Am J Transl Res 2021;13(3):1051-1063
www.ajtr.org /ISSN:1943-8141/AJTR0123998

Original Article

Identification and validation of significant gene mutations to predict clinical benefit of immune checkpoint inhibitors in lung adenocarcinoma

Ying Chen[1], Shaoyi Miao[2], Wancheng Zhao[3]

[1]*Department of Ultrasound, Xiaoshan Traditional Chinese Medical Hospital, Hangzhou 311200, China;* [2]*Department of Respiratory, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, Henan, China;* [3]*Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Shenyang 110000, China*

Received October 12, 2020; Accepted December 31, 2020; Epub March 15, 2021; Published March 30, 2021

Abstract: Objective: Immune checkpoint inhibitors (ICI) has achieved remarkable clinical benefit in advanced lung adenocarcinoma (LUAD). However, effective clinical use of ICI agents is encumbered by the high rate of innate resistance. The aim of our research is to identify significant gene mutations which can predict clinical benefit of immune checkpoint inhibitors in LUAD. Methods: The "mafComapre" function of "MafTools" package was used to screen the differentially mutated genes between durable clinical benefit (DCB) group and no durable clinical benefit (NDB) group based on the somatic mutation data from NSCLc_PD1_mSK_2018. Machine learning was performed to select significantly mutated genes to accurately classify patients into DCB group and NDB group. A nomogram model was constructed based on the significantly mutated genes to predict the susceptibility of patients to ICI. Finally, we explored the correlation between two classifications of immune cell infiltration, PD-1 and PD-L1 expression, tumor mutational burden (TMB) and prognosis. Results: Through utilize machine learning, 6 significantly mutated genes were obtained from 8 differentially mutated genes and used to accurately classify patients into DCB group and NDB group. The DCA curve and clinical impact curve revealed that the patients can benefit from the decisions made based on the nomogram model. Patients highly sensitive to ICI have elevated immune activity, higher expression of PD-1 and PD-L1, increased TMB, and well prognosis if they accept ICI treatment. Conclusions: Our research selected 6 significantly mutated genes that can predict clinical benefit of ICI in LUAD patients.

Keywords: Lung adenocarcinoma, immune checkpoint inhibitors, somatic mutations, nomogram, random forest model

Introduction

Lung cancer is one of the malignant tumors with the highest morbidity and mortality in the world [1]. Non-small cell lung cancer (NSCLC) accounts for 80-85% of all lung cancers, and adenocarcinoma is one of the important types of non-small cell lung cancer [2, 3]. Surgical combined with chemoradiotherapy is still the main treatment option for NSCLC. However, both the loss of surgical opportunity in advanced stage patients and the poor response to chemotherapy in patients with postoperative recurrence have promoted the development of non-operative and non-chemoradiotherapy treatment [4]. Among them, immune checkpoint inhibitors (ICI) are one of the main research hotspots at present. Immune checkpoint inhibitors, nivolumab and pembrolizumab have been approved for treatment in advanced and relapsed lung cancer patients by the Food and Drug Administration (FDA) and both of them can improve the survival outcomes of lung cancer patients [5-8]. However, ICI can also enhance normal immune response of the body while enhancing cellular immunity against tumor, leading to immune tolerance disorder of the body. Therefore, it is very important to find molecular biomarkers to predict clinical benefit of ICI in lung adenocarcinoma (LUAD).

Many potential biomarkers have been explored to predict the response of lung cancer patients to ICI, such as lymphocyte infiltration, the

expression level of PD-L1, and tumor mutation burden [9-12]. However, these biomarkers are also in a dynamic state as the tumor progresses [13, 14]. Therefore, it is urgent to develop clinically useful tools to identify patients most likely to benefit from ICB treatment. Neoantigens derived from gene mutations have recently become one of the hotspots in immunotherapy efficacy studies [15-17]. Malignant tumors with genetic mutations produce new sequences of peptides that can be processed by antigen-presenting cells (APC), and presented to T lymphocytes [18]. In general, patients with higher neoantigens are more likely to benefit from ICI [19, 20]. However, it is not clear neoantigens derive from which gene mutations will increase the response of LUAD patients to ICI. The aim of our research is to identify significant gene mutations to predict clinical benefit of ICI in LUAD.

Materials and methods

Data source

The Training dataset of this research, NSCLc_PD1_mSK_2018 is derived from cbioportal database (https://www.cbioportal.org/study/summary?id=nsclc_pd1_msk_2018). The data set contains 240 NSCLC samples, including 192 LUAD samples. Among the 192 LUAD patients, 179 LUAD patients received anti-PD-1/PD-L1 and (or) anti-CTLA4 treatment, including 55 patients with durable clinical benefit (DCB) and 124 patients with no durable clinical benefit (NDB). The Validation dataset, MIXed_allen_2018 is also available from the cbioportal database (https://www.cbioportal.org/study/summary?id=mixed_allen_2018). The data set contains somatic mutation data from 47 LUAD patients receiving anti-PD-1/PD-L1 treatment, including 20 patients with DCB and 27 patients with NDB. Another validation dataset, TCGA-LUAD dataset downloaded from GDC Data Portal, including 594 LUAD patients (https://portal.gdc.cancer.gov/projects/TCGA-LUAD). Somatic mutation data, mRNA expression profiling data and clinical information of LUAD patients were obtained from this database. Patients in TCGA-LUAD dataset were not treated with anti-PD-1/PD-L1 or anti-CTLA4.

Screening of differentially mutated genes

Patients in training dataset were divided into DCB and NDB groups. Somatic mutation data files were converted into MAF files. The "mafComapre" function of R software's "MafTools" package was used to screen the differentially mutated genes. Fisher test was performed and $P < 0.05$ was considered statistically significant [21].

Construction of the classification models using machine learning

Machine learning is a branch of artificial intelligence. It involves two important steps, first training with existing data, and then predicting unknown data [22]. In our research, we construct classification models to distinguish DCB and NDB groups using machine learning. We used "caret" package in R software to construct the random forest (RF) model, generalized linear model (GLM) and support vector machine model (SVM) respectively based on the training dataset. We then used the explain function of the "DALEX" package to perform an explanatory analysis of the three models. Finally, we choose the model with the smallest residual error as the best model.

Establishment and evaluation of the nomogram model

The "Nomogram" function of the "rms" package in R software was invoked to establish the nomogram model based on the variables selected by machine learning. The predicted result can be obtained by adding the corresponding score on the score axis of each variable. The calibration curve was used to verify the performance of the model. Decision curve analysis (DCA) and clinical impact curve were plotted to evaluate the clinical value of the nomogram model [23].

Estimate of tumor immune cell infiltration by CIBERSORT algorithm

We first used CIBERSORT algorithm to analyze the RNA expression profile data of LUAD samples from TCGA-LUAD dataset. We then downloaded the expression profiles of 547 genes of 22 immune cell types from CIBERSORT website as reference data (http://cibersort.stanford.edu/). Expression profile data of TCGA-LUAD dataset was compared with the reference data by deconvolution and support vector machine, the infiltration score of 22 immune cell types in each sample were obtained [24, 25].

Calculation of tumor mutational burden in LUAD patients

Tumor mutational burden (TMB) is the total number of non-synonymous mutations in each coding region of the tumor genome. In our research, the TMB of LUAD patients was calculated using the following formula: TMB = Sn × 1,000,000/n (Sn means the absolute value of somatic mutations, n means the number of exon base coverage depth ≥ 100×) [25].

Statistical analysis

One-way ANOVA and Kruskal-Wallis tests were used to compare the differences between the groups. A two-sided *p*-value of less than 0.05 was considered statistically significant. The survival curve was plotted by the Kaplan-Meier method, and the log-rank test was performed to compare the statistical significance of the survival difference between the two groups. All statistical analyses were performed using R 4.0.0.

Results

Somatic mutations in LUAD patients from the training dataset

We transferred the somatic mutation data files from training dataset into MAF files. The "maftools" package in R software was used to visualize the mutation information of LUAD hidden in the MAF file. Oncoplot shows the top 30 mutated genes in LUAD, with TP53 being the most frequently mutated gene (**Figure 1A**). Stacked bar plot shows the fraction of conversions in each LUAD sample. We can find that C > T is the main type of mutation in LUAD, and "Transversions" (Tv) was higher than "Transitions" (Ti) (**Figure 1B**).

Screening of differentially mutated genes

LUAD samples were classified into DCB and NDB groups. We found 8 differentially mutated genes between DCB and NDB groups, and the results were visualized by Forest plots (**Figure 2A**) and Co-onco plots (**Figure 2B**). Among the 8 differentially mutated genes, the mutation rate of ERBB4, ATRX, FAT1, KDM5C, ASXL1, AR and MGA in DCB group was higher than in NDB group, while SMAD4 was the opposite. Lollipop plot indicated that the group with a higher mutation rate holds more sequences of peptides (**Figure 3**).

Construction of the classification models by using machine learning

We established the RF, GLM and SVM model based on the 8 differentially mutated genes respectively to classify the LUAD samples in training dataset into DCB and NDB groups. The residual represents the deviation from the true value. The smaller the absolute value of residual, the more accurate the model is. The correct interpretation of "Reverse cumulative distribution of residual" graph is that a small number of samples contribute a lot of residuals. If the line is above, then it represents a large number of samples with a large residual. In **Figure 4A**, we found that a large number of samples have relatively small residuals in RF model compared to the other two models. The red dot of "Boxplots of residual" represents the mean value of residuals of all samples. **Figure 4B** revealed that the mean value of residuals in RF model was smaller than other models. The mean value of residuals in RF, GLM and SVM model was 0.510, 0.512, 0.612 respectively. ROC curves was plotted and the AUC of RF, GLM and SVM model was 0.90, 0.88, 0.83 respectively, which again demonstrates the superiority of the RF model. Therefore, the RF model was finally selected as the best model. The RF model ranks eight differentially mutated genes according to their contribution to classification (**Figure 4D**). In order to find the best combination of categorizing variables, we performed 10 fold cross validation. The cross validation curve revealed that the RF model has the highest accuracy when the top six variables (ERBB4, ATRX, FAT1, KDM5C, ASXL1 and AR) are selected (**Figure 4E**).

Construction of the nomogram model

We used the selected mutated genes by RF model to construct a nomogram model to predict the susceptibility of LUAD patients to ICI (**Figure 5A**). Calibration curve indicated that the well accuracy of the nomogram model (**Figure 5B**). The red lines of the DCA curve are above the gray and black lines from 0 to 1, which revealed that the LUAD patients can benefit from the decisions made based on the nomogram model (**Figure 5C**). The clinical impact curve indicated that the predicted number of

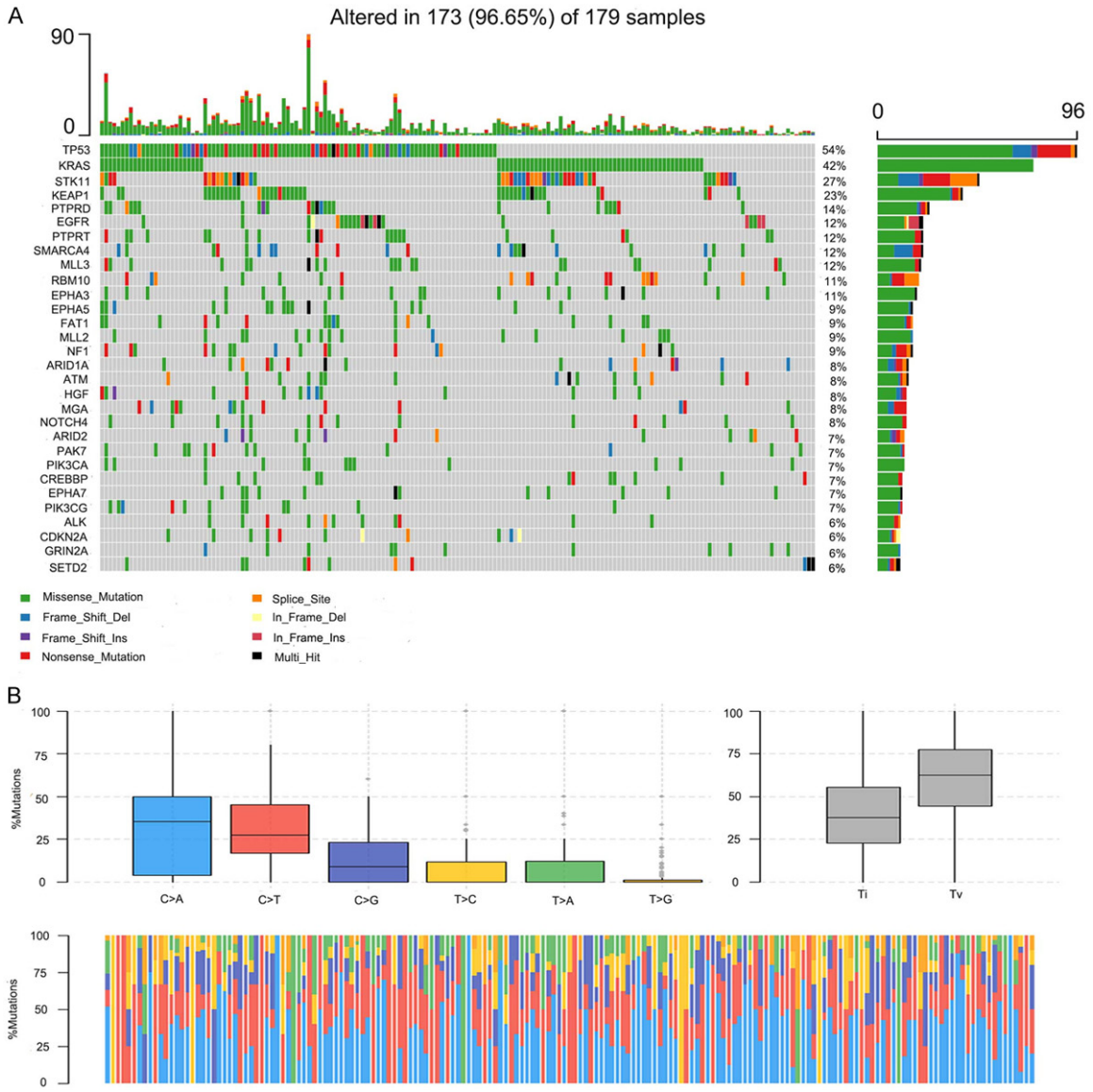


Figure 1. Somatic mutation profiling of lung adenocarcinoma (LUAD) patients in NSCLc_PD1_mSK_2018. A. Oncoplot shows the top 30 genes with the highest mutation frequency in 179 patients with LUAD from the training cohort. Each column shows one patient. The upper barplot represents the TMB of each sample. The right barplot represents the sample size with corresponding gene mutation. B. Transition and Transversions plot shows the distribution of SNPs in LUAD. Stacked bar plot shows the fraction of conversions in each sample.

high-risk patients was similar to the number of high-risk patients with an event from 0.25 to 1, which showed that the predictive power of the nomogram model was remarkable (**Figure 5D**).

Correlation between two classifications of immune cell infiltration, PD-1 and PD-L1 expression

According to the constructed nomogram model, patients in TCGA-LUAD dataset were divided into Sub1 group and Sub2 group. Patients in Sub2 group were predicted having higher sensitivity to ICI. We found that patients in Sub2 group have higher infiltration of CD8+T cells, activated CD4+T cells, activated NK cells and Macrophages M1 (**Figure 6**). In addition, the expression of PD-1 and PD-L1 was higher in Sub2 group than Sub1 group (**Figure 7**).

The DCB group have elevated TMB and prognosis

Multiple studies have shown that TMB is positively correlated with tumor immune checkpoint inhibitors response [26, 27]. Indeed, we

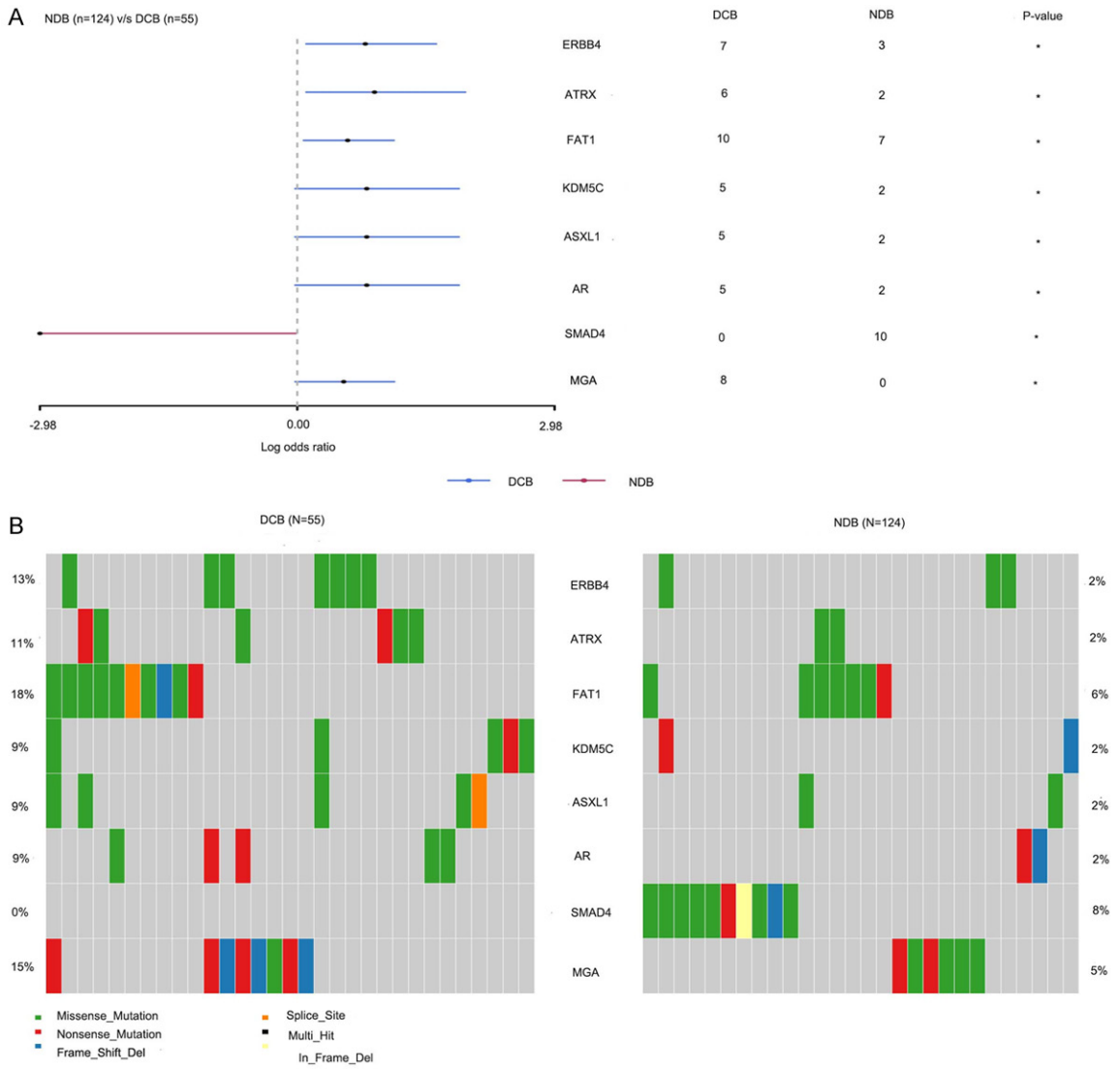


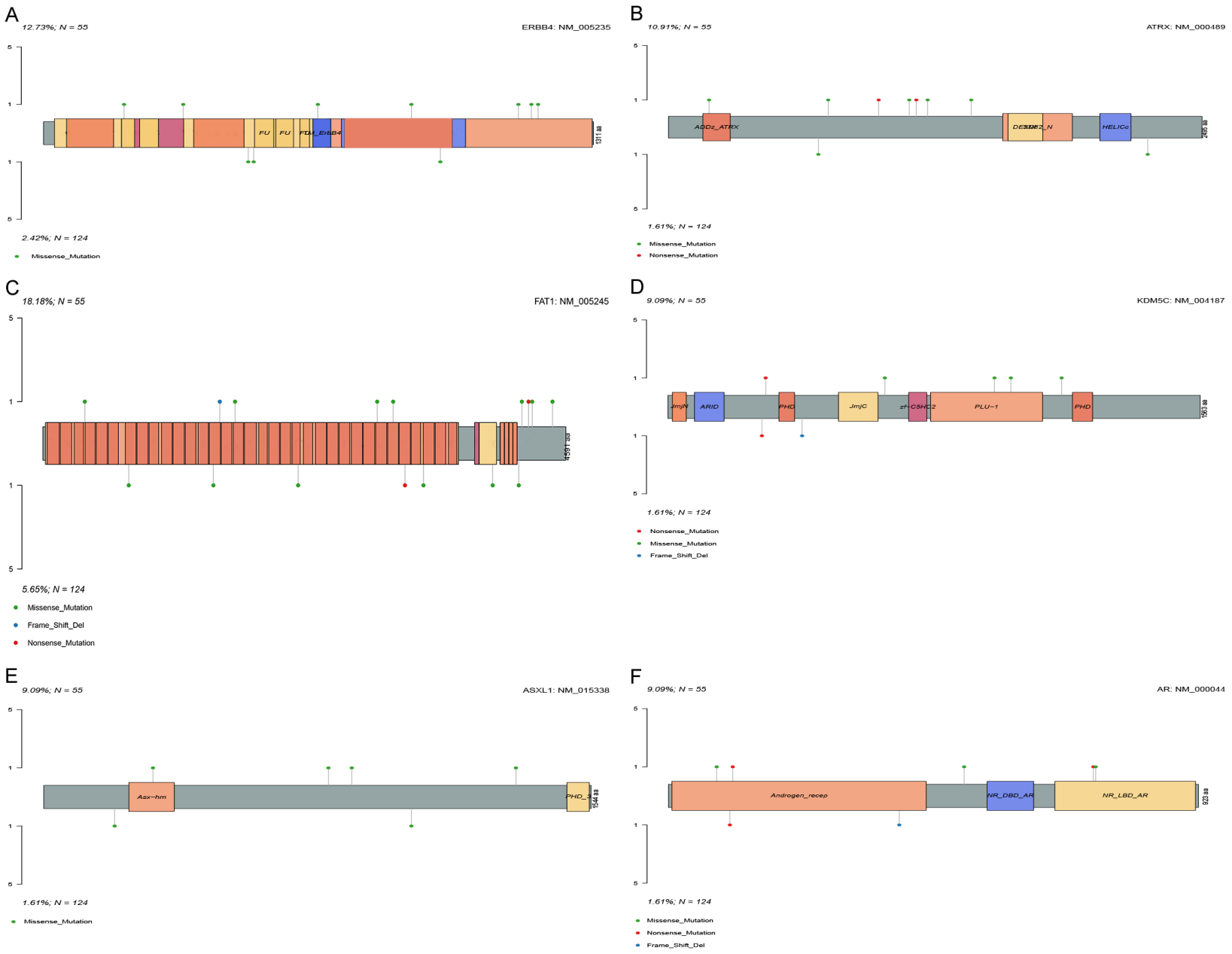
Figure 2. Differentially mutated genes. A. Forest plots show 8 differentially mutated genes (ERBB4, ATRX, FAT1, KDM5C, ASXL1, AR, MGA and SMAD4) between durable clinical benefit (DCB) group (n = 55) and no durable clinical benefit (NDB) group. (n = 124) using the training corhort. B. Co-onco plots show 8 differentially mutated genes between DCB group and NDB group. (*P < 0.05; **P < 0.01; ***P < 0.001).

observed that patients in DCB group have higher TMB than in NDB group (**Figure 8A** and **8B**). Patients in Sub2 group also have higher TMB than in Sub1 group, which proves the excellent accuracy of our prediction model from the side (**Figure 8C**). The Kaplan-Meier curve showed that the patients in DCB group have a better prognosis than in NDB group, while the patients in Sub2 group and Sub1 group have no statistical difference (**Figure 8D-F**).

Discussion

In this research, we selected 6 significantly mutated genes that can effectively distinguish LUAD patients received ICI therapy into DCB group (Sub2) and NDB (Sub1) group based on RF classifications. Patients in Sub2 group were significantly associated with elevated immune activity, increased TMB, and over-expression of PD-1 and PD-L1. Patients highly sensitive to ICI can significantly extend their survival time by receiving ICI therapy. These results suggest that the classification model constructed based on the 6 significantly mutated genes can effectively identify a group of patients which benefit from ICI therapy.

Previous research has reported a set of biomarkers to predict the response of ICI to thera-


A
12.73%; N = 55
ERBB4: NM_005235
2.42%; N = 124
Missense_Mutation
B
10.91%; N = 55
ATRX: NM_000489
ADDz_ATRX
HELICc
1.61%; N = 124
Missense_Mutation
Nonsense_Mutation
C
18.18%; N = 55
FAT1: NM_005245
5.65%; N = 124
Missense_Mutation
Frame_Shift_Del
Nonsense_Mutation
D
9.09%; N = 55
KDM5C: NM_004187
JmjN
ARID
PHD
JmjC
PLU-1
PHD
1.61%; N = 124
Nonsense_Mutation
Missense_Mutation
Frame_Shift_Del
E
9.09%; N = 55
ASXL1: NM_015338
Asx-hm
1.61%; N = 124
Missense_Mutation
F
9.09%; N = 55
AR: NM_000044
Androgen_recep
NR_LBD_AR
1.61%; N = 124
Missense_Mutation
Nonsense_Mutation
Frame_Shift_Del

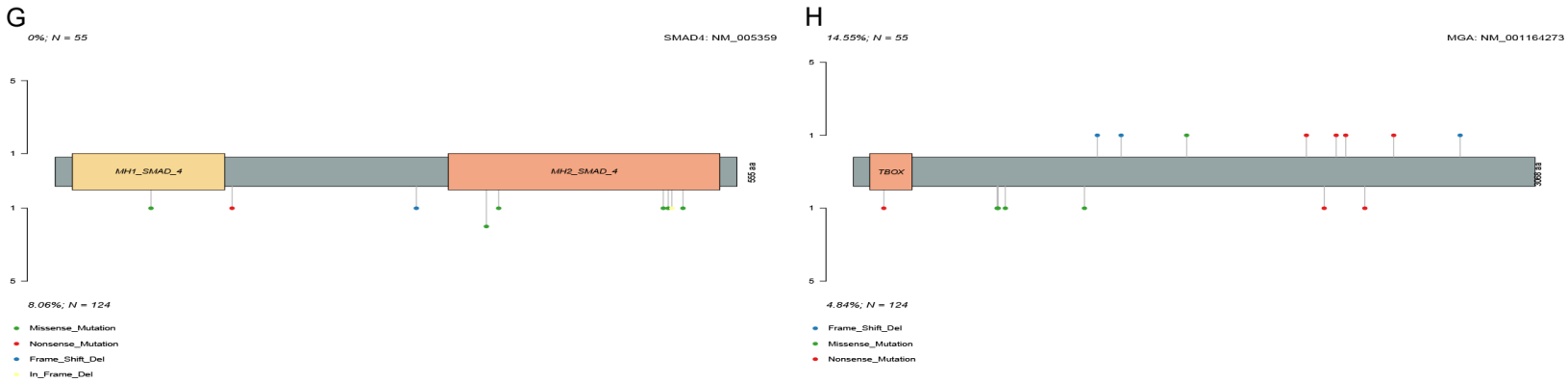


Figure 3. Lollipop Plot shows the differentially sequences of peptides of 8 mutated genes between DCB group and NDB group using the training corhort. The DCB group is above the horizontal axis, and NDB group is below the horizontal axis. The lollipops on the horizontal axis represents different variant type (A). ERBB4 (B). ATRX (C). FAT1 (D). KDM5C (E). ASXL1 (F). AR (G). SMAD4 (H). MGA.

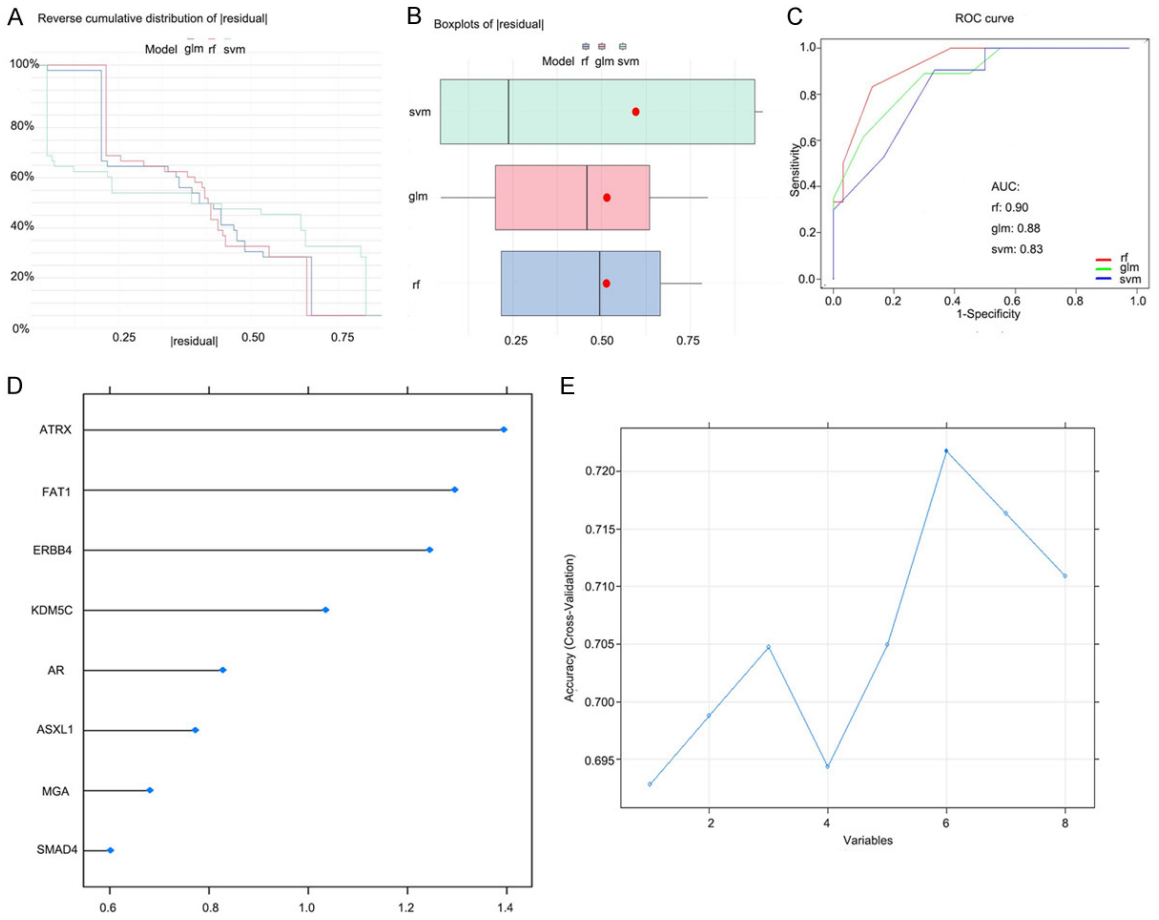


Figure 4. Construction of the classification models using machine learning based on the training dataset. A. The residual represents the deviation from the true value. The smaller the absolute value of residual, the more accurate the model is. Reverse cumulative distribution of residual was plotted to show the residual distribution of samples in random forest (RF) model, generalized linear model (GLM) and support vector machine model (SVM). B. Boxplots of residual was plotted to show the residual distribution of RF, GLM and SVM model. The red dot represents the mean value of residuals of all samples. C. ROC curves was plotted and the AUC of RF, GLM and SVM model was 0.90, 0.88, 0.83 respectively. D. RF model was performed to show the importance of 8 differentially mutated genes to distinguish the DCB group and NDB group. The horizontal axis shows the importance score of genes. E. Cross validation curve was plotted to show the accuracy of the RF model from variable = 1 to 8. The horizontal axis shows the number of variables. The vertical axis shows the accuracy of the model corresponding to the number of variables.

py [28-30]. However, most research was focused on single variable, and the non-unified cutoff values make the research results impractical. Considering the heterogeneity of tumor, the predictive ability of a single gene mutation was weak. We used differentially mutated genes to train the performance of the classification model, and finally found the most accurate mutated gene set to predict the sensitivity of LUAD patients with ICI therapy. We found the RF classifications can not only accurately predict DCB and NDB in LUAD, but also not be constrained by the optimal cut-off value. It suggests that the new classification algorithm is not only easy to generalize, but also can identify potentially beneficiaries of ICI therapy.

Several studies reported that tumors with presence CD8 T-Cell infiltration and higher expression of PD-L1 can benefit from ICI therapy [31, 32]. Research of KEYNOTE-010 revealed that the advanced lung cancer patients with high expression of PD-L1 had a significantly better efficacy to pembrolizumab than those with low expression of PD-L1 [33]. In our research, we also proved that patient sensitivity to ICI therapy has high expression of PD-1 and PD-L1. As expected, patients in Sub2 group have higher

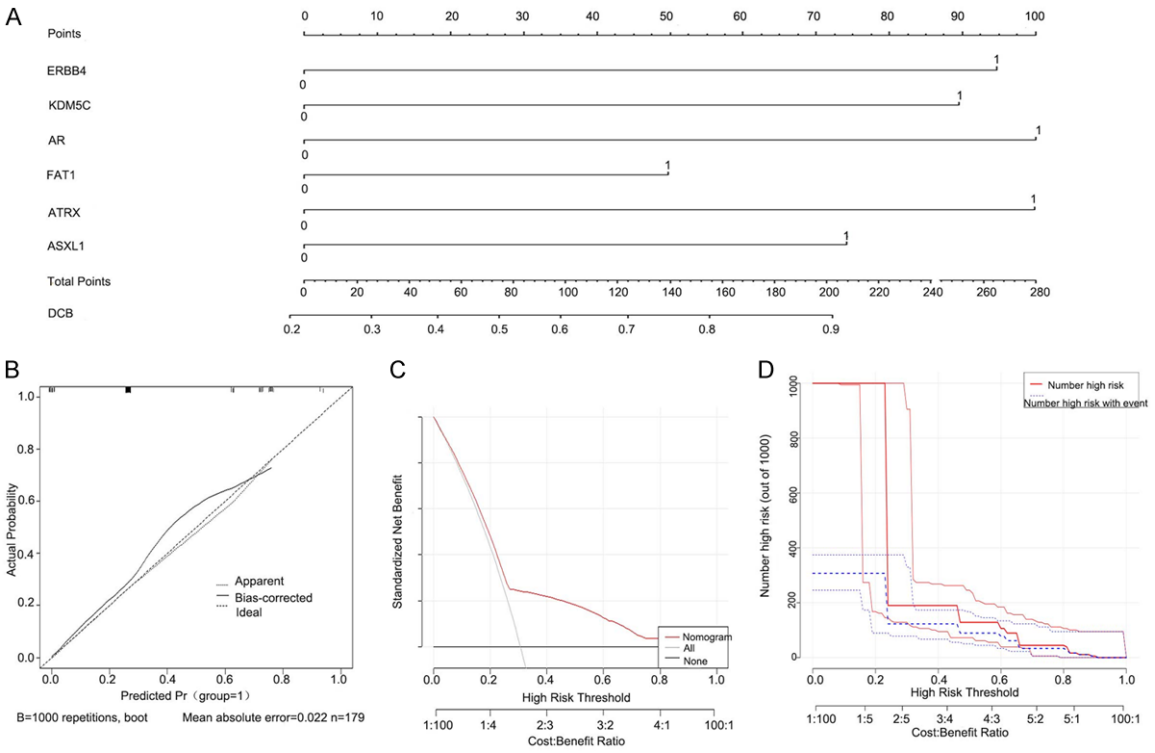


Figure 5. Construction of the nomogram model. A. Construction of the nomogram model based on the 6 significantly mutated genes (ERBB4, ATRX, FAT1, KDM5C, ASXL1 and AR). Vertical line passes through the value of each gene, and obtains a integral on the integral line at the top of the nomogram. The integrals of all the genes are added to obtain a total score to predict the benefit to ICIs at the bottom of the nomogram. B. The calibration curve revealed the predictiveness of the nomogram model. The "Apparent" curve closed to the "Ideal" curve represents that the nomogram model is accurate in predicting the sensitivity of patients to ICIS treatment. C. The DCA curve evaluated the clinical value of the nomogram model. The x-axis indicates the predicted probability, and the y-axis represents the net benefit. D. The clinical impact curve was used to assess the clinical impact of the nomogram model. The "Number high risk" curve was closed to the "Number high risk with event" curve at high risk threshold from 0.23 to 1.

infiltration of CD8+T cells than in Sub1 group [34]. Consistent with the research of Jie Peng, activated NK cells and Macrophages M1 were high infiltration in Sub2 group than Sub1 group [35]. CD4+T cells, as a driving factor of antitumor immunity, were elevated activated in Sub2 group than in Sub1 group, which is consistent with previous research results [34]. Moreover, we found that patients in DCB group have increased TMB than in NDB group, which indicated that the patients in DCB group exhibit higher neoantigens. The RF model based on the 6 significantly mutated genes also found that the patients in Sub2 group have increased TMB, which verifies the accuracy of our classification model to some extent. Finally, the survival curve showed that the patients in DCB group have longer survival time than in NDB group. However, patients in Sub2 group and Sub1 group have no statistical difference, which suggested that the patients highly sensitive to ICI can only improve their survival time by receiving ICI therapy.

Conclusions

In conclusion, our research revealed that the classification model of RF based on 6 significantly mutated genes can predict the sensitivity of the LUAD patients with ICI therapy. The research may provide a new anticancer immunotherapy strategy to LUAD.

Acknowledgements

We acknowledge the authors who provided information for the public database in cbioportal database and TCGA database. This article was funded by the corresponding author Hongyun Li.

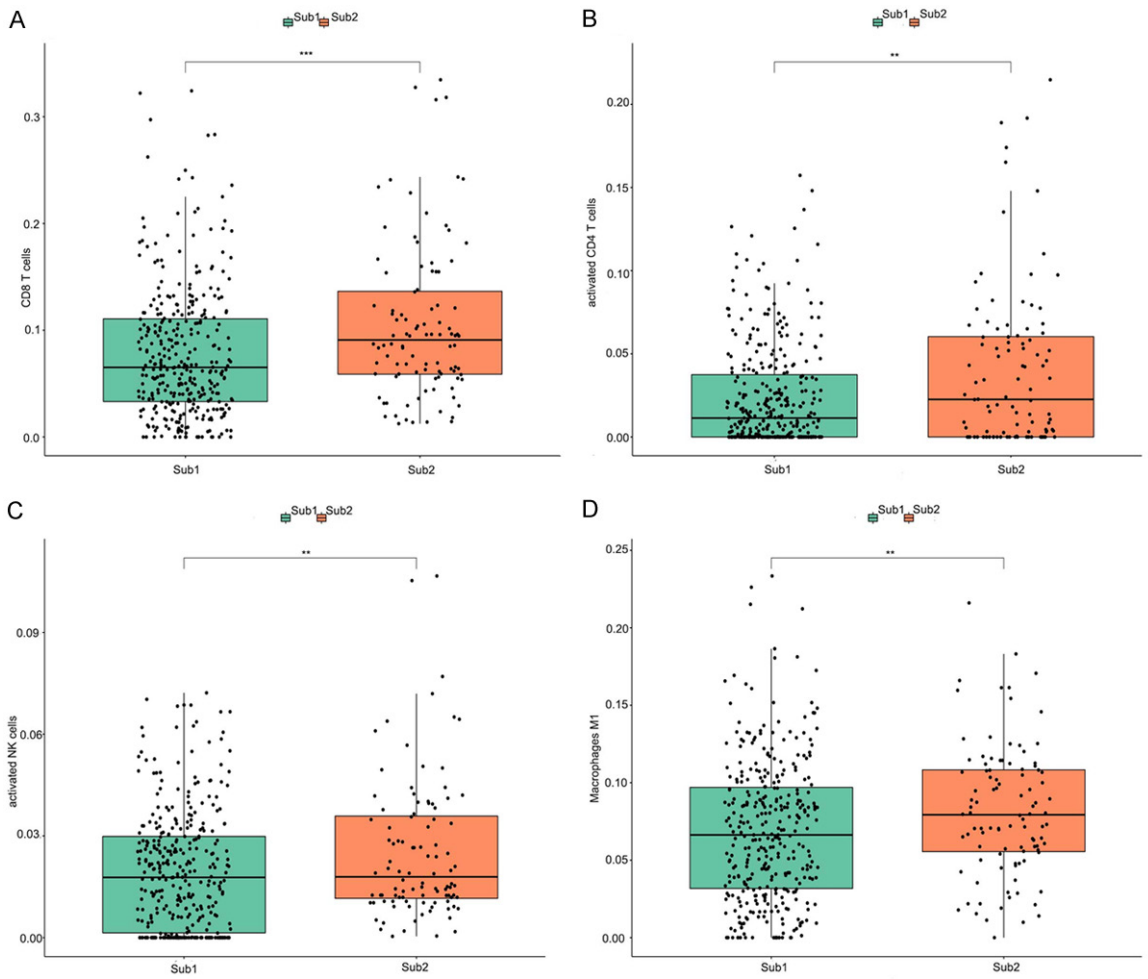


Figure 6. Sub2 group is associated with elevated immune activity in LUAD. (A) CD8+ T cells (B) activated CD4+ T cells (C) activated NK cells (D). Macrophages M1 showed significantly higher enrichment levels in Sub2 group than in Sub1 group. (*P < 0.05; **P < 0.01; ***P < 0.001).

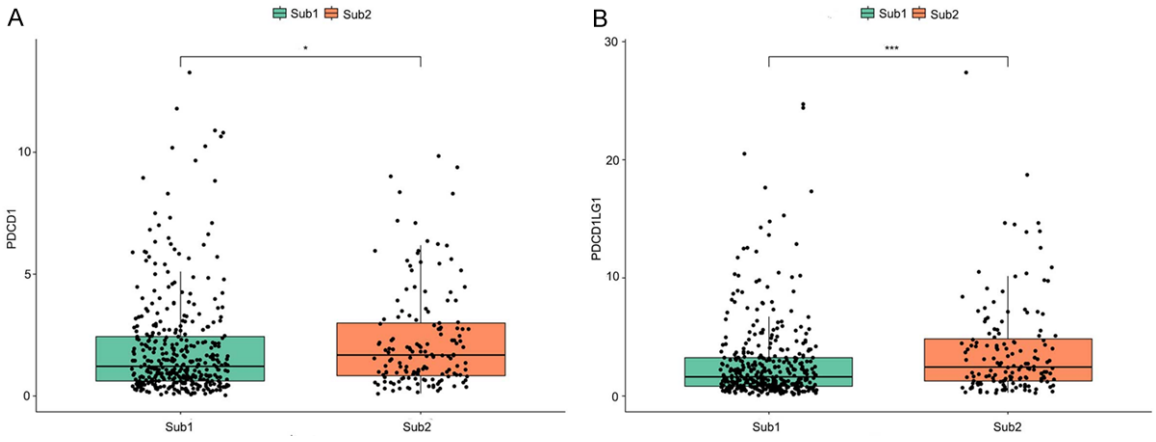


Figure 7. Sub2 group is associated with elevated PD-1 and PD-L1 Expression in LUAD. A. Sub2 group patients have significantly higher PD-1 expression levels than Sub2 group patients. B. Sub2 group patients have significantly higher PD-L1 expression levels than Sub2 group patients. (*P < 0.05; **P < 0.01; ***P < 0.001).

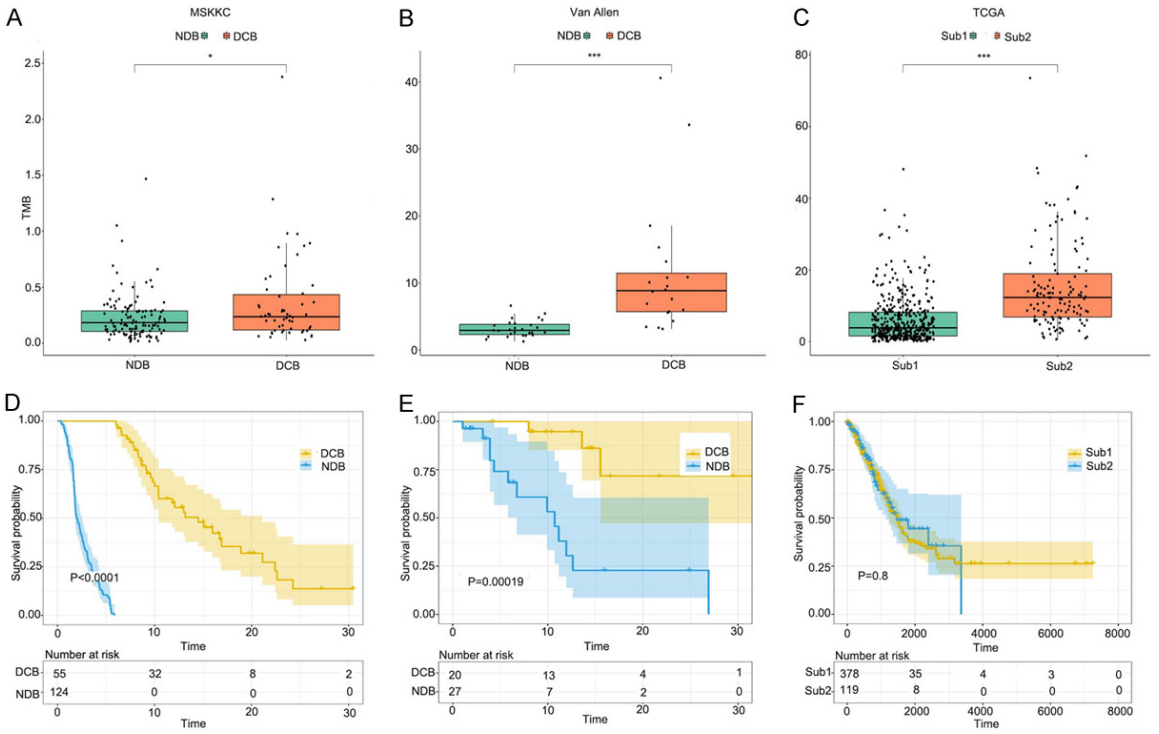


Figure 8. DCB group has increased tumor mutational burden (TMB) and well prognosis. A. DCB group patients have significantly higher TMB than NDB group patients in NSCLc_PD1_mSK_2018. B. DCB group patients have significantly higher TMB than NDB group patients in MIXed_allen_2018. C. Sub2 group patients have significantly higher TMB than Sub2 group patients in TCGA-LUSD corhort. D. DCB group patients have a longer survival time than NDB group patients in NSCLc_PD1_mSK_2018. E. DCB group patients have a longer survival time than NDB group patients in MIXed_allen_2018. F. Survival time have no statistically significant between Sub2 group and Sub1 group. (*P < 0.05; **P < 0.01; ***P < 0.001).

Disclosure of conflict of interest

None.

Address correspondence to: Wancheng Zhao, Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, NO. 36, Sanhao Road, Shenyang 110000, China. E-mail: 373-986045@qq.com

References

[1] Wu YL, Lu S, Cheng Y, Zhou C, Wang J, Mok T, Zhang L, Tu HY, Wu L, Feng J, Zhang Y, Luft AV, Zhou J, Ma Z, Lu Y, Hu C, Shi Y, Baudelet C, Cai J and Chang J. Nivolumab versus docetaxel in a predominantly chinese patient population with previously treated advanced nsclc: checkmate 078 randomized phase iii clinical trial. J Thorac Oncol 2019; 14: 867-875.

[2] Yu H, Pang Z, Li G and Gu T. Bioinformatics analysis of differentially expressed miRNAs in non-small cell lung cancer. J Clin Lab Anal 2020; e23588.

[3] Tian B, Han X, Li G, Jiang H, Qi J, Li J, Tian Y and Wang C. A long intergenic non-coding RNA, LINC01426, promotes cancer progression via AZGP1 and predicts poor prognosis in patients with LUAD. Mol Ther Methods Clin Dev 2020; 18: 765-780.

[4] Liu S, Zhou F, Liu Z, Xiong A, Jia Y, Zhao S, Zhao C, Li X, Jiang T, Han R, Qiao M, Liu Y, He Y, Li J, Li W, Gao G, Ren S, Su C and Zhou C. Predictive and prognostic significance of M descriptors of the 8th TNM classification for advanced NSCLC patients treated with immune checkpoint inhibitors. Transl Lung Cancer Res 2020; 9: 1053-1066.

[5] Antonia SJ, Villegas A, Daniel D, Vicente D, Murakami S, Hui R, Yokoi T, Chiappori A, Lee KH, de Wit M, Cho BC, Bourhaba M, Quantin X, Tokito T, Mekhail T, Planchard D, Kim YC, Karapetis CS, Hiret S, Ostoros G, Kubota K, Gray JE, Paz-Ares L, de Castro Carpeño J, Wadsworth C, Melillo G, Jiang H, Huang Y, Dennis PA and Özgüroğlu M. Durvalumab after chemoradiotherapy in stage iii non-small-cell lung cancer. N Engl J Med 2017; 377: 1919-1929.

[6] Carbone DP, Reck M, Paz-Ares L, Creelan B, Horn L, Steins M, Felip E, van den Heuvel MM, Ciuleanu TE, Badin F, Ready N, Hiltermann TJN, Nair S, Juergens R, Peters S, Minenza E, Wrangle JM, Rodriguez-Abreu D, Borghaei H, Blumenschein GR Jr, Villaruz LC, Havel L, Krejci J, Corral Jaime J, Chang H, Geese WJ, Bhagavatheeswaran P, Chen AC and Socinski MA. First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. N Engl J Med 2017; 376: 2415-2426.

[7] Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, Domine M, Clingan P, Hochmair MJ, Powell SF, Cheng SY, Bischoff HG, Peled N, Grossi F, Jennens RR, Reck M, Hui R, Garon EB, Boyer M, Rubio-Viqueira B, Novello S, Kurata T, Gray JE, Vida J, Wei Z, Yang J, Raftopoulos H, Pietanza MC and Garassino MC. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. N Engl J Med 2018; 378: 2078-2092.

[8] Pai-Scherf L, Blumenthal GM, Li H, Subramaniam S, Mishra-Kalyani PS, He K, Zhao H, Yu J, Paciga M, Goldberg KB, McKee AE, Keegan P and Pazdur R. FDA approval summary: pembrolizumab for treatment of metastatic non-small cell lung cancer: first-line therapy and beyond. Oncologist 2017; 22: 1392-1399.

[9] Shukuya T and Carbone DP. Predictive markers for the efficacy of anti-PD-1/PD-L1 antibodies in lung cancer. J Thorac Oncol 2016; 11: 976-988.

[10] Martinez P, Peters S, Stammers T and Soria JC. Immunotherapy for the first-line treatment of patients with metastatic non-small cell lung cancer. Clin Cancer Res 2019; 25: 2691-2698.

[11] Patel SP and Kurzrock R. PD-L1 Expression as a predictive biomarker in cancer immunotherapy. Mol Cancer Ther 2015; 14: 847-856.

[12] Tang H, Wang Y, Chlewicki LK, Zhang Y, Guo J, Liang W, Wang J, Wang X and Fu YX. Facilitating T cell infiltration in tumor microenvironment overcomes resistance to PD-L1 blockade. Cancer Cell 2016; 29: 285-296.

[13] Caswell DR and Swanton C. The role of tumour heterogeneity and clonal cooperativity in metastasis, immune evasion and clinical outcome. BMC Med 2017; 15: 133.

[14] Blomberg OS, Spagnuolo L and de Visser KE. Immune regulation of metastasis: mechanistic insights and therapeutic opportunities. Dis Model Mech 2018; 11: 10.

[15] Bobisse S, Foukas PG, Coukos G and Harari A. Neoantigen-based cancer immunotherapy. Ann Transl Med 2016; 4: 262.

[16] Balachandran VP, Łuksza M, Zhao JN, Makarov V, Moral JA, Remark R, Herbst B, Askan G, Bhanot U, Senbabaoglu Y, Wells DK, Cary CIO, Grbovic-Huezo O, Attiyeh M, Medina B, Zhang J, Loo J, Saglimbeni J, Abu-Akeel M, Zappasodi R, Riaz N, Smoragiewicz M, Kelley ZL, Basturk O; Australian Pancreatic Cancer Genome Initiative; Garvan Institute of Medical Research; Prince of Wales Hospital; Royal North Shore Hospital; University of Glasgow; St Vincent's Hospital; QIMR Berghofer Medical Research Institute; University of Melbourne, Centre for Cancer Research; University of Queensland, Institute for Molecular Bioscience; Bankstown Hospital; Liverpool Hospital; Royal Prince Alfred Hospital, Chris O'Brien Lifehouse; Westmead Hospital; Fremantle Hospital; St John of God Healthcare; Royal Adelaide Hospital; Flinders Medical Centre; Envoi Pathology; Princess Alexandria Hospital; Austin Hospital; Johns Hopkins Medical Institutes; ARC-Net Centre for Applied Research on Cancer, Gönen M, Levine AJ, Allen PJ, Fearon DT, Merad M, Gnjatic S, Iacobuzio-Donahue CA, Wolchok JD, DeMatteo RP, Chan TA, Greenbaum BD, Merghoub T and Leach SD. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. Nature 2017; 551: 512-516.

[17] Pan D, Kobayashi A, Jiang P, Ferrari de Andrade L, Tay RE, Luoma AM, Tsoucas D, Qiu X, Lim K, Rao P, Long HW, Yuan GC, Doench J, Brown M, Liu XS and Wucherpfennig KW. A major chromatin regulator determines resistance of tumor cells to T cell-mediated killing. Science 2018; 359: 770-775.

[18] van Rooij N, van Buuren MM, Philips D, Velds A, Toebes M, Heemskerk B, van Dijk LJ, Behjati S, Hilkmann H, El Atmioui D, Nieuwland M, Stratton MR, Kerkhoven RM, Kesmir C, Haanen JB, Kvistborg P and Schumacher TN. Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. J Clin Oncol 2013; 31: e439-442.

[19] Matsushita H, Sato Y, Karasaki T, Nakagawa T, Kume H, Ogawa S, Homma Y and Kakimi K. Neoantigen load, antigen presentation machinery, and immune signatures determine prognosis in clear cell renal cell carcinoma. Cancer Immunol Res 2016; 4: 463-471.

[20] Howitt BE, Shukla SA, Sholl LM, Ritterhouse LL, Watkins JC, Rodig S, Stover E, Strickland KC, D'Andrea AD, Wu CJ, Matulonis UA and Konstantinopoulos PA. Association of polymerase e-mutated and microsatellite-instable endometrial cancers with neoantigen load, number of tumor-infiltrating lymphocytes, and expression of PD-1 and PD-L1. JAMA Oncol 2015; 1: 1319-1323.

[21] Madan V, Shyamsunder P, Han L, Mayakonda A, Nagata Y, Sundaresan J, Kanojia D, Yoshida K, Ganesan S, Hattori N, Fulton N, Tan KT, Alpermann T, Kuo MC, Rostami S, Matthews J, Sanada M, Liu LZ, Shiraishi Y, Miyano S, Chen-

damarai E, Hou HA, Malnassy G, Ma T, Garg M, Ding LW, Sun QY, Chien W, Ikezoe T, Lill M, Biondi A, Larson RA, Powell BL, Lübbert M, Chng WJ, Tien HF, Heuser M, Ganser A, Koren-Michowitz M, Kornblau SM, Kantarjian HM, Nowak D, Hofmann WK, Yang H, Stock W, Ghavamzadeh A, Alimoghaddam K, Haferlach T, Ogawa S, Shih LY, Mathews V and Koeffler HP. Comprehensive mutational analysis of primary and relapse acute promyelocytic leukemia. Leukemia 2016; 30: 1672-1681.

[22] Deo RC. Machine learning in medicine. Circulation 2015; 132: 1920-1930.

[23] Park SY. Nomogram: an analogue tool to deliver digital knowledge. J Thorac Cardiovasc Surg 2018; 155: 1793.

[24] Ali HR, Chlon L, Pharoah PD, Markowetz F and Caldas C. Patterns of immune infiltration in breast cancer and their clinical implications: a gene-expression-based retrospective study. PLoS Med 2016; 13: e1002194.

[25] Bi F, Chen Y and Yang Q. Significance of tumor mutation burden combined with immune infiltrates in the progression and prognosis of ovarian cancer. Cancer Cell Int 2020; 20: 373.

[26] Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, Stephens PJ, Daniels GA and Kurzrock R. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. Mol Cancer Ther 2017; 16: 2598-2608.

[27] Wang X and Li M. Correlate tumor mutation burden with immune signatures in human cancers. BMC Immunol 2019; 20: 4.

[28] Campbell JD, Alexandrov A, Kim J, Wala J, Berger AH, Pedamallu CS, Shukla SA, Guo G, Brooks AN, Murray BA, Imielinski M, Hu X, Ling S, Akbani R, Rosenberg M, Cibulskis C, Ramachandran A, Collisson EA, Kwiatkowski DJ, Lawrence MS, Weinstein JN, Verhaak RG, Wu CJ, Hammerman PS, Cherniack AD, Getz G, Artyomov MN, Schreiber R, Govindan R and Meyerson M. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat Genet 2016; 48: 607-616.

[29] Brahmer JR, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, Gottfried M, Peled N, Tafreshi A, Cuffe S, O'Brien M, Rao S, Hotta K, Zhang J, Lubiniecki GM, Deitz AC, Rangwala R and Reck M. Health-related quality-of-life results for pembrolizumab versus chemotherapy in advanced, PD-L1-positive NSCLC (KEYNOTE-024): a multicentre, international, randomised, open-label phase 3 trial. Lancet Oncol 2017; 18: 1600-1609.

[30] Biton J, Mansuet-Lupo A, Pécuchet N, Alifano M, Ouakrim H, Arrondeau J, Boudou-Rouquette P, Goldwasser F, Leroy K, Goc J, Wislez M, Germain C, Laurent-Puig P, Dieu-Nosjean MC, Cremer I, Herbst R, Blons H and Damotte D. TP53, STK11, and EGFR mutations predict tumor immune profile and the response to anti-PD-1 in lung adenocarcinoma. Clin Cancer Res 2018; 24: 5710-5723.

[31] Teng MW, Ngiow SF, Ribas A and Smyth MJ. Classifying cancers based on T-cell infiltration and PD-L1. Cancer Res 2015; 75: 2139-2145.

[32] Ock CY, Keam B, Kim S, Lee JS, Kim M, Kim TM, Jeon YK, Kim DW, Chung DH and Heo DS. Pan-cancer immunogenomic perspective on the tumor microenvironment based on PD-L1 and CD8 T-cell infiltration. Clin Cancer Res 2016; 22: 2261-2270.

[33] Herbst RS, Baas P, Kim DW, Felip E, Pérez-Gracia JL, Han JY, Molina J, Kim JH, Arvis CD, Ahn MJ, Majem M, Fidler MJ, de Castro G Jr, Garrido M, Lubiniecki GM, Shentu Y, Im E, Dolled-Filhart M and Garon EB. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet 2016; 387: 1540-1550.

[34] Tunger A, Sommer U, Wehner R, Kubasch AS, Grimm MO, Bachmann MP, Platzbecker U, Bornhäuser M, Baretton G and Schmitz M. The evolving landscape of biomarkers for anti-PD-1 or anti-PD-L1 therapy. J Clin Med 2019; 8: 1534.

[35] Peng J, Zou D, Gong W, Kang S and Han L. Deep neural network classification based on somatic mutations potentially predicts clinical benefit of immune checkpoint blockade in lung adenocarcinoma. Oncoimmunology 2020; 9: 1734156.