Original Article

Molecular gene mutation profiles, TMB and the impact of prognosis in Caucasians and east Asian patients with lung adenocarcinoma

Weiquan Gu[1#], Na Wang[2,3,4#], Weiguang Gu[5,6#], Yuan Qiu[7], Hua Zhang[8], Jianmiao Liang[8], Tongfei Zhou[8], Liheng Ma[9], Weijing Cai[10], Weineng Feng[8], Likun Chen[2,3,4]

[1]Department of Thoracic Surgery, The First People's Hospital of Foshan, Foshan 528000, China; [2]Department of Medical Oncology, [3]State Key Laboratory of Oncology in South China, [4]Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, China; [5]Department of Medical Oncology, Nanhai District People's Hospital of Foshan, Foshan 528200, China; [6]The Second School of Clinical Medical, Southern Medical University, Guangzhou 510280, China; [7]The First Affiliated Hospital of Guangzhou Medical University, Guangzhou 510120, China; [8]Department of Head and Neck Cancer, the First People's Hospital of Foshan, Foshan 528000, China; [9]Pfizer pharmaceuticals Ltd., New York, USA; [10]Shanghai Tongshu Biotechnology Co., Ltd., Shanghai 200444, China

Contributions: (I) Conception and design: L Chen, W Feng; (II) Administrative support: L Chen; (III) Provision of study materials or patients: W Gu; (IV) Collection and assembly of data: N Wang; (V) Data analysis and interpretation: W Gu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Likun Chen. Department of Medical Oncology, Sun Yat-sen University Cancer Center, No. 651, Dongfeng East Road, Yuexiu District, Guangzhou 510060, China. Email: chenlk@sysucc.org.cn; Weineng Feng. Department of Head and Neck Cancer, The First People's Hospital of Foshan, No. 81, Lingnan Avenue North, Chancheng District, Foshan 528000, China. Email: fengweineng@163.com.

Background: The difference in molecular gene mutation profile, tumor mutational burden (TMB) and their prognostic effects in lung adenocarcinoma between different ethnic groups are still unknown. A retrospective analysis was used to investigate the differences in lung adenocarcinoma driver gene mutations, TMB, and their impact on prognosis across different ethnic groups.

Methods: The incidence of epidermal growth factor receptor (*EGFR*) mutations and follow-up data of 647 Chinese lung adenocarcinoma patients were compared with the data from 522 Caucasian patients in The Cancer Genome Atlas (TCGA) database. Moreover, a comprehensive analysis was performed to compare the differences in gene mutation frequency, signaling pathway variation, and TMB using the whole-exome sequencing (WES) data of Chinese patients with that of Caucasian patients.

Results: A comparison of tumor signaling pathways and gene mutation profiles between Caucasians and Chinese revealed ethnic variations in the incidence of mutations in TGF-β and RTK-RAS signaling pathways, with P values of 0.012 and 0.016, respectively. In the Caucasian population, the mutations in 5 signaling pathways and 18 genes were all significantly correlated with TMB, whereas in the Chinese population, only mutations in the Notch pathway and 6 genes were found to be associated with TMB-high. *EGFR* mutations showed a better prognosis in Chinese patients with lung adenocarcinoma, while the opposite was found in Caucasians patients.

Conclusions: Variations in the incidence of mutations in signaling pathways involved in lung adenocarcinoma and the correlation of the signaling pathways with TMB may exist across different ethnic groups.

Keywords: Whole-exome sequencing (WES); tumor mutational burden (TMB); gene mutation profile; lung adenocarcinoma

Submitted Feb 22, 2020. Accepted for publication Apr 22, 2020.
doi: 10.21037/tlcr-20-457
View this article at: http://dx.doi.org/10.21037/tlcr-20-457

Introduction

Lung cancer is the leading cause of cancer-related deaths worldwide. The primary pathological type of lung cancer is non-small cell lung cancer (NSCLC), and most NSCLC is classified as lung adenocarcinoma, which accounts for more than 50% of lung cancers (1). With a better understanding of the genetic profiles of lung adenocarcinoma, targeted therapy has become an essential treatment for this cancer type. Epidermal growth factor receptor (*EGFR*) mutations are the most critical driver mutations. The incidence of EGFR mutations is approximately 11% in Caucasian patients with lung cancer and 27–34% in Asian patients, and as high as 40–60% in Asian lung adenocarcinoma patients (2,3). The clinical application of EGFR-tyrosine kinase inhibitors (EGFR-TKIs), including erlotinib and gefitinib, has greatly prolonged the survival time of NSCLC patients with EGFR mutations (4,5).

Tumor mutational burden (TMB), is the total number of nonsynonymous mutations (including point mutation and insertion/deletion mutation) per megabase somatic base in the tumor tissue, that is, the mutation density of the tumor gene (6). Generally speaking, the more mutations the lesion harbors, the higher the TMB and the better the tumor response to immunotherapy (7). For most tumors, the effectiveness of treatment with PD-1 inhibitors was almost linear with TMB (8). Moreover, for certain drugs (such as Opdivo), compared to PD-L1 level, the choice of TMB as a biomarker for predicting the efficacy of NSCLC treatment can better differentiate the beneficial population.

With the popularity of immunotherapy and detailed studies of TMB, a significant negative correlation has been found between EGFR mutation status and high TMB (TMB-H). However, the incidence of mutations in the genes of each signaling pathway and their correlation with TMB is still unclear. The role of *EGFR* mutations and TMB in prognosis is also inconclusive (9).

Therefore, this study reviewed the *EGFR* test results and survival data of 647 Chinese patients with lung adenocarcinoma. The whole-exome sequencing (WES) and survival data of 454 cases of lung adenocarcinoma were downloaded from The Cancer Genome Atlas (TCGA) database. These two groups were stratified according to disease stage and ethnic group, followed by a comparative analysis. Since the above conclusions are based on different genetic testing methods, in this study, WES was performed on tumor tissues from 49 Chinese patients with advanced lung adenocarcinoma. Those data were compared with the WES data from Caucasian patients in the TCGA database for a further head-to-head comparative analysis. The results showed that tumor signaling pathways and their relationship with TMB exhibit variations in different ethnic groups, and immunotherapy strategies might also differ according to TMB.

Methods

Data from Chinese patients with advanced-stage lung adenocarcinoma

All clinical data and the driver gene mutation status of Chinese patients with NSCLC were retrospectively collected from Shanghai Pulmonary Hospital. The institutional review board approved this study of Shanghai Pulmonary Hospital. *EGFR* mutation status was determined by amplification refractory mutation system (ARMS, AmoyDx).

Data from patients in the TCGA database

All data analyzed were downloaded using TCGAbiolinks (8), which can access the National Cancer Institute (NCI) Genomic Data Commons (GDC) through its GDC Application Programming Interface (API), in order to search, download and prepare relevant data for analysis in R. Briefly, the somatic mutation data of lung adenocarcinoma from the Cancer Genome Atlas Lung Adenocarcinoma (TCGA-LUAD) project (10) were searched and downloaded with GDCquery and GDCprepare, and the clinical data of patients were downloaded with GDCquery_clinic. Then, the data were merged and filtered using several packages in R, such as dplry and stringr (11).

The Kaplan-Meier curves were generated, and Cox regression analyses were performed using survminer (12) and survival package that provides functions for facilitating survival analyses and visualization in R.

Next-generation sequencing

We generated targeted capture pulldown and exon-wide libraries from native DNA using the xGen® Exome Research Panel (Integrated DNA Technologies, Inc., Illinois, USA) and the TruePrep DNA Library Prep Kit V2 for Illumina (#TD501, Vazyme, Nanjing, China). We also generated paired-end sequence data using an Illumina

HiSeq machine. The sequence data were aligned to the human reference genome (NCBI build 37) using BWA (13) and were sorted; duplicate PCR data were removed using SAMtools (14). Somatic mutation calling was performed using VarDict (15).

Annotation and analysis of variants

The vcf files containing the somatic mutations generated by VarDict were annotated with ANNOVAR (16) and then converted to MAF files using maftools (17).

Mutation counts and pathway analysis

In the MAF files of East Asian patients and the TCGA variant data, all synonymous mutations were removed. The variant numbers of each patient were counted using maftools. TMB (mutations/MB) calculation only included non-synonymous mutations. Patients whose TMB was > the median TMB were determined to have a "TMB-H" (TMB-high) status, whereas patients whose TMB was ≤ the median TMB were determined to have a "TMB-L" (TMB-low) status. Genes involved in oncogenic signaling pathways were selected and underwent pathway analysis. The oncogenic pathways were summarized in previous work (18). The OncoPrint was generated with ComplexHeatmap (19).

Statistical analysis

Fisher's exact test was used to compare the frequency data between two or more groups. The Kruskal-Wallis test was performed for the comparison of three or more groups of the variant counts data, and the Wilcoxon test was used for the comparison of two groups of the variant counts data. For multiple test correction, the FDR was measured for the overall accuracy.

Results

Chinese lung adenocarcinoma patient data and the TCGA database

The Chinese lung adenocarcinoma data was derived from a total of 647 cases from Shanghai Pulmonary Hospital. *EGFR* mutations were detected in 57% of the patients by reverse transcription-polymerase chain reaction (RT-PCR), and early- and late-stage patients accounted for 64.5% and 35.5% of all cases with mutations, respectively (*Table S1*). In the WES data of 454 lung adenocarcinoma patients from the TCGA, Asians, Caucasians, and African Americans, each accounted for 1.8%, 86.6%, and 11.7%, respectively. Among the Caucasian patients, early-stage patients accounted for 78.4%.

WES data analysis in different ethnicities: distribution of gene mutations in each signaling pathway

Of 454 patients whose WES data were downloaded from the TCGA database, only 79 Caucasian patients were diagnosed with advanced lung adenocarcinoma. Among 61 Chinese NSCLC patients with WES data, 12 lung squamous cell carcinoma (LUSC) patients were removed, leaving 49 lung adenocarcinoma. However, because Chinese patients with lung adenocarcinoma all had an advanced-stage disease, this study compared the WES data of these two groups of patients (*Figure S1*). Genes were classified into 11 signaling pathways (RTK-RAS, NRF2, TGF-β, PI3K, WNT, MYC, Hippo, p53, cell cycle, Notch and DDR), and the incidence of mutations in the genes of each signaling pathway was analyzed (*Figure S2*). Overall, the incidence of mutations was significantly different among patients of different ethnicities, especially in the *NOTCH2* and *TGFβR2* (*Figure S2*). The incidence of mutations in the TGF-β and RTK-RAS pathways in Chinese patients was significantly higher than that in Caucasian patients, with P values of 0.012 and 0.016, respectively (*Figure 1A*).

No significant differences were observed in the median variant numbers in the 11 signaling pathways in the Chinese patients (P=0.258, *Figure 1B*). The median variant numbers did not show significant differences among the mutant genes, whose incidence was greater than 20% (P=0.568, *Figure 1C*). However, in Caucasian patients, the median variant numbers of those with mutations in the MYC, DDR, and Notch signaling pathways were significantly higher than that of patients with mutations in the RTK-RAS pathway (P<0.001, *Figure 1D*). The median variant numbers segregated according to mutations in individual genes with an incidence of more than 20% also exhibited significant differences (P=0.041, *Figure 1E*), which is likely due to the lower median variant numbers of patients with *TP53* mutations. No significant correlation was observed between the mutation status of these 11 signaling pathways and prognosis (*Figure S3*).

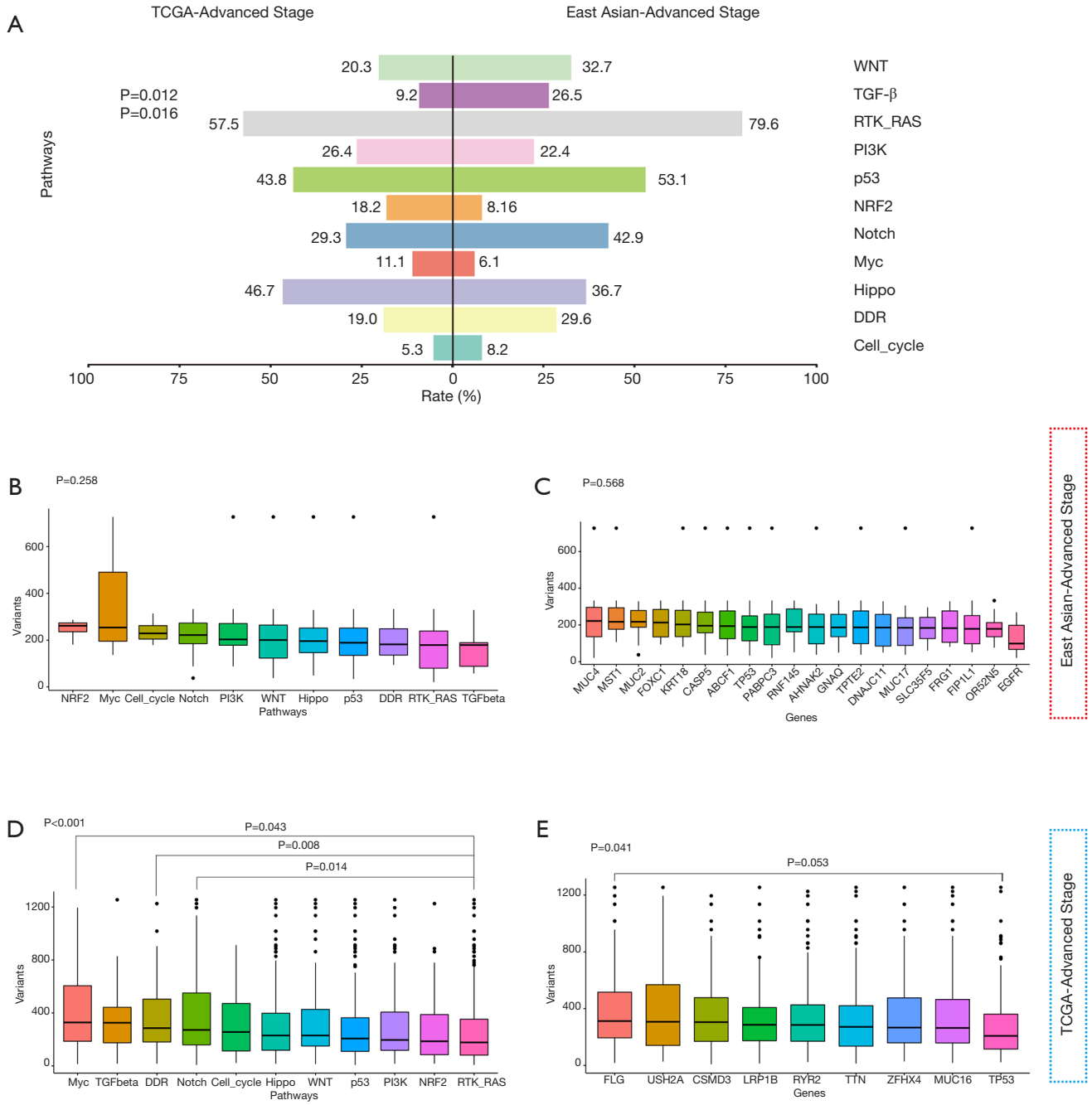

Figure 1 The mutation frequency of 11 signal pathways and the differences in the median variant numbers among the 11 signaling pathways between TCGA and East Asian. (A) The mutation frequency of TGFβ and RTK-RAS pathway in East Asian patients was significantly higher than TCGA (all stages). (B) Variants numbers of signal pathways in East Asian. (C) Variants numbers of genes with a frequency of more than 20% in East Asian. (D) Variants numbers of signal pathways in data from TCGA. (E) Variants numbers of genes with a frequency of more than 20% in data from TCGA. TCGA, The Cancer Genome Atlas.

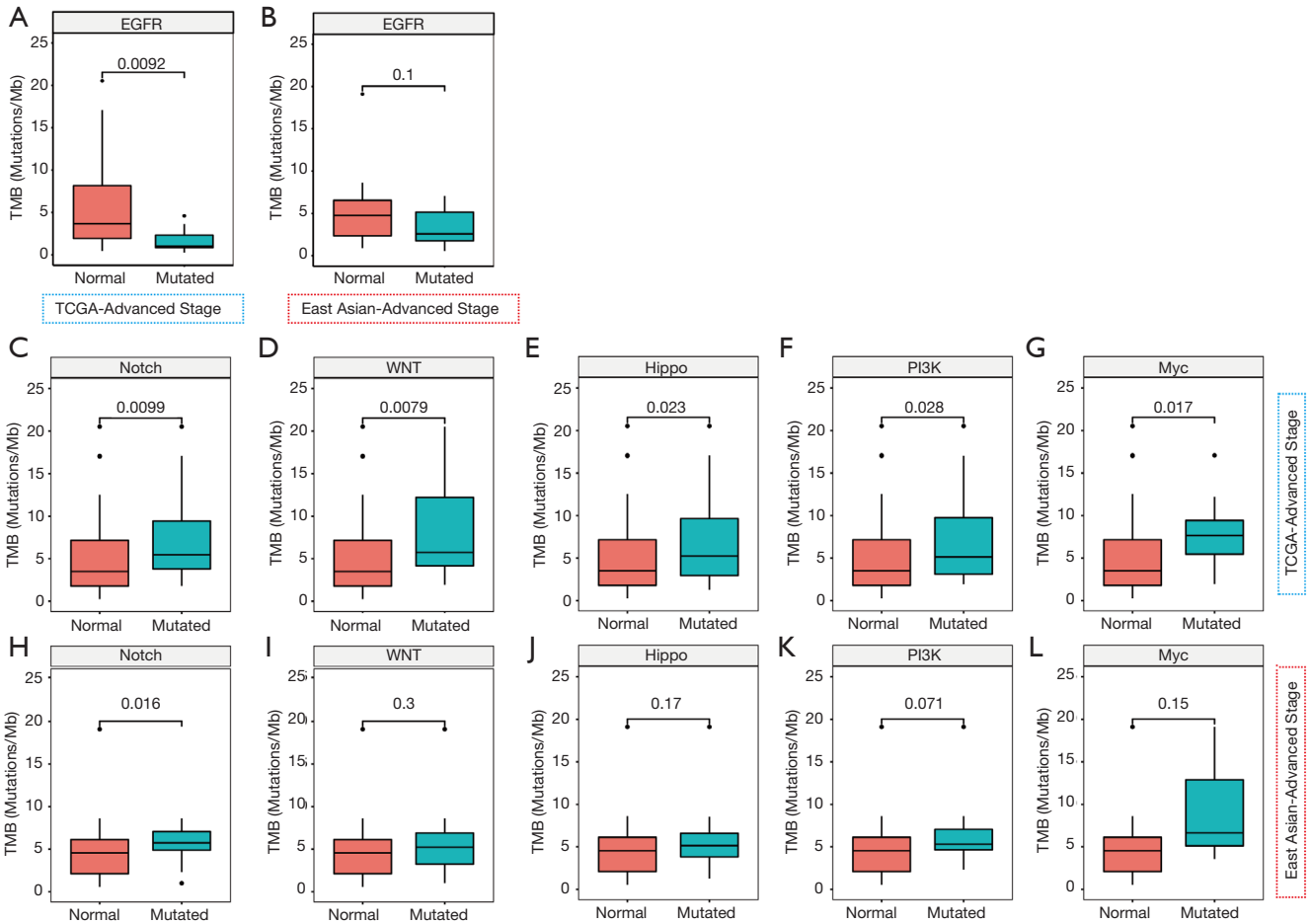

Figure 2 The significant relationship between mutated signal pathways and TMB between TCGA and East Asian. (A) The median TMB was lower in mutated EGFR patients from TCGA, with a statistical significance. (B) The median TMB was lower in mutated EGFR pts from East Asian, without a statistical significance. (C,D,E,F,G) The median TMB was higher in mutated Notch, WNT, Hippo, PI3K, and Myc signaling pathway patients from TCGA, with a statistical significance. (H,I,J,K,L) The median TMB was higher in East Asian patients with mutated WNT, Hippo, PI3K, and Myc signaling pathway, without a statistical significance, except the Notch pathway. EGFR, epidermal growth factor receptor; TMB, tumor mutational burden; TCGA, The Cancer Genome Atlas.

Eighteen genes from the TCGA database and 6 genes from the Chinese population in 11 signaling pathways were significantly correlated with TMB

This study analyzed the association of TMB with all genes in the 11 signaling pathways in patients of different ethnicities (*Figure S4*). The results revealed a significant correlation between TMB and the mutation status of 18 genes in the patients from the TCGA database. *EGFR* mutations were significantly associated with TMB-L, while mutations in the remaining 17 genes were significantly correlated with TMB-H in TCGA (*Figure S4A*, *Figure 2A*,*B*). Mutations of 6 genes showed a significant correlation with TMB-H in the Chinese population (*Figure S4B*). Genes outside of the 11 signaling pathways that are significantly associated with TMB are summarized in *Table S2*.

Impact of TMB status on the prognosis of patients with different disease stages

The median TMB of Chinese patients was 4.61 mutations/MB, while the median TMB of the Caucasian patients was 3.26 mutations/MB (*Table S1*). The difference between the two groups was not statistically significant. No significant

differences were seen in overall survival (OS) between the patients of TMB-H and TMB-L status patients (*Figure S5A,B,C,D*). The Caucasian patients at different stages did not show significant differences in TMB compared with the Chinese patients (P=0.86, *Figure S5E*). In the patients from the TCGA database, the mutation status of the Notch, WNT, Hippo, PI3K, and Myc signaling pathways were significantly associated with TMB-H, whereas in the Chinese patients, only the mutation status of the Notch pathway was significantly correlated with TMB-H (*Figure 2C,D,E,F,G,H,I,J,K,L*).

EGFR mutations have different effects on the prognosis of patients of different ethnic groups with early-stage lung adenocarcinoma

Because of the insufficient number of Asian patients in the TCGA database we combined the data of Chinese NSCLC patients with the data of the Asian population in the TCGA database and then compared the combined data with the data of other ethnic groups. We found that a significantly better prognosis existed in early-stage lung adenocarcinoma of Asian patients with *EGFR*-mutant when compared those with wild-type *EGFR* (P=0.022) (*Figure 3A*). but not in advanced-stage lung adenocarcinoma (P=0.17) (*Figure 3B*). In contrast, out of the early-stage Caucasian patients, the prognosis of patients with wild-type *EGFR* was significantly better than those with *EGFR* mutations (P=0.037) (*Figure 3A*). In Asian patients with early-stage lung adenocarcinoma, the prognosis of light smokers was significantly better than that of heavy smokers (P=0.00021) (*Figure 3C*). (Heavy smoker had a smoking load ≥100, which is calculated by the number of cigarettes smoked per day × years as a smoker, while light smoker had a smoking index <100). However, this phenomenon was not found in other ethnic groups and advanced-stage lung adenocarcinoma (*Figure 3C,D*).

Discussion

In this study, a comprehensive comparative analysis was performed on a total of 11 signaling pathways, including 10 crucial tumor signaling pathways in the TCGA database and the DDR signaling pathway (18). The mutation rates of the TGF-β pathway and the RTK-RAS pathway were significantly different in the East Asian patients. A TGF-β signaling pathway takes part in many cellular processes. The primary genes that are expressed in these pathways are TGFβR1, TGFβR2, SMAD2, SMAD3, and SMAD4 (20,21). These genes have higher alteration rates in pancreatic cancer, kidney cancer, and brain cancer, but have relatively low alteration rates in Caucasian patients with lung adenocarcinoma (22). In terms of the overall alteration rate of the signaling pathway components, the alteration rate of TGF-β in East Asian patients with lung adenocarcinoma is approximately 26.5%, which indicates that TGF-β is one of the pathways with higher alteration rates. The alteration rate of the RTK-RAS pathway is 57.5% in Caucasian patients with lung adenocarcinoma and 79.6% in East Asian patients. The main reason for this difference is the significantly higher frequency of *EGFR* mutations in East Asian patients compared with Caucasian patients. The variations in the two pathways mentioned above are of great significance for the development of anticancer drugs and the exploration of related molecular markers.

TMB may be the most critical molecular marker for anti-PD-1 therapy (7,23). We conducted a comprehensive analysis of the correlation of TMB with 11 signaling pathways and their associated genes in patients of different ethnic groups. We found no differences in TMB in patients with different cancer stages or different ethnic groups. Differences in TMB were present in patients with variations in different pathways. In Caucasians, if mutations occurred in Notch, WNT, Hippo, PI3K, Myc signaling pathways, the TMB of these patients was significantly better than that of patients whose tumors did not have the mutations. However, in the East Asian population, only variations in Notch were related to TMB. The Notch pathway is closely related to immunotherapy response. It has been reported that activation of the Notch pathway inhibited the activation of T cells in tumors and inhibition; the pathway can increase the killing effect of T cells (24). Patients with *NOTCH1* gene mutations exhibit better treatment responses after anti-PD-1 therapy (25). Some reports have also shown that the WNT pathway modulated PD-L1 expression in triple-negative breast cancer and melanoma (26,27). Therefore, these studies supply valuable information for combination therapies, including anti-PD-1 treatment.

Controversial results have been reported on whether TMB is a prognostic factor. Devarakonda *et al.* (28) reported that TMB was an indicator of a good prognosis in all NSCLC cases, despite that no significant correlation was observed between TMB and disease-free survival (DFS) (P=0.171) or OS (P=0.091) according to the data from the 908 cases of NSCLC in the Lung Adjuvant Cisplatin

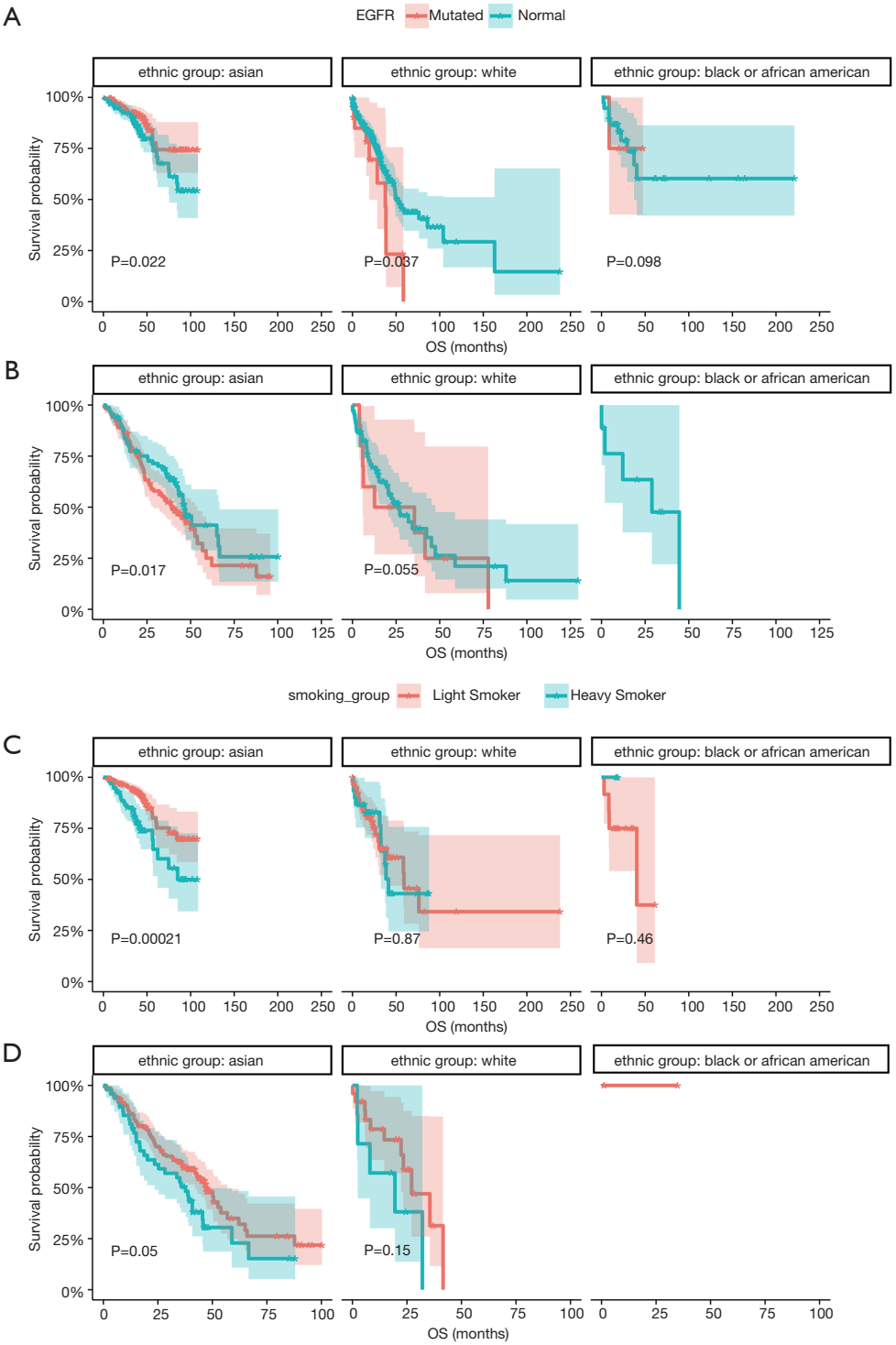

Figure 3 The impact of EGFR mutations on OS between different ethnic groups with lung adenocarcinoma. (A) The impact of EGFR mutations on OS between different ethnic groups with early-stage lung adenocarcinoma; (B) the impact of EGFR mutations on OS between different ethnic groups with advanced-stage lung adenocarcinoma; (C) the impact of smoking on OS between different ethnic groups with early-stage lung adenocarcinoma; (D) the impact of smoking on OS between different ethnic groups with advanced-stage lung adenocarcinoma. EGFR, epidermal growth factor receptor; OS, overall survival.

Evaluation (LACE)-Bio-II study. However, studies from Japan showed poorer postoperative prognosis in patients with higher TMB (HR =12.3, P=0.019) (29). Our results are like the LACE study. TMB demonstrates no correlation with the prognosis of patients with either early- or late-stage disease. The difference described above is caused by different TMB test methods and cut-off values. The gold standard of TMB is WES. However, the WES probes used in numerous studies cover different regions, which may cause inconsistency in the standards of the TMB calculation. Furthermore, the cut-off values for TMB have not been determined. In the abovementioned LACE study, 8 mutations/M was used, while the Japanese study used 62 mutations/patient. Therefore, the establishment of standards for TMB testing is urgently needed.

Whether EGFR mutations are prognostic factors for patients of different ethnic groups is currently inconclusive. Kosaka *et al.* found that *EGFR* mutations did not show a correlation with postoperative OS (P=0.9933), and several other studies reported the same findings (30-33). However, in a study of 1,118 Caucasian patients with lung adenocarcinoma (34), patients with *EGFR* mutations had a better prognosis, HR =0.51 (P<0.001). In our study, among patients with early-stage lung adenocarcinoma, Caucasian patients with *EGFR*-mutant had a poor prognosis, while East Asian patients with EGFR-mutant had a better prognosis. The reason may be that East Asian patients have a higher *EGFR* mutation frequency, and a higher proportion of patients received EGFR-TKIs therapy.

For the first time, we systematically analyzed the genetic variations associated with TMB, but this study still has several limitations. The differences in the TMB assessment criteria, WES test methods, and differences between the distinct stages of TMB may all affect the evaluation of the results. Thus, the results of this study should be confirmed in a more prospective study.

In conclusion, first, the comparison of tumor signaling pathways and gene mutation profiles between Caucasians and East Asians showed racial differences in terms of variations in the TFGβ and RTK-RAS pathways. Second, the mutations in 5 signaling pathways and 18 genes were all significantly associated with the differences in TMB in the Caucasian population, only variations in the Notch pathway and 6 genes were correlated with an increase in TMB in the East Asian population. Finally, we found that TMB was not correlated with patient prognosis, and *EGFR* mutations showed a better prognosis in East Asian patients, whereas the inverse was observed in Caucasian patients. These findings reveal differences in genetic variations between Caucasian and East Asian patients with lung adenocarcinomas and ease our research on therapeutic strategies, especially immunotherapeutic strategies, and the identification of potential biomarkers.

Acknowledgments

This work was sponsored by Shanghai Tongshu Biotechnology Co., Ltd., and supported by Pfizer Investment.

Funding: This work was supported in part by the National Natural Science Foundation of China (No. 81572270, No. 81602000).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tlcr-20-457). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The institutional review board approved this study of Shanghai Pulmonary Hospital and individual consent for this retrospective analysis was waived.

References

1. Ettinger DS, Wood DE, Akerley W, et al. NCCN Guidelines Insights: Non-Small Cell Lung Cancer, Version 4.2016. J Natl Compr Canc Netw 2016;14:255-64.
2. Arenberg D. Update on screening for lung cancer. Transl

Lung Cancer Res 2019;8:S77-87.
3. Nahar R, Zhai W, Zhang T, et al. Elucidating the genomic architecture of Asian EGFR-mutant lung adenocarcinoma through multi-region exome sequencing. Nat Commun 2018;9:216.
4. Chen J, Soudy H. Osimertinib in the treatment of leptomeningeal disease in T790M-negative, epidermal growth factor receptor-mutated non-small cell lung cancer: a case report. Chin Clin Oncol 2019;8:29.
5. Mountzios G. Making progress in epidermal growth factor receptor (EGFR)-mutant non-small cell lung cancer by surpassing resistance: third-generation EGFR tyrosine kinase inhibitors (EGFR-TKIs). Ann Transl Med 2018;6:140.
6. Merino DM, McShane LM, Fabrizio D, et al. Establishing guidelines to harmonize tumor mutational burden (TMB): in silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of Cancer Research TMB Harmonization Project. J Immunother Cancer 2020;8:e000147.
7. Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 2015;348:124-8.
8. Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 2016;44:e71.
9. Motono N, Funasaki A, Sekimura A, et al. Prognostic value of epidermal growth factor receptor mutations and histologic subtypes with lung adenocarcinoma. Med Oncol 2018;35:22.
10. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014;511:543-50.
11. Wickham H. Stringr: simple, Consistent Wrappers for Common String Operations. 2018.
12. Kassambara A, Kosinski M. Survminer: drawing Survival Curves using 'ggplot2'. 2018.
13. Slater PM, Grivell R, Cyna AM. Labour management of a woman with carnitine palmitoyl transferase type 2 deficiency. Anaesth Intensive Care 2009;37:305-8.
14. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011;27:2987-93.
15. Lai Z, Markovets A, Ahdesmaki M, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res 2016;44:e108.
16. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
17. Mayakonda A, Koeffler HP. Maftools: Efficient Analysis, Visualization and Summarization of MAF Files from Large-Scale Cohort Based Cancer Studies. bioRxiv 2016. doi: https://doi.org/10.1101/052662.
18. Sanchez-Vega F, Mina M, Armenia J, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell 2018;173:321-337.e10.
19. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847-9.
20. Wrana JL, Attisano L, Carcamo J, et al. TGF beta signals through a heteromeric protein kinase receptor complex. Cell 1992;71:1003-14.
21. Runyan CE, Schnaper HW, Poncelet AC. The role of internalization in transforming growth factor beta1-induced Smad2 association with Smad anchor for receptor activation (SARA) and Smad2-dependent signaling in human mesangial cells. J Biol Chem 2005;280:8300-8.
22. Schaub FX, Dhankani V, Berger AC, et al. Pan-cancer Alterations of the MYC Oncogene and Its Proximal Network across the Cancer Genome Atlas. Cell Syst 2018;6:282-300.e2.
23. Carbone DP, Reck M, Paz-Ares L, et al. First-Line Nivolumab in Stage IV or Recurrent Non-Small-Cell Lung Cancer. N Engl J Med 2017;376:2415-26.
24. Sierra RA, Thevenot P, Raber PL, et al. Rescue of notch-1 signaling in antigen-specific CD8+ T cells overcomes tumor-induced T-cell suppression and enhances immunotherapy in cancer. Cancer Immunol Res 2014;2:800-11.
25. Champiat S, Dercle L, Ammari S, et al. Hyperprogressive Disease Is a New Pattern of Progression in Cancer Patients Treated by Anti-PD-1/PD-L1. Clin Cancer Res 2017;23:1920-8.
26. Castagnoli L, Cancila V, Cordoba-Romero SL, et al. WNT signaling modulates PD-L1 expression in the stem cell compartment of triple-negative breast cancer. Oncogene 2019;38:4047-60.
27. Kakavand H, Rawson RV, Pupo GM, et al. PD-L1 Expression and Immune Escape in Melanoma Resistance to MAPK Inhibitors. Clin Cancer Res 2017;23:6054-61.
28. Devarakonda S, Rotolo F, Tsao MS, et al. Tumor Mutation Burden as a Biomarker in Resected Non-Small-Cell Lung Cancer. J Clin Oncol 2018;36:2995-3006.
29. Owada-Ozaki Y, Muto S, Takagi H, et al. Prognostic

Impact of Tumor Mutation Burden in Patients With Completely Resected Non-Small Cell Lung Cancer: Brief Report. J Thorac Oncol 2018;13:1217-21.
30. Kosaka T, Yatabe Y, Endoh H, et al. Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. Cancer Res 2004;64:8919-23.
31. Sasaki H, Shimizu S, Endo K, et al. EGFR and erbB2 mutation status in Japanese lung cancer patients. Int J Cancer 2006;118:180-4.
32. Kobayashi N, Toyooka S, Ichimura K, et al. Non-BAC component but not epidermal growth factor receptor gene mutation is associated with poor outcomes in small adenocarcinoma of the lung. J Thorac Oncol 2008;3:704-10.
33. Lim KH, Huang MJ, Liu HC, et al. Lack of prognostic value of EGFR mutations in primary resected non-small cell lung cancer. Med Oncol 2007;24:388-93.
34. D'Angelo SP, Janjigian YY, Ahye N, et al. Distinct clinical course of EGFR-mutant resected lung cancers: results of testing of 1118 surgical specimens and effects of adjuvant gefitinib and erlotinib. J Thorac Oncol 2012;7:1815-22.

Cite this article as: Gu W, Wang N, Gu W, Qiu Y, Zhang H, Liang J, Zhou T, Ma L, Cai W, Feng W, Chen L. Molecular gene mutation profiles, TMB and the impact of prognosis in Caucasians and east Asian patients with lung adenocarcinoma. Transl Lung Cancer Res 2020;9(3):629-638. doi: 10.21037/tlcr-20-457

Supplementary

Table S1 The clinical features of patients with lung adenocarcinoma from East Asian and The Cancer Genome Atlas (TCGA) database

Features	Data from East Asian (n=696)				Data from TCGA (n=454)		
	RT-PCR		WES		WES		
	n=647	P*	n=49	P#	Asian (n=8)	White (n=393)	Black (n=53)
Gender							
Male	288 (44.5%)	0.951	35 (71.4%)	<0.001	4 (50.0%)	176 (44.8%)	23 (43.4%)
Female	359 (55.5%)		14 (28.6%)		4 (50.0%)	217 (55.2%)	30 (56.6%)
Age (yr, median, range)	59 [27–82]	<0.001	60 [30–81]	<0.001	65 [49–81]	68 [33–89]	61 [40–68]
Smoking							
Never/light smoker	486 (75.1%)	1			NA	105 (26.7%)##	14 (26.4%)###
Smoker	161 (24.9%)				NA	37 (9.4%)##	2 (3.8%)###
Tumor stage							
Stage I/II	421 (65.1%)	<0.001	0 (0%)	<0.001	6 (75%)	308 (78.4%)	42 (79.2%)
Stage III/IV	226 (34.9%)		49 (100%)		2 (25%)	78 (19.8%)	9 (17.0%)
EGFR							
Mutation	369 (57.0%)	<0.001	15 (30.6%)	<0.001	1 (12.5%)	31 (7.9%)	5[a] (9.4%)
Stage I/II	238				1	21	4
Stage III/IV	131				0	10	0
Wildtype	278 (43.0%)		34 (70.4%)		7 (87.5%)	362[b] (92.1%)	48[a] (90.6%)
Stage I/II	183				5	287	38
Stage III/IV	95				2	69	9
Tumor mutation burden (mutations/MB) (median, range)			4.61 (0.55–19.13)		2.70 (0.5–10)	3.26 (0.08–33.03)	6.05 (0.18–32.92)
Low			0.66	0.881	0.11	5.05	0.68
High			0.63		0.11	5.29	0.71

*, the RT-PCR data of east Asian patients were compared with the white race from TCGA. #, the WES data of east Asian patients were compared with the white race from TCGA. ##, 251 missing. ###, 37 missing. [a], 1 missing. [b], 6 missing. RT-PCR, reverse transcription-polymerase chain reaction; WES, whole-exome sequencing; NA, not available.

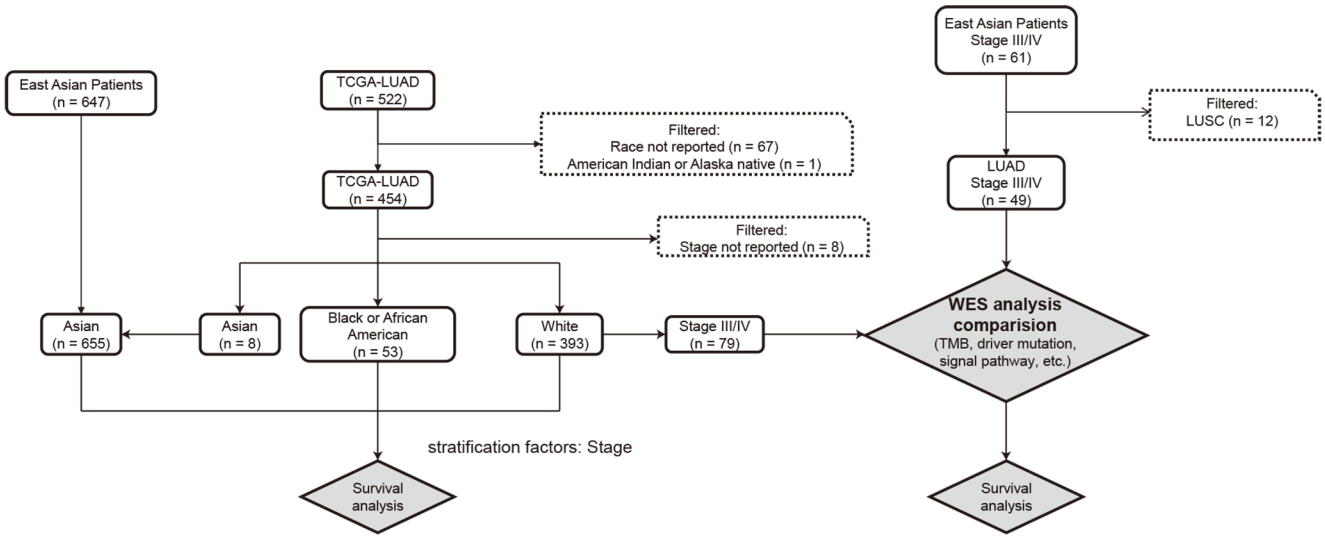

Figure S1 The workflow of this study. TCGA-LUAD, The Cancer Genome Atlas Lung Adenocarcinoma; LUSC, lung squamous cell carcinoma; WES, whole-exome sequencing; TMB, tumor mutational burden.

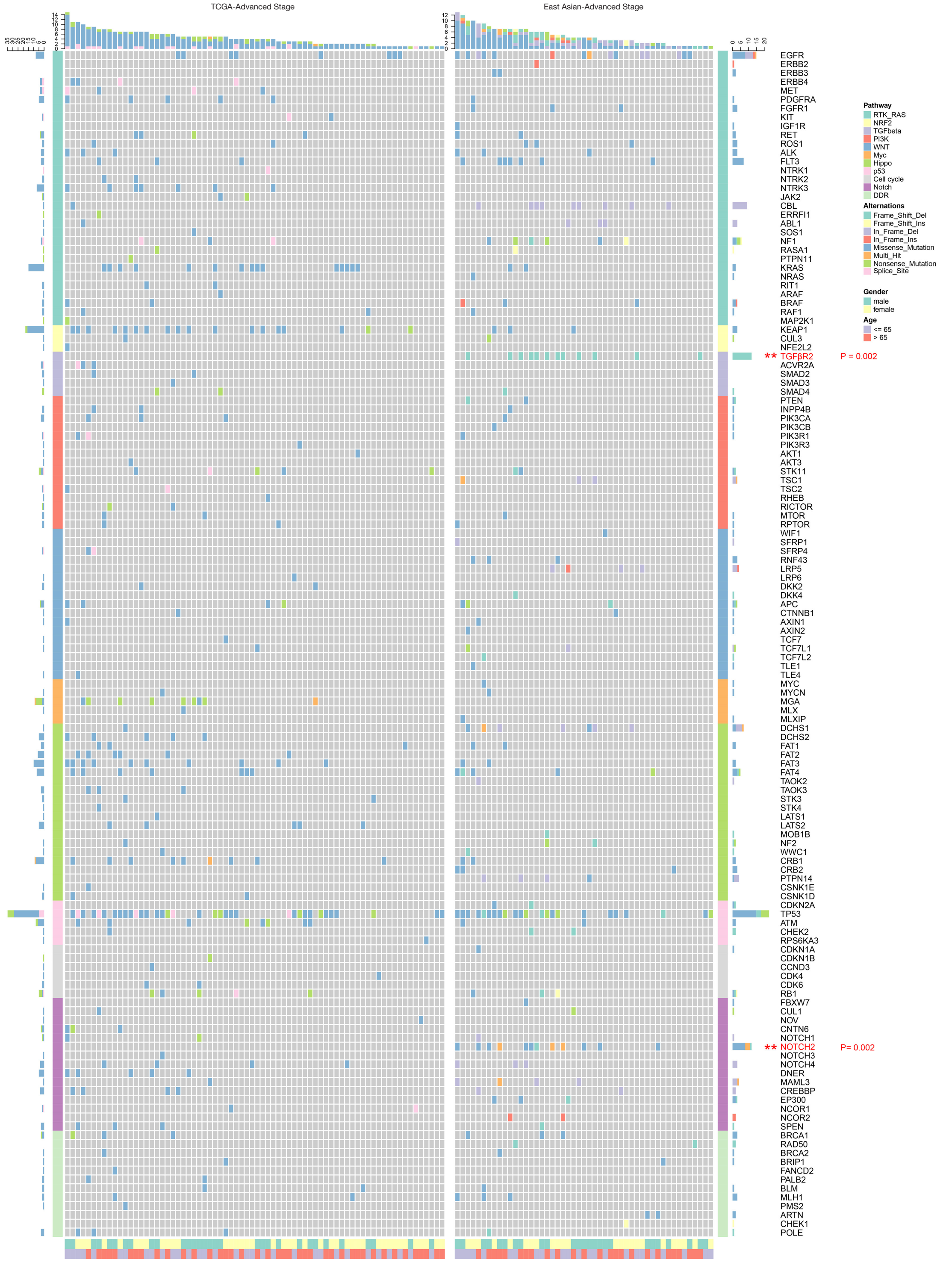

Figure S2 The frequency of genes among 11 signal pathways between East Asian and White with advanced lung adenocarcinoma. TMB, tumor mutational burden; TCGA, The Cancer Genome Atlas.

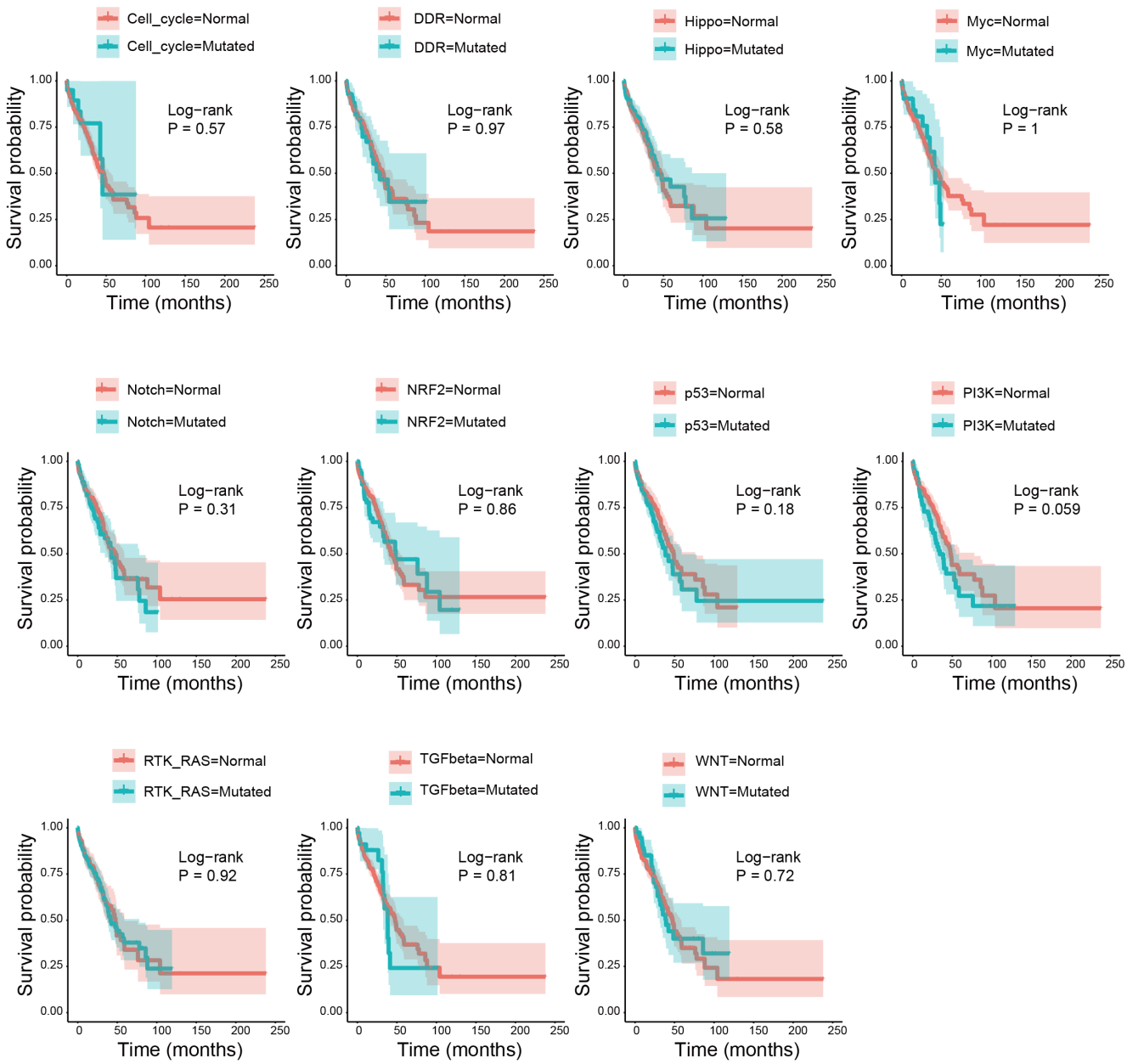

Figure S3 The OS analysis between mutation and wild-type group in 11 signal pathways. OS, overall survival.

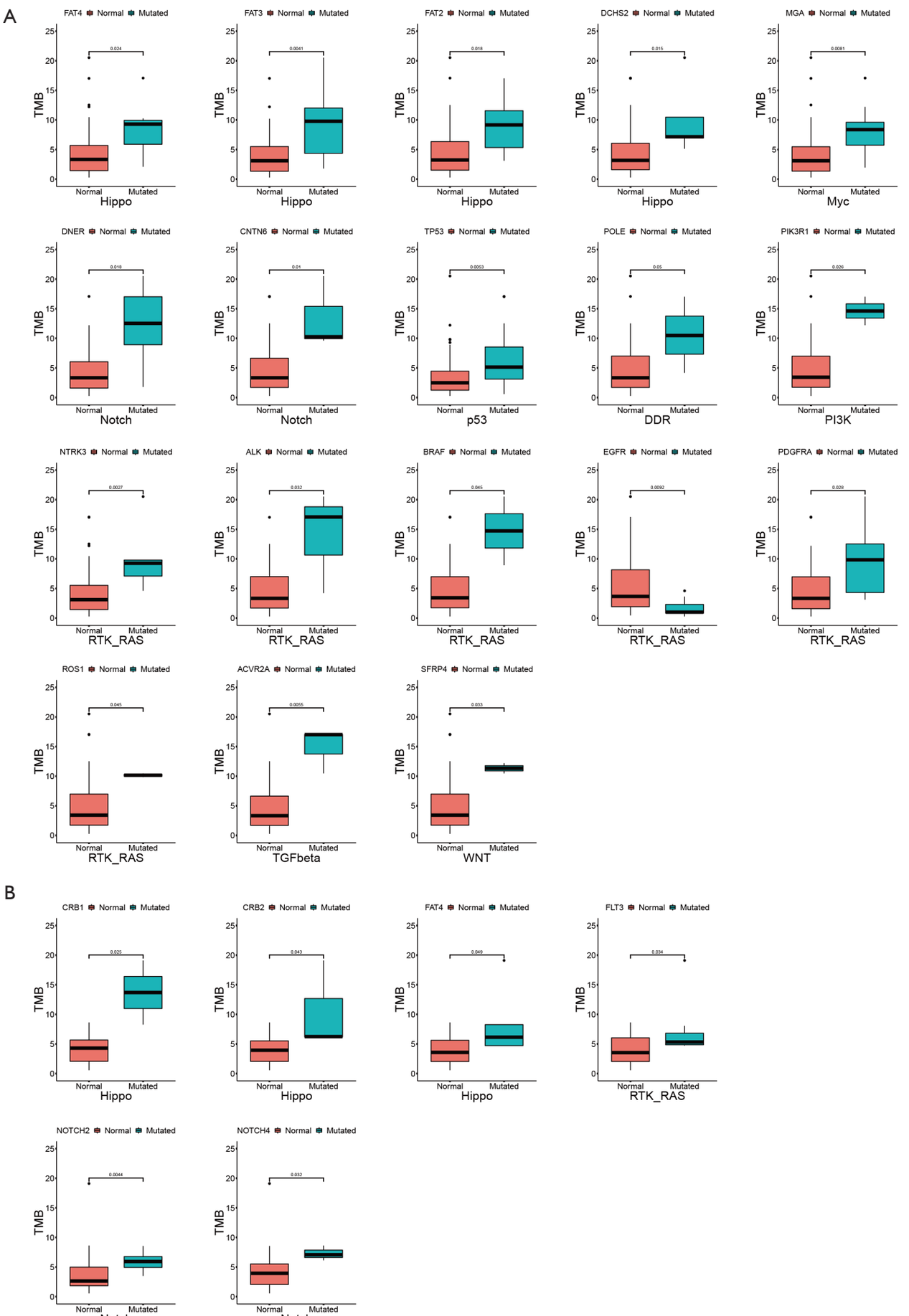

Figure S4 (A) Eighteen genes from TCGA database and (B) 6 genes from the Chinese population in 11 signaling pathways were significantly correlated with TMB. TCGA, The Cancer Genome Atlas; TMB, tumor mutational burden.

Table S2 Genes, not included in 11 signal pathways, with different variant frequencies between The Cancer Genome Atlas (TCGA) and East Asian patients with adjusted P value <0.05

Hugo_Symbol	Altered samples TCGA late	Normal samples TCGA late	Altered samples East Asian	Normal samples East Asian	Fisher's P value	Adjusted P value
FIP1L1	1	76	23	26	1.146E-10	1.065E-06
KRT18	0	77	19	30	1.186E-09	5.509E-06
GNAQ	0	77	18	31	4.131E-09	9.597E-06
MST1	0	77	18	31	4.131E-09	9.597E-06
MUC4	2	75	21	28	1.275E-08	2.179E-05
RNF145	0	77	17	32	1.407E-08	2.179E-05
PABPC3	1	76	19	30	1.740E-08	2.310E-05
ABCF1	2	75	20	29	4.187E-08	4.843E-05
DNAJC11	0	77	16	33	4.690E-08	4.843E-05
FOXC1	0	77	15	34	1.531E-07	1.423E-04
OR52N5	1	76	17	32	1.830E-07	1.546E-04
CASP5	1	76	16	33	5.722E-07	3.128E-04
FRG1	1	76	16	33	5.722E-07	3.128E-04
MUC2	1	76	16	33	5.722E-07	3.128E-04
PABPC1	0	77	14	35	4.900E-07	3.128E-04
SLC35F5	1	76	16	33	5.722E-07	3.128E-04
TPTE2	1	76	16	33	5.722E-07	3.128E-04
MARCKS	0	77	13	36	1.538E-06	7.940E-04
DKC1	0	77	12	37	4.739E-06	1.915E-03
FUZ	0	77	12	37	4.739E-06	1.915E-03
TREML2	0	77	12	37	4.739E-06	1.915E-03
CEL	0	77	11	38	1.434E-05	4.859E-03
CRLF1	0	77	11	38	1.434E-05	4.859E-03
DSPP	1	76	13	36	1.516E-05	4.859E-03
FOXO3	0	77	11	38	1.434E-05	4.859E-03
MYL1	1	76	13	36	1.516E-05	4.859E-03
SUZ12	0	77	11	38	1.434E-05	4.859E-03
ADAMTS7	1	76	12	37	4.315E-05	9.780E-03
ANKRD36	1	76	12	37	4.315E-05	9.780E-03
ASTE1	0	77	10	39	4.266E-05	9.780E-03
CD3EAP	0	77	10	39	4.266E-05	9.780E-03
DDX11	1	76	12	37	4.315E-05	9.780E-03
KRT8	0	77	10	39	4.266E-05	9.780E-03
NT5C3A	0	77	10	39	4.266E-05	9.780E-03
OR7E24	0	77	10	39	4.266E-05	9.780E-03
PRSS3	0	77	10	39	4.266E-05	9.780E-03
SLC22A9	0	77	10	39	4.266E-05	9.780E-03
TNFAIP6	1	76	12	37	4.315E-05	9.780E-03
ZNF705B	0	77	10	39	4.266E-05	9.780E-03
BRD7	0	77	9	40	1.248E-04	1.999E-02
EVPLL	0	77	9	40	1.248E-04	1.999E-02
FKBP9	1	76	11	38	1.200E-04	1.999E-02
GOLGA6L6	0	77	9	40	1.248E-04	1.999E-02
HRCT1	0	77	9	40	1.248E-04	1.999E-02
IGFBP2	1	76	11	38	1.200E-04	1.999E-02
NAP1L2	0	77	9	40	1.248E-04	1.999E-02
OR13C9	0	77	9	40	1.248E-04	1.999E-02
OR2T29	0	77	9	40	1.248E-04	1.999E-02
PBOV1	0	77	9	40	1.248E-04	1.999E-02
RHPN2	0	77	9	40	1.248E-04	1.999E-02
SLC35G6	0	77	9	40	1.248E-04	1.999E-02
TAF1B	0	77	9	40	1.248E-04	1.999E-02
UNC93A	0	77	9	40	1.248E-04	1.999E-02
UTP3	0	77	9	40	1.248E-04	1.999E-02
ZBTB40	0	77	9	40	1.248E-04	1.999E-02
ZFHX3	1	76	11	38	1.200E-04	1.999E-02
ARID1B	2	75	12	37	2.111E-04	3.269E-02
SVEP1	2	75	12	37	2.111E-04	3.269E-02
DNHD1	3	74	14	35	2.237E-04	3.353E-02
TCF20	3	74	14	35	2.237E-04	3.353E-02
ANKRD20A4	0	77	8	41	3.591E-04	4.171E-02
AR	1	76	10	39	3.258E-04	4.171E-02
ATAD3B	1	76	10	39	3.258E-04	4.171E-02
ATXN1	1	76	10	39	3.258E-04	4.171E-02
CCDC168	0	77	8	41	3.591E-04	4.171E-02
CNOT2	1	76	10	39	3.258E-04	4.171E-02
GOLGA6D	0	77	8	41	3.591E-04	4.171E-02
IL32	0	77	8	41	3.591E-04	4.171E-02
KRTAP4-7	0	77	8	41	3.591E-04	4.171E-02
MED15	0	77	8	41	3.591E-04	4.171E-02
OTOP1	0	77	8	41	3.591E-04	4.171E-02
RNPC3	0	77	8	41	3.591E-04	4.171E-02
SLC8A2	0	77	8	41	3.591E-04	4.171E-02
THOC3	0	77	8	41	3.591E-04	4.171E-02
ZNF185	0	77	8	41	3.591E-04	4.171E-02
ZNF384	0	77	8	41	3.591E-04	4.171E-02
ZNF705D	0	77	8	41	3.591E-04	4.171E-02
ZNF727	0	77	8	41	3.591E-04	4.171E-02
USH2A	19	58	1	48	3.691E-04	4.234E-02
LRP1B	25	52	3	46	3.799E-04	4.305E-02
TTN	35	42	7	42	4.047E-04	4.531E-02
NAV3	15	62	0	49	4.227E-04	4.622E-02
TNR	15	62	0	49	4.227E-04	4.622E-02

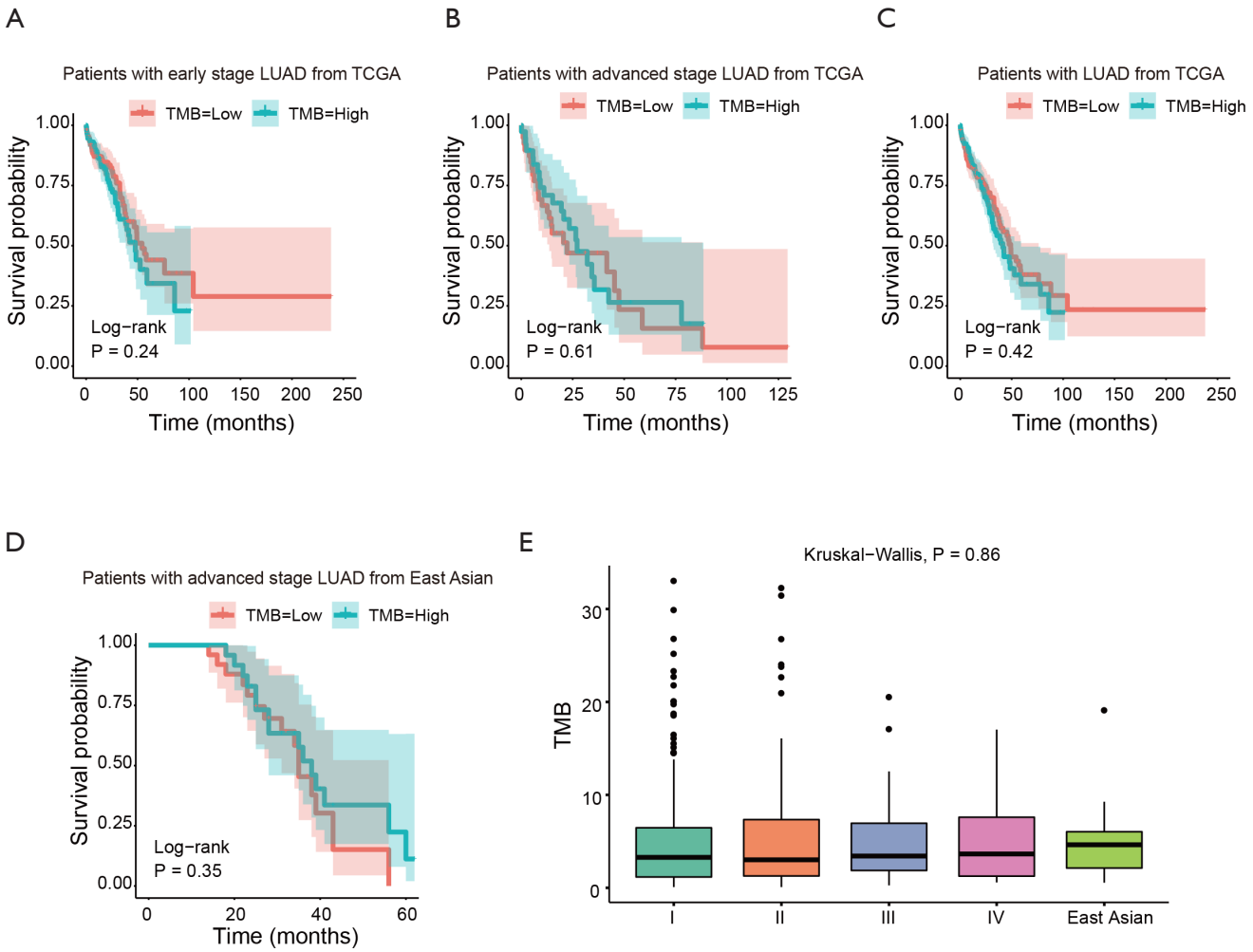

Figure S5 The impact of TMB on OS between The Cancer Genome Atlas (TCGA) and East Asian lung adenocarcinoma. OS between the patients (A) with early stage LUAD from TCGA; (B) with advanced stage LUAD from TCGA; (C) with LUAD from TCGA; (D) with advanced stage LUAD from East Asian of TMB-H and TMB-L status; (E) the TMB of Caucasian patients at different stages compared with the Chinese patients. TMB, tumor mutational burden; OS, overall survival; LUAD, lung adenocarcinoma.