it consists of immune cells, mesenchymal cells, endothelial cells, extracellular matrix (ECM) molecules, and inflammatory mediators [45]. High TMB indicates the presence of more neoantigens in tumor microenvironment, which promotes the inflammatory response and results in the alteration of transcriptomic and epigenetic changers [45]. It has been
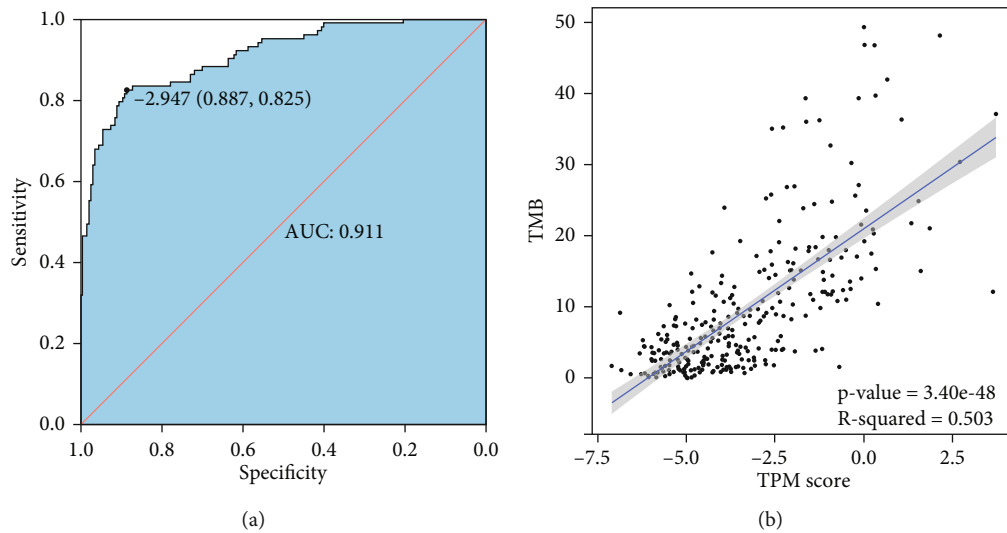
(a)

(b)

Figure 7: The performance of TPM in the training cohort. (a) ROC analysis of the TPM score in the training cohort; (b) the TPM score is highly correlated with TMB. TPM: TMB prediction model; AUC: area under curve.
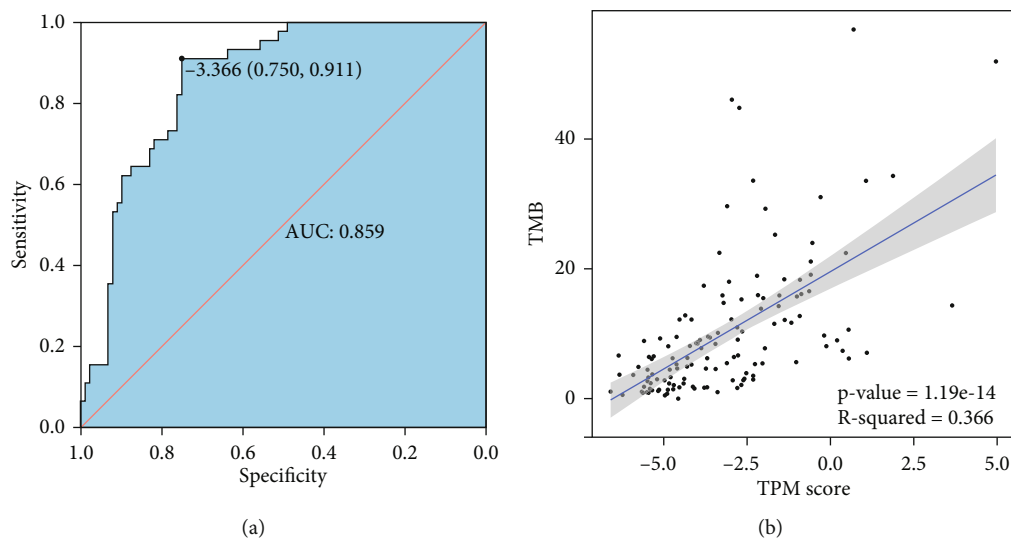

(a)

(b)

Figure 8: The performance of TPM in the validation cohort. (a) AUC of ROC analysis was 0.859 showing the great predictive accuracy of TPM; (b) the TPM score is highly correlated with TMB with p value $= 1.19e - 14$. TPM: TMB prediction model; AUC: area under curve.

proved that gene expression signatures in the tumor microenvironment were associated with the prognosis in NSCLC [46–49]. In agreement with previous studies, the differentially expressed genes between TMB-H and TMB-L patients identified in this study were found to enrich in the immune-related damaged DNA binding, nuclear division, nuclear chromosome segregation, organelle fission, single-stranded DNA binding, ribonucleoprotein complex binding, and pyrimidine metabolism (Supplementary Figure S2) [50–54]. The 7 genes used in constructing TPM might be involved in carcinogenesis; for instance, Y box binding protein 2 (YBX2) was differentially expressed between different subtypes of breast cancer and was one of RNA processing factors which contribute to subtype-specific splicing [55]. Meanwhile, it was found that LINC00958

promoted cell proliferation and migration in oral squamous cell carcinoma through the miR-627-5p/YBX2 axis [56]. Moreover, it was reported that the wild type alleles of kinesin light chain 3 (KLC3) Lys751Gln were significantly correlated with greater smoking intensity, and genetic variations may influence the progression of lung cancer [57]. In addition, the expression of CDC28 protein kinase regulatory subunit 1 (CKS1B) in lung cancer cells developed the chemoresistance through the Hsp90 and MEK1/2 pathway [58].

miRNA expression in tumor microenvironment plays a crucial role in mediating and controlling several immune and cell interactions and convolutes in the regulation of immune checkpoints, PD1 and PD-L1 [59]. It was reported that a 25 miRNA-based signature classifier could predict

the TMB level with high accuracy [60]. A cluster of highly expressed miRNA including hsa-miR-492, hsa-miR-498, and hsa-miR-320 were found to be correlated with tumorigenesis of retinoblastoma [61]. Moreover, the invasion, proliferation, and migration of cervical cancer cells were found to be promoted by hsa-miR-6727-5p, which might play an important role in cervical cancer progress [62]. In this study, we mapped the differentially expressed miRNAs between TMB-H and TMB-L patients to their target genes, and enrichment analysis of the target genes suggested that DNA-binding transcription activator, single-stranded RNA binding, MAP kinase activity, and glycerolipid metabolism that related to lung cancer were affected in a tumor microvironment. The 7 miRNAs used in constructing the TPM in this study include hsa-miR-571, hsa-miR-586, hsa-miR-151b, hsa-miR-378i, hsa-miR-6727-5p, hsa-miR-502-3p, and hsa-miR-6798-3p, among which hsa-miR-378i and miR-502-3p were demonstrated to be important for colorectal cancer carcinoma metastasis [63, 64].

Changes in DNA methylation are one of the most important epigenetic alterations in a tumor microvironment. A multicenter study in 15 hospitals suggested epigenomic profile based on a microarray DNA methylation signature (EPIMMUNE) could serve as an effective biomarker in predicting the outcomes of NSCLC patients treated with PD-1 inhibitors [65], and the FOXP1 could be a predictive biomarker for better-selecting patients to benefit with immunotherapy [65]. The CpG site signature also had a relatively high predictive performance (measure = 0.861) of TMB in our study, suggesting its great value in NSCLC prognosis. cg02849937 located in the TSS1500 region of C7orf13 and its expression level were negatively associated with promoter methylation using whole-genome integrative analysis [66]. In addition, cg27281030 is located in the TSS1500 region of NLRP12, which has been demonstrated regulate inflammation, and it is believed that hepatocellular carcinoma was negatively regulated by NLRP12 through suppression of inflammation and proliferation of hepatocytes [67].

Through multiomics analysis, we integrated gene/miRNA expression and DNA methylation data to reflect subtle alterations of the tumor microvironment to precisely predict TMB for better prognosis of patients with LUAD in immunotherapy. Fragments per kilobase per million mapped reads (FPKM) of YBX2, HLTF, KLC3, WRNIP1, CKS1B, RNF26, ZYG11A, and RPM of hsa-miR-571, hsa-miR-586, hsa-miR-151b, hsa-miR-378i, hsa-miR-6727-5p, hsa-miR-502-3p, and hsa-miR-6798-3p as well as beta-value of cg02031308, cg03286742, cg04046889, cg12095807, cg16794961, and cg24553235 were extracted from RNA-seq, miRNA-seq and Illumina HumanMethylation450 BeadChip, respectively, for calculating TPM score. Although the FPKM, RPM, and beta value involved in the TPM were based on high-throughput sequencing or chip analysis, it is feasible to convert them to cycle threshold (Ct) value in qRT-PCR analysis and thus simplify the prediction of TMB by using benchtop qRT-PCR instrument. The conversion of FPKM in different samples to Ct values could be probably through comparison of the targeted gene expression to reference gene expressions, such as actin and eukary-otic elongation factor (eEF), which have relative consistent expression under different tumor microvironment, and beta value of CpG sites could also be converted into Ct value though the quantitative MethyLight technology [68]. To our best knowledge, this is the first time to construct the TPM for patients with LUAD from multiomics view.

5. Conclusion

In summary, the present study developed a multiomics risk model with high specificity and sensitivity in predicting TMB for patients with LUAD and laid the base for establishing a more simplified and cost-effective TMB test assay. Nevertheless, this study was solely bioinformatics research, and clinical sample validation for the TPM had not been implemented. The training cohort and the validation cohort used in this study were relatively small in size and required further expansion to increase the accuracy.

Data Availability

Somatic mutation profiles of 567 samples, gene expression profiles of 594 samples, DNA methylation profiles of 507 samples, and miRNA expression profiles of 495 samples were obtained from TCGA database.

Conflicts of Interest

The author(s) declare(s) that they have no conflicts of interest.

Authors' Contributions

Jun Wang and Peng Chen contributed equally to this work.

Acknowledgments

Supplementary Materials

Supplementary Table S1: summary of TCGA-LUAD data. Supplementary Table S2: five CpG sites located in the TSS1500 and TSS200 region of the genes. Supplementary Table S3: different biomarker signatures used in the training cohort. Supplementary Table S4: the performance of TPM in the validation cohort. Supplementary Figure S1: the differences in the abundance of tumor-infiltrating immune cells between TMB-H and TMB-L patients. Student's t-test was used; error bars indicated standard deviation; $^*p < 0.05$. TMB-H: TMB-high; TMB-L: TMB-low. Supplementary

Figure S2: GO and KEGG analysis for the differentially expressed genes. Supplementary Figure S3: GO and KEGG analysis for the target genes of differentially expressed miRNAs. Supplementary Figure S4: distribution of differentially methylated CpG sites in genes. Supplementary Figure S5: principal component analysis (PCA) for the multiomics signature after LASSO variable reduction. (Supplementary Materials)

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: a Cancer Journal for Clinicians*, vol. 69, no. 1, pp. 7–34, 2019.

[2] W. Chen, K. Sun, R. Zheng et al., "Cancer incidence and mortality in China, 2014," *Chinese Journal of Cancer Research*, vol. 30, no. 1, pp. 1–12, 2018.

[3] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[4] R. Siegel, E. Ward, O. Brawley, and A. Jemal, "Cancer statistics, 2011," *CA: a Cancer Journal for Clinicians*, vol. 61, no. 4, pp. 212–236, 2011.

[5] S. Takamori, G. Toyokawa, K. Takada, F. Shoji, T. Okamoto, and Y. Maehara, "Combination Therapy of Radiotherapy and Anti-PD-1/PD-L1 Treatment in Non- Small-cell Lung Cancer: A Mini-review," *Clinical Lung Cancer*, vol. 19, no. 1, pp. 12–16, 2018.

[6] D. E. Meyers, P. M. Bryan, S. Banerji, and D. G. Morris, "Targeting the PD-1/PD-L1 axis for the treatment of non-small-cell lung cancer," *Current oncology*, vol. 25, no. 4, pp. 324–e334, 2018.

[7] W. Luo, Z. Wang, P. Tian, and W. Li, "Safety and tolerability of PD-1/PD-L1 inhibitors in the treatment of non-small cell lung cancer: a meta-analysis of randomized controlled trials," *Journal of Cancer Research and Clinical Oncology*, vol. 144, no. 10, pp. 1851–1859, 2018.

[8] J. X. Li, J. M. Huang, Z. B. Jiang et al., "Current clinical progress of PD-1/PD-L1 immunotherapy and potential combination treatment in non-small cell lung cancer," *Integrative Cancer Therapies*, vol. 18, 2019.

[9] M. Khunger, P. Jain, S. Rakshit et al., "Safety and efficacy of PD-1/PD-L1 inhibitors in treatment-naive and chemotherapy-refractory patients with non-small-cell lung cancer: a systematic review and meta-analysis," *Clinical Lung Cancer*, vol. 19, no. 3, pp. e335–e348, 2018.

[10] J. Brahmer, K. L. Reckamp, P. Baas et al., "Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer," *The New England Journal of Medicine*, vol. 373, no. 2, pp. 123–135, 2015.

[11] R. S. Herbst, P. Baas, D. W. Kim et al., "Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial," *The Lancet*, vol. 387, no. 10027, pp. 1540–1550, 2016.

[12] E. J. Aguilar, B. Ricciuti, J. F. Gainor et al., "Outcomes to first-line pembrolizumab in patients with non-small-cell lung cancer and very high PD-L1 expression," *Annals of Oncology*, vol. 30, no. 10, pp. 1653–1659, 2019.

[13] M. Reck, D. Rodríguez-Abreu, A. G. Robinson et al., "Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer," *The New England Journal of Medicine*, vol. 375, no. 19, pp. 1823–1833, 2016.

[14] T. Landre, G. Des Guetz, K. Chouahnia, C. Taleb, A. Vergnenegre, and C. Chouaid, "First-line PD-1/PD-L1 inhibitor plus chemotherapy vs chemotherapy alone for negative or < 1% PD-L1-expressing metastatic non-small-cell lung cancers," *Journal of Cancer Research and Clinical Oncology*, vol. 146, no. 2, pp. 441–448, 2020.

[15] P. Martinez, S. Peters, T. Stammers, and J. C. Soria, "Immunotherapy for the first-line treatment of patients with metastatic non-small cell lung cancer," *Clinical Cancer Research*, vol. 25, no. 9, pp. 2691–2698, 2019.

[16] J. R. Brahmer, R. Govindan, R. A. Anders et al., "The Society for Immunotherapy of Cancer consensus statement on immunotherapy for the treatment of non-small cell lung cancer (NSCLC)," *Journal for Immunotherapy of Cancer*, vol. 6, no. 1, p. 75, 2018.

[17] A. Rittmeyer, F. Barlesi, D. Waterkamp et al., "Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial," *Lancet*, vol. 389, no. 10066, pp. 255–265, 2017.

[18] L. Fehrenbacher, A. Spira, M. Ballinger et al., "Atezolizumab versus docetaxel for patients with previously treated non-small- cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial," *The Lancet*, vol. 387, no. 10030, pp. 1837–1846, 2016.

[19] R. Hui, E. B. Garon, J. W. Goldman et al., "Pembrolizumab as first-line therapy for patients with PD-L1-positive advanced non-small cell lung cancer: a phase 1 trial," *Annals of Oncology*, vol. 28, no. 4, pp. 874–881, 2017.

[20] N. A. Rizvi, M. D. Hellmann, A. Snyder et al., "Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer," *Science*, vol. 348, no. 6230, pp. 124–128, 2015.

[21] M. D. Hellmann, T. E. Ciuleanu, A. Pluzanski et al., "Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden," *The New England Journal of Medicine*, vol. 378, no. 22, pp. 2093–2104, 2018.

[22] R. M. Samstein, C. H. Lee, A. N. Shoushtari et al., "Tumor mutational load predicts survival after immunotherapy across multiple cancer types," *Nature Genetics*, vol. 51, no. 2, pp. 202–206, 2019.

[23] M. Allgäuer, J. Budczies, P. Christopoulos et al., "Implementing tumor mutational burden (TMB) analysis in routine diagnostics-a primer for molecular pathologists and clinicians," *Transl Lung Cancer Res*, vol. 7, no. 5, pp. 703–715, 2018.

[24] R. Buttner, J. W. Longshore, F. Lopez-Rios et al., "Implementing TMB measurement in clinical practice: considerations on assay requirements," *ESMO Open*, vol. 4, no. 1, 2019.

[25] A. Colaprico, T. C. Silva, C. Olsen et al., "TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data," *Nucleic Acids Research*, vol. 44, no. 8, 2016.

[26] C. Peng, W. Jun, S. Mingyang, Z. Guocheng, Z. Shasha, and G. Deming, "Integrative modeling of multi-omics data for predicting tumor mutation burden in lung cancer patients," *Research Square*, 2021.

[27] J. Reimand, T. Arak, P. Adler et al., "G:profiler-a web server for functional interpretation of gene lists (2016 update)," *Nucleic Acids Research*, vol. 44, no. W1, pp. W83–W89, 2016.

[28] M. E. Ritchie, B. Phipson, D. Wu et al., "Limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, 2015.

[29] J. P. Fortin, T. J. Triche Jr., and K. D. Hansen, "Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi," *Bioinformatics*, vol. 33, no. 4, pp. 558–560, 2017.

[30] J. Sandoval, J. Mendez-Gonzalez, E. Nadal et al., "A prognostic DNA methylation signature for stage I non-small-cell lung cancer," *Journal of Clinical Oncology*, vol. 31, no. 32, pp. 4140–4147, 2013.

[31] M. E. Price, A. M. Cotton, L. L. Lam et al., "Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array," *Epigenetics & Chromatin*, vol. 6, no. 1, 2013.

[32] Y. A. Chen, M. Lemire, S. Choufani et al., "Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray," *Epigenetics*, vol. 8, no. 2, pp. 203–209, 2013.

[33] N. Ready, M. D. Hellmann, M. M. Awad et al., "First-line nivolumab plus ipilimumab in advanced non-small-cell lung cancer (CheckMate 568): outcomes by programmed death ligand 1 and tumor mutational burden as biomarkers," *Journal of Clinical Oncology*, vol. 37, no. 12, pp. 992–1000, 2019.

[34] T. Li, J. Fan, B. Wang et al., "TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells," *Cancer Research*, vol. 77, no. 21, pp. e108–e110, 2017.

[35] C. Sticht, C. De La Torre, A. Parveen, and N. Gretz, "miRWalk: an online resource for prediction of microRNA binding sites," *PLoS One*, vol. 13, no. 10, 2018.

[36] G. Yu, L. G. Wang, Y. Han, and Q. Y. He, "clusterProfiler: an R package for comparing biological themes among gene clusters," *OMICS*, vol. 16, no. 5, pp. 284–287, 2012.

[37] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.

[38] X. Robin, N. Turck, A. Hainard et al., "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics*, vol. 12, no. 1, 2011.

[39] D. Hanahan and R. A. Weinberg, "The hallmarks of cancer," *Cell*, vol. 100, no. 1, pp. 57–70, 2000.

[40] S. Gettinger, N. A. Rizvi, L. Q. Chow et al., "Nivolumab monotherapy for first-line treatment of advanced non-small-cell lung cancer," *Journal of Clinical Oncology*, vol. 34, no. 25, pp. 2980–2987, 2016.

[41] D. P. Carbone, M. Reck, L. Paz-Ares et al., "First-line nivolumab in stage IV or recurrent non-small-cell lung cancer," *The New England Journal of Medicine*, vol. 376, no. 25, pp. 2415–2426, 2017.

[42] D. R. Gandara, S. M. Paul, M. Kowanetz et al., "Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab," *Nature Medicine*, vol. 24, no. 9, pp. 1441–1448, 2018.

[43] C. Kandoth, M. D. McLellan, F. Vandin et al., "Mutational landscape and significance across 12 major cancer types," *Nature*, vol. 502, no. 7471, pp. 333–339, 2013.

[44] A. S. D. Patel, D. Soneji, P. Parikh, and M. Kumar, "Biomarkers in immuno-oncology: a review article," *International Journal of Molecular & Immuno Oncology*, vol. 4, no. 2, pp. 41–49, 2019.

[45] N. K. Altorki, G. J. Markowitz, D. Gao et al., "The lung microenvironment: an important regulator of tumour growth and metastasis," *Nature Reviews. Cancer*, vol. 19, no. 1, pp. 9–31, 2019.

[46] G. Chen, Z. Dong, D. Wu, and Y. Chen, "Profiles of immune infiltration in lung adenocarcinoma and their clinical significant: a gene-expression-based retrospective study," *Journal of Cellular Biochemistry*, vol. 121, no. 11, pp. 4431–4439, 2020.

[47] C. Yue, H. Ma, and Y. Zhou, "Identification of prognostic gene signature associated with microenvironment of lung adenocarcinoma," *PeerJ*, vol. 7, 2019.

[48] J. Li, X. Li, C. Zhang, C. Zhang, and H. Wang, "A signature of tumor immune microenvironment genes associated with the prognosis of non-small cell lung cancer," *Oncology Reports*, vol. 43, no. 3, pp. 795–806, 2020.

[49] J. Sun, Z. Zhang, S. Bao et al., "Identification of tumor immune infiltration-associated lncRNAs for improving prognosis and immunotherapy response of patients with non-small cell lung cancer," *Journal for Immunotherapy of Cancer*, vol. 8, no. 1, 2020.

[50] B. K. Slowikowski, B. Galecki, W. Dyszkiewicz, and P. P. Jagodzinski, "Decreased expression of cytochrome p450 1B1 in non-small cell lung cancer," *Biomedicine & Pharmacotherapy*, vol. 95, pp. 339–345, 2017.

[51] C. A. Wassenaar, Y. Ye, Q. Cai et al., "CYP2A6 reduced activity gene variants confer reduction in lung cancer risk in African American smokers—findings from two independent populations," *Carcinogenesis*, vol. 36, no. 1, pp. 99–103, 2015.

[52] D. Veale, N. Kerr, G. J. Gibson, P. J. Kelly, and A. L. Harris, "The relationship of quantitative epidermal growth factor receptor expression in non-small cell lung cancer to long term survival," *British Journal of Cancer*, vol. 68, no. 1, pp. 162–165, 1993.

[53] G. V. Scagliotti, G. Selvaggi, S. Novello, and F. R. Hirsch, "The biology of epidermal growth factor receptor in lung cancer," *Clinical Cancer Research*, vol. 10, no. 12, pp. 4227s–4232s, 2004.

[54] J. Ogawa, M. Iwazaki, H. Inoue, S. Koide, and A. Shohtsu, "Immunohistochemical study of glutathione-related enzymes and proliferative antigens in lung cancer. Relation to cisplatin sensitivity," *Cancer*, vol. 71, no. 7, pp. 2204–2209, 1993.

[55] T. P. Stricker, C. D. Brown, C. Bandlamudi et al., "Robust stratification of breast cancer subtypes using differential patterns of transcript isoform expression," *PLoS Genetics*, vol. 13, no. 3, 2017.

[56] F. Chen, M. Liu, Y. Yu et al., "LINC00958 regulated miR-627-5p/YBX2 axis to facilitate cell proliferation and migration in oral squamous cell carcinoma," *Cancer Biology & Therapy*, vol. 20, no. 9, pp. 1270–1280, 2019.

[57] Z. Verde, L. Reinoso, L. M. Chicharro et al., "Are SNP-smoking association studies needed in controls? DNA repair gene polymorphisms and smoking intensity," *PLoS One*, vol. 10, no. 5, 2015.

[58] H. Wang, M. Sun, J. Guo et al., "3-O-(Z)-coumaroyloleanolic acid overcomes *Cks1b*-induced chemoresistance in lung cancer by inhibiting Hsp90 and MEK pathways," *Biochemical Pharmacology*, vol. 135, pp. 35–49, 2017.

[59] M. A. Iqbal, S. Arora, G. Prakasam, G. A. Calin, and M. A. Syed, "MicroRNA in lung cancer: role, mechanisms, pathways and therapeutic relevance," *Molecular Aspects of Medicine*, vol. 70, pp. 3–20, 2019.

[60] Y. Lv, Z. Huang, Y. Lin et al., "MiRNA expression patterns are associated with tumor mutational burden in lung adenocarcinoma," *Oncoimmunology*, vol. 8, no. 10, 2019.

[61] J. J. Zhao, J. Yang, J. Lin et al., "Identification of miRNAs associated with tumorigenesis of retinoblastoma by miRNA microarray analysis," *Child's Nervous System*, vol. 25, no. 1, pp. 13–20, 2009.

[62] H. Liu, G. Pei, M. Song, S. Dai, and Y. Wang, "Influence of hsa-miR-6727-5p on the proliferation, apoptosis, invasion and migration of Caski, Hela and SiHa cervical cancer cells," *Journal of BUON*, vol. 22, no. 4, pp. 973–978, 2017.

[63] D. F. Pellatt, J. R. Stevens, R. K. Wolff et al., "Expression profiles of miRNA subsets distinguish human colorectal carcinoma and normal colonic mucosa," *Clinical and Translational Gastroenterology*, vol. 7, no. 3, 2016.

[64] Y. P. Hu, Y. P. Jin, X. S. Wu et al., "LncRNA-HGBC stabilized by HuR promotes gallbladder cancer progression by regulating miR-502-3p/SET/AKT axis," *Molecular Cancer*, vol. 18, no. 1, p. 167, 2019.

[65] M. Duruisseaux, A. Martínez-Cardús, M. E. Calleja-Cervantes et al., "Epigenetic prediction of response to anti-PD-1 treatment in non-small-cell lung cancer: a multicentre, retrospective analysis," *Respiratory Medicine*, vol. 6, no. 10, pp. 771–781, 2018.

[66] A. Etcheverry, M. Aubry, M. de Tayrac et al., "DNA methylation in glioblastoma: impact on gene expression and clinical outcome," *BMC Genomics*, vol. 11, no. 1, 2010.

[67] S. N. Udden, Y. T. Kwak, V. Godfrey et al., "NLRP12 suppresses hepatocellular carcinoma via downregulation of cJun N-terminal kinase activation in the hepatocyte," *eLife*, vol. 8, 2019.

[68] C. A. Eads, K. D. Danenberg, K. Kawakami et al., "MethyLight: a high-throughput assay to measure DNA methylation," *Nucleic Acids Research*, vol. 28, no. 8, pp. 32e–320, 2000.