Molecular Therapy
Methods & Clinical Development
Original Article

Evaluating the Potential of T Cell Receptor Repertoires in Predicting the Prognosis of Resectable Non-Small Cell Lung Cancers

Zhengbo Song,[1,8] Xiangbin Chen,[2,8] Yi Shi,[3] Rongfang Huang,[4] Wenxian Wang,[1] Kunshou Zhu,[5] Shaofeng Lin,[5] Minxian Wang,[6] Geng Tian,[2] Jialiang Yang,[2] and Gang Chen[4,7]

[1]Department of Medical Oncology, Cancer Hospital of University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou 310022, China; [2]Geneis Beijing, Beijing 100102, China; [3]Department of Molecular Pathology, Fujian Medical University Cancer Hospital & Fujian Cancer Hospital, Fuzhou 350014, China; [4]Department of Pathology, Fujian Medical University Cancer Hospital & Fujian Cancer Hospital, Fuzhou 350014, China; [5]Department of Thoracic Surgery, Fujian Cancer Hospital & Fujian Medical University Cancer Hospital, Fuzhou 350014, China; [6]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02139, USA; [7]Fujian Provincial Key Laboratory of Translational Cancer Medicine, Fuzhou 350014, China

For resectable cancer patients, a method that could precisely predict the risk of postoperative recurrence would be crucial for guiding adjuvant treatment. Since T cell receptor (TCR) repertoires had been shown to be closely related to the dynamics of cancers, here we enrolled a cohort of patients to evaluate the potential of TCR repertoires in predicting the prognosis of resectable non-small cell lung cancers. Specifically, TCRβ repertoires were analyzed in surgical tumor tissues and matched adjacent non-tumor tissues from 39 patients enrolled with resectable non-small cell lung cancer, through target enrichment and high-throughput sequencing. As a result, there are significant differences between the TCR repertories of tumor samples and those of matched adjacent non-tumor samples as evaluated by criteria like the number of clonotypes. In addition, TCR repertoires were significantly associated with a few clinical features, as well as somatic mutations. Finally, certain TCRβ variable-joining (V-J) pairings were featured to build a logistic regression model in predicting postoperative recurrence of resectable non-small cell lung cancers with a testing area under the receiver operating characteristic curve (AUC) of around 0.9. Thus, we hypothesize that TCR repertoires could be potentially used to predict prognosis after curative surgery for non-small cell lung cancer patients.

INTRODUCTION

Surgical resection is considered to be a preferable option for patients with many kinds of solid cancers. However, a major problem after curative surgery is the postoperative recurrence for most cancer treatments. For example, relevant statistics show that more than half of the resectable lung cancer patients had recurrence following curative surgery.[1] Thus, an effective method that can predict the risk of postoperative recurrence would play crucial roles in guiding adjuvant treatment for cancer patients.

Clinical factors, such as tumor-node-metastasis (TNM) stage, have potential in predicting the prognosis of lung cancer patients after surgical resection.[2] Recently, the TNM stage classification has been widely used to estimate the postoperative outcome and guide adjuvant therapy.[3] A general assumption is that disease progression is a tumor cell-autonomous process.[4] However, this assumption ignores the effects of host immune responses, which have been demonstrated to be useful in predicting tumor progression or clinical outcomes in multiple tumor types.[5–7] As an indication, it might be better to predict postoperative outcome by integrating both clinical factors and immune-related markers.[8]

It is known that the repertoires of T cell receptor (TCR) play critical roles in anti-tumor immune responses for oncology.[9,10] The diversity of TCRs is resulted from the genomic rearrangement of variable (V), diversity (D), and joining (J) regions, as well as palindromic and random incorporation of nucleotides, which is important for the adaptive immunity.[11] Latest methodological advancements, especially high-throughput sequencing technology, have allowed fast yet accurate identification and quantification of the TCR repertoire for any biological sample.[12]

Recently, the repertoires of TCR had been found to associate with the survival of patients of many cancers, such as breast cancer,[13–15] hepatocellular carcinoma,[16,17] pancreatic ductal adenocarcinoma,[18] lung cancer,[19,20] gastric cancer,[21] and cervical cancer.[22] A recent study reported that the dynamic change of TCR repertoire profiling of peripheral blood during anti-cancer treatment was a good indicator of clinical outcome of advanced lung cancer patients.[19] However, the anti-cancer treatment of patients in that study included chemotherapy,

Received 24 February 2020; accepted 19 May 2020;
https://doi.org/10.1016/j.omtm.2020.05.020.

[8]These authors contributed equally to this work.

Correspondence: Gang Chen, Department of Pathology, Fujian Medical University Cancer Hospital & Fujian Cancer Hospital, Fuzhou 350014, China.
E-mail: naichengang@126.com
Correspondence: Jialiang Yang, Geneis Beijing Co., Ltd., Beijing 100102, China.
E-mail: yangjl@geneis.cn

Table 1. The Clinical Pathological Characteristics of Enrolled Patients

Clinical Features	All Samples (n = 39)	Training Set (n = 25)	Testing Set (n = 14)
Sex			
Male	26	15	11
Female	13	10	3
Age (years)			
≤60	18	13	5
>60	21	12	9
Pathological type			
Adenocarcinoma	29	19	10
Squamous cell cancer	10	6	4
TNM Stage			
I	18	10	8
II	11	5	6
III	8	8	0
IV	2	2	0
Preoperative C-REACTIVE PROTEIN			
median (range), mg/L	1.16 (0.08∼79.54)	1.07 (0.08∼79.54)	1.265 (0.27∼14.46)
Preoperative neutrophil			
median (range), %	59.6 (45∼78.1)	59.3 (45∼78.1)	59.75 (46.7∼74.2)
Preoperative white blood cell			
median (range), '10^9	5.7 (4.2∼10.7)	5.7 (4.2∼10.7)	5.95 (4.4∼9.6)
PD-L1			
median (range), %	3.5 (0∼90)	1.5 (0∼90)	15 (0∼75)
CD8+			
median (range), %	5 (1∼70)	5 (1∼70)	5 (1∼50)
Disease status at last follow-up			
Non-recurrence	16	11	5
Recurrence	23	14	9

radiotherapy, tyrosine kinase inhibitor (TKI) therapy, surgery, anti-angiogenic therapy, or combinations of several treatment,[19] and only 2 out of 48 patients received surgical resection. As a consequence, the results and conclusions were mainly based on multiple treatments for advanced lung cancer, instead of resectable lung cancer. Another study by Reuben et al.[20] showed that T cell repertoire intratumor heterogeneity was associated with the risk of postsurgical relapse and disease-free survival in localized lung adenocarcinomas. However, multi-region (2 to 5 regions per tumor) TCR profiling was necessary to investigate the heterogeneity of TCR repertoires,[20] which would undoubtedly bring great difficulties to future clinical application.

In this study, we aimed to evaluate the possibility of predicting the prognosis of post-surgical non-small cell cancer patients, based on TCR repertoires from single region (instead of multiple regions) of the cancer tissue. By combining TCRβ-targeted high throughput sequencing and whole-exome sequencing, we investigated the characteristics of TCR repertoires and their potential relationship with clinical features and somatic mutations. Importantly, our study assessed the potential of TCR repertoires in predicting postoperative recurrence for non-small cell cancer patients.

RESULTS

Clinical Characteristics of Patients

A total of 39 patients with non-small cell lung cancer (NSCLC) who received curative surgery at the Zhejiang Cancer Hospital were enrolled. The age of all patients ranged from 41 to 71 years old, among which 66.7% (26/39) were male and 74.4% (29/39) were adenocarcinoma subtype besides squamous cell subtype (Table 1). According to the Union for International Cancer Control (UICC, 2017) staging system, 46.2% (18/39), 28.2% (11/39), and 25.6% (10/39) of patients were diagnosed with stage I, stage II, and stage III/IV, respectively.

The preoperative biochemical indices including C-reaction protein, neutrophil ratio, white blood cell, $CD8^+$, and PD-L1 level were also tested (Table 1). As can be seen, 59.0% (23/39) of patients had postoperative recurrence at the last follow-up, while 41% (16/39) had no evidence of recurrence.

Comparisons of TCR Repertoires of Tumor Tissues with Adjacent Non-Tumors

To explore the characteristics of TCR repertoires, we conducted high-throughput sequencing of TCRβ repertoires using each tumor and adjacent non-tumor tissues derived from these 39 enrolled patients. As a result, a total of 61 V gene segments and 13 J gene segments were identified in these 39 paired samples, and subsequently 703 distinct V-J combinations were obtained. Principal-component analysis (PCA) was then applied to all samples based on the 703 V-J combinations. There is a well separation between tumor and normal samples, indicating that the TCR patterns in the two types of samples are quite different and the samples have low chance of contamination (Figure S1).

Several indices were calculated to describe the repertoires of tumor sample and paired adjacent non-tumor tissues of each patient. These indices include the number of clonotypes, which represents unique CDR3 amino acid (aa) sequences translated by coding sequences that contain V gene, J gene, and C (constant) gene,[23] and diversity index (the Shannon-Wiener index was calculated in this study). Because of the skewed distribution of T cell clones, the TCR repertoires could be dominated by a small fraction of highly expanded T cell clones. Therefore, highly expanded clones (HECs) ratio representing the proportion of highly expanded clones whose frequency was more than 0.1%, and the cumulative percentage of top 100 frequent TCRβ in each sample, were also included as the repertoires' indices.

These repertoires' indices mentioned above were compared between tumor and adjacent non-tumor tissues using the paired t test. The

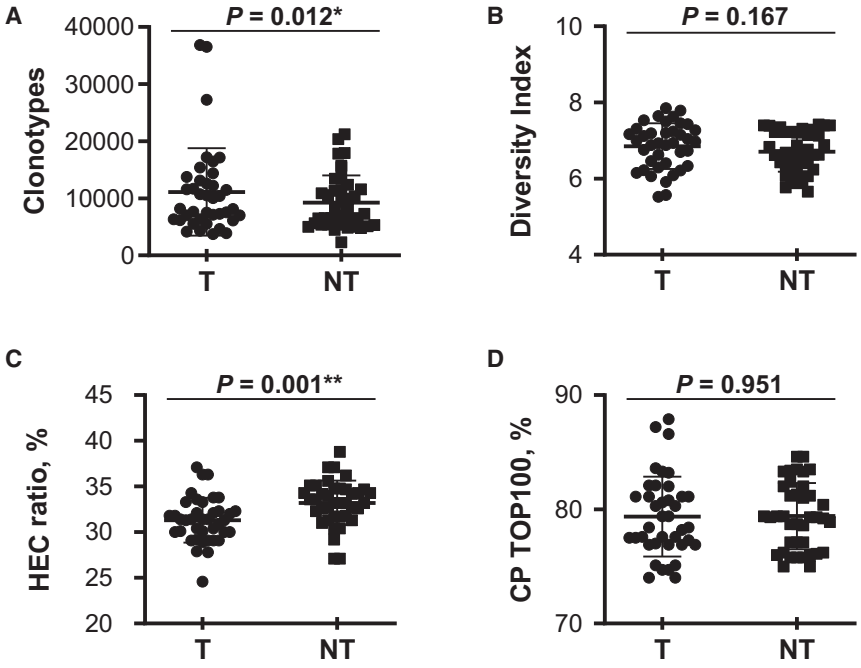


Figure 1. Comparison of T Cell Receptor β Repertoires between Tumor Tissues and Paired Adjacent Non-Tumor Tissues

(A) Clonotypes. (B) Shannon-Wiener Index. (C) HECs (highly expanded clones, >0.1%) ratio. (D) CP TOP100, cumulative percentage of top 100 frequent TCRβ in each sample. The differences were compared using paired t test between groups. T, tumor tissues; NT, paired adjacent non-tumor tissues.

results showed that clonotype numbers in tumor tissues were significantly higher than those in paired adjacent non-tumor tissues ($p = 0.012$, Figure 1A). Similarly, the diversity indices for tumor samples were also greater than those for adjacent non-tumor samples, though the statistic is not significant ($p = 0.167$, Figure 1B). Consistent with the increased clonotypes and diversity (Figures 1A and 1B), the frequency distribution of TCR repertoires was also more dispersive in tumors, which was reflected by the significantly decreased ratio of highly expanded clones (Figure 1C). These results indicated that there were more divergent TCR repertoires in tumors than in adjacent non-tumor tissues for NSCLC, although the cumulative percentage of top 100 frequent TCRβ showed no difference between them (Figure 1D).

TCRβ V-J combinations usage frequency between tumors and adjacent non-tumor tissues were compared using the paired t test method and the p values were adjusted for multiple testing by the Benjamini Hochberg method.[24] Finally, the usage frequency of 185 V-J combinations were found significantly different between tumors and matched adjacent non-tumor tissues ($p < 0.05$). The usage distribution of these significantly differentiated V-J genes across these samples were shown in Figure 2A, and the top upregulated or downregulated V-J genes were reflected by fold change of T versus NT (tumor tissues versus paired adjacent non-tumor tissues; Figure 2B).

To explore the potential clinical meaning of differential tumor TCR repertoires, we analyzed the relation between the 185 significantly differentiated V-J genes and clinical pathogenic features by the association analysis of V-J genes usage frequency and the value of these clinical characteristics. The results showed that there were some specific V-J combinations positively or negatively associated with clinical characteristics significantly (Figure S2).

Association Analysis of Tumor TCR Repertoires with Clinical Characteristics

To further characterize the profiling of tumor TCR repertoires, we investigated associations between clinical pathogenic features and intratumoral TCR repertoires indices. First, all patients were divided into subgroups according to basic clinical features, such as age (≤ 60 and >60), gender (male and female), pathological type (adenocarcinoma and squamous cell cancer), and TNM stage (I/II and III/IV), respectively. Then the repertoires indices of TCR including clonotype numbers, diversity index, HECs ratio, and cumulative percentage of the top 100 frequent TCRβ were compared between each subgroups by the unpaired t test. The results showed that there was a significant association between clonotype numbers and TNM stage (Figure S3), i.e., the intratumoral TCR clonotype numbers in late-stage patients were significantly less than those in early stage ($p = 0.034$, Figure S3). Similarly, for biochemical indices, all patients were divided into low subgroups and high subgroups using the median value as cut-off, respectively. The repertoires indices of TCR were compared between the two subgroups using unpaired t test, however, no significant associations were found (Figure S4).

Overlap ratio was often used to access the similarity of repertoires between samples.[14,21,25] The T cell clone overlap ratio between tumors and paired adjacent non-tumor tissues was defined as the number of shared unique TCRβ V-J combinations divided by the number of unique TCRβ V-J combinations detected in tumors. Here, overlap ratio of all TCRβ detected and that of the top 100 frequent TCRβ were both calculated and compared between subgroups divided by basic clinical features and biochemical indices. The results showed neither overlap ratio of all TCRβ detected nor that of the top 100 frequent TCRβ have a significant association with clinical pathogenic features (Figures S3 and S4).

Weighted gene co-expression network analysis (WGCNA) is one of the most popular methods to identify clusters (modules) of highly correlated genes, and relate co-expression modules to external sample traits.[26] We adjusted the WGCNA algorithm to investigate the

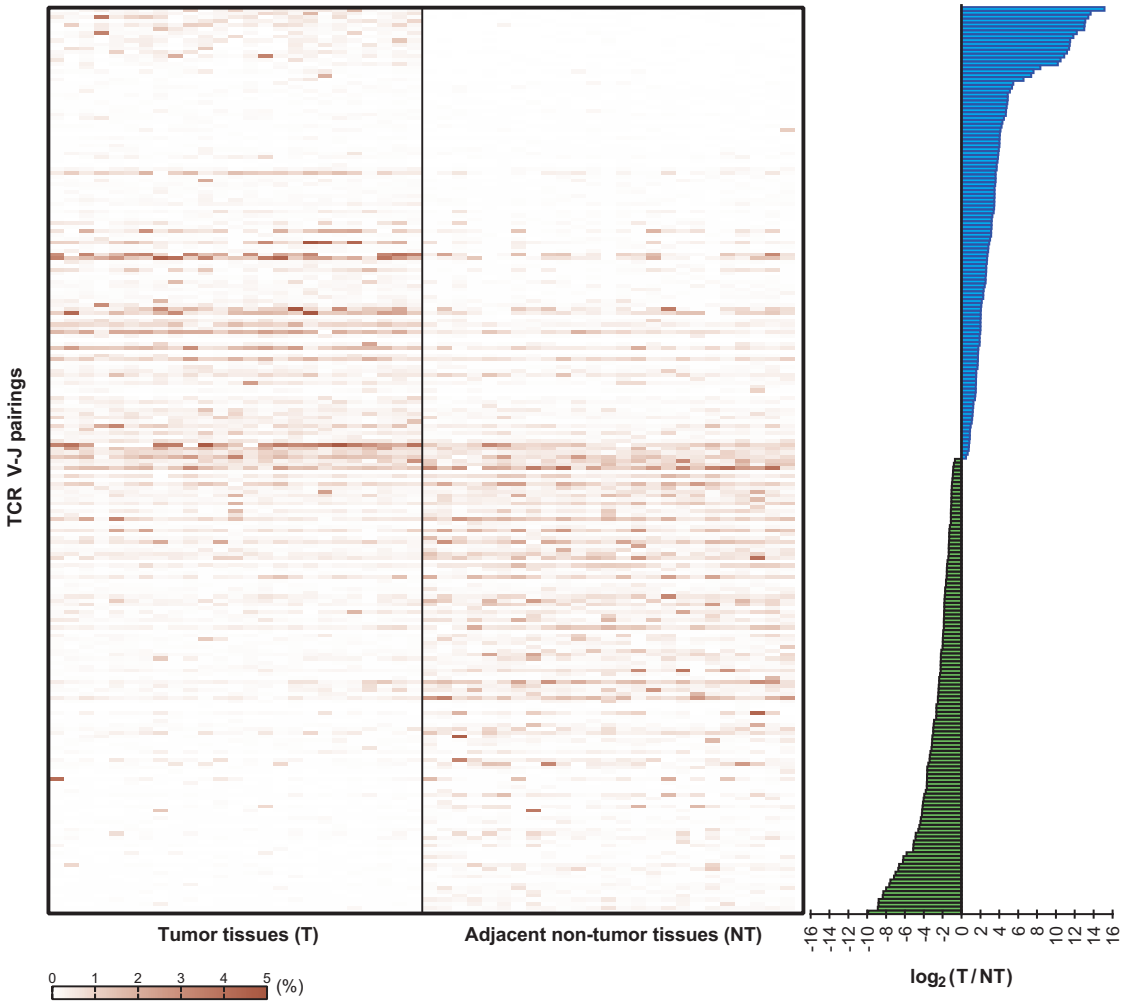


Figure 2. The Analysis of V-J Combination in Tumor Tissues and Paired Adjacent Non-Tumor Tissues

Left panel shows heatmap of the usage frequency of V-J significantly differentiated between tumor tissues and paired adjacent non-tumor tissues, and right panel showes the fold change of the significantly differentiated V-J pairs. The fold change was calculated by the usage frequency in tumor tissue divided by that in paired adjacent non-tumor tissue. T, tumor tissues; NT, paired adjacent non-tumor tissues.

networks of V-J genes in TCR repertoires for better understanding the biological meaning of V-J correlation. Specifically, we examined the pairwise correlations between V-J genes in their usage frequency across the enrolled subjects. Because the curve of soft threshold (SFT) graph reached a saturation point when the SFT power was 5 (Figure 3A), this value was chosen to subsequently construct network and identify modules. Hierarchical clustering tree was established to detect outliers, and ultimately 1 highly associated V-J module was identified (Figure S5A; Figures 3B and 3C). This identified V-J gene module presents a cluster of highly interconnected V-J combinations, and the V-J genes of this module have high mutual correlation coefficients. To analyze the association between V-J gene module and the clinical features of enrolled patients, we performed the module-trait relationship analysis with several clinical pathogenic characteristics including age, sex, type, TNM stage, CRP, $CD8^+$, PD-L1, CRP, neutrophil ratio, and white blood cell content. The results showed the identified V-J gene module was significantly associated with age (Figure 3D).

The co-expression network was exported into Cytoscape software,[27] which could help us to further depict the V-J genes network of module related to clinical features. Each node represents an individual V-J gene in the network, and each edge represents the interactions between genes. The results of V-J genes interaction network visualized by Cytoscape showed there were 105 nodes and 491 edges in the identified module (Figure S5B).

Association Analysis of Tumor TCR Repertoires with Somatic Mutations

Whole-exome sequencing was performed for 25 subjects randomly selected from all 39 patients. On average, 7.61 Gb of clean bases

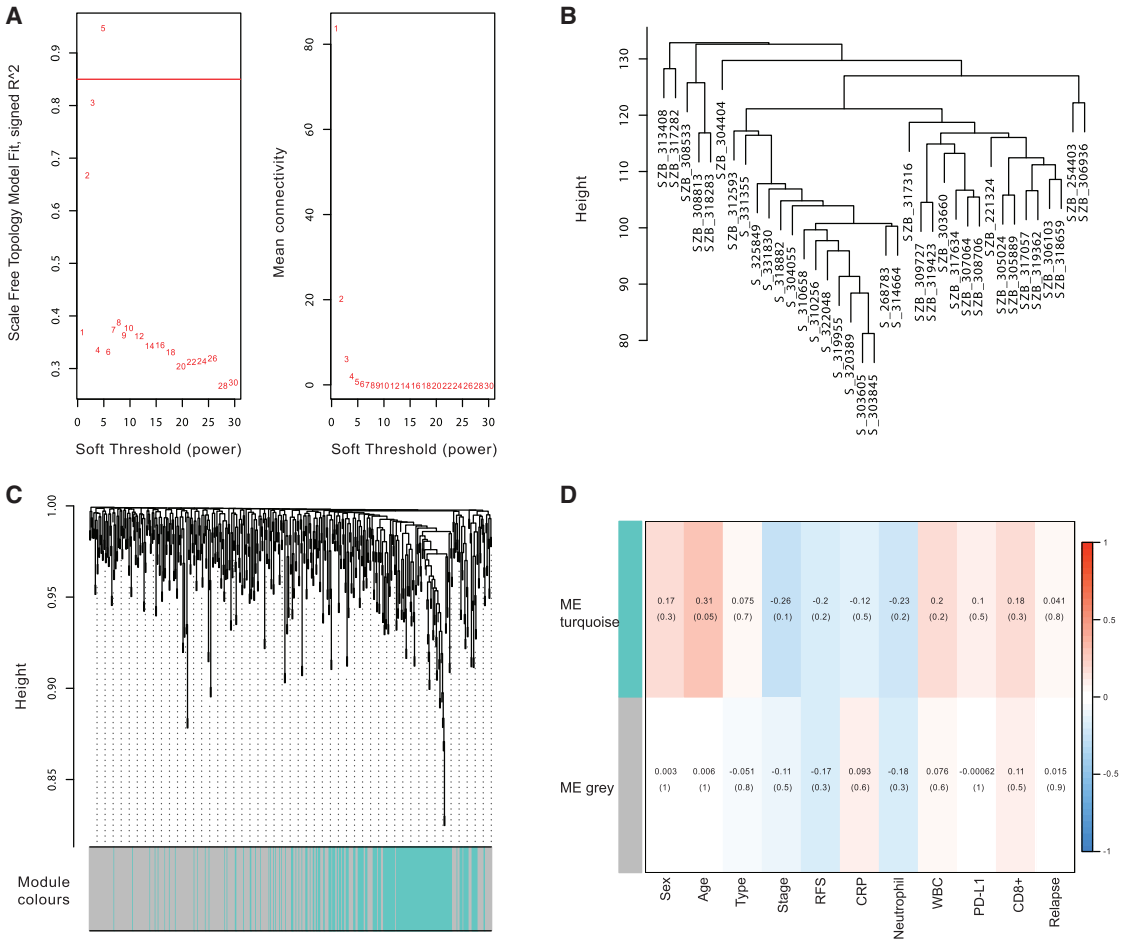


Figure 3. Weighted Gene Co-Expression Network Analysis of V-J Combination of T Cell Receptor β Genes

(A) Soft threshold value calculated by the standard of the scale-free network. (B) Hierarchical clustering tree established to detect outliers. (C) Identification of a co-expression module in V-J genes of TCR repertoires. (D) Module-trait relationship analysis between the identified module and clinical features. Each row represents a module, and each column represents one clinical feature. The correlation coefficient between the module and each clinical feature was indicated by the color of each cell at the row-column intersection.

were generated from each sample, and 94.73% of reads were uniquely aligned to reference genome. The depth on target ranged from 30.77 to ~241.26× with the average 110.40× depth, and mean ratio of target coverage was 99.67% (Table S1). Somatic mutations were identified by the exome-targeted sequencing of these tumor tissues and paired adjacent non-tumor tissues. The top frequently mutated genes such as *EGFR*, *TP53*, and *TTN*, which were well-known highly mutated in NSCLC in previous studies, were also identified in these 25 subjects (Figure S6).

It is known that T cells preferably target the neoantigens encoded by tumor-specific somatic mutations. To explore the potential roles of these frequent mutation identified genes in affecting the profiling of TCR repertoires, we analyzed the association between the somatic mutation status of genes and the status of intratumoral TCR repertoires. First, 25 subjects were divided into WT (wild-type) subgroup and MT (mutation) subgroup, according to the somatic mutation status of each gene, respectively. Then the repertoires indices of TCR including clonotype numbers, diversity index, HECs ratio, and cumulative percentage of TOP100 frequent TCRβ were simultaneously compared between each subgroups by unpaired t test method. The results showed there were no significant differences of TCR repertoire profiling between tumors with *EGFR* mutations and those without *EGFR* mutations, as well as *TP53* mutation subgroups (Figures 4A–4H). Notably, although both the clonotypes number and HECs ratio of tumors with *TTN* mutations were non-significantly less than those without *TTN* mutations ($p = 0.284$, Figure 4I; $p = 0.220$, Figure 4L), the TCRβ diversity index of tumors were significantly reduced when *TTN* gene was mutated ($p = 0.002$, Figure 4J). Moreover, the cumulative percentage of TOP100 frequent TCRβ of tumors were simultaneously increased significantly while *TTN* gene had mutations ($p = 0.015$, Figure 4K). Correspondingly, neither TCRβ diversity

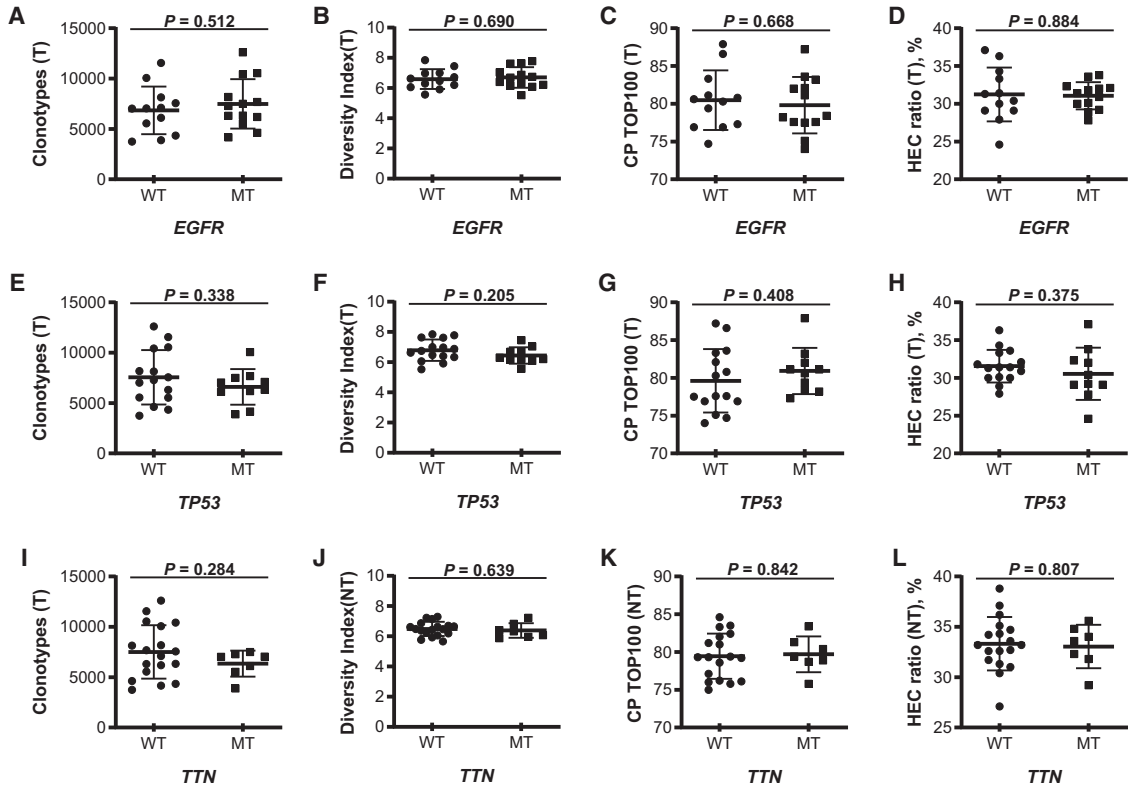


Figure 4. Association Analysis between Tumor TCRβ Repertoires and Top Mutated Genes

Groups were respectively divided by the mutation status of (A–D) *EGFR*, (E–H) *TP53*, and (I–L) *TTN*. Differences between groups were compared by unpaired t test. WT, wild-type. MT, mutation.

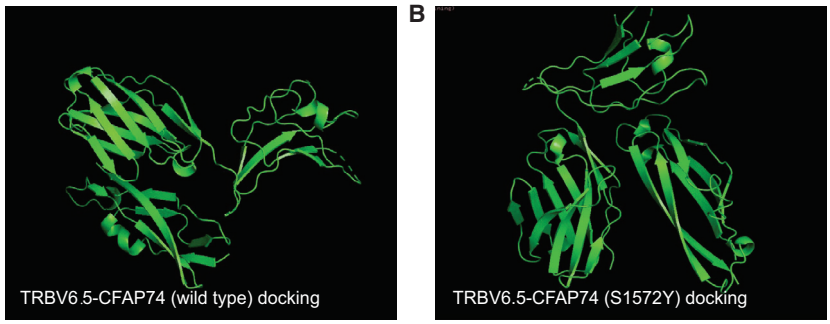
index nor cumulative percentage of TOP100 frequent TCRβ of adjacent non-tumor tissues showed significant changes regardless of the tumor-mutation status of *TTN* gene (Figures S7A–S7D).

In addition, neither overlap ratio of all detected TCRβ nor overlap ratio of TOP100 frequent TCRβ, were found significant changes related to the somatic mutation status of *EGFR* (Figures S8A and S8D) or *TP53* (Figures S8B and S8E). The overlap ratio of all detected TCRβ between tumors and adjacent non-tumor tissues was also not obviously affected by *TTN* mutation or not (Figure S8C), but the overlap ratio of TOP100 frequent TCRβ was significantly decreased when *TTN* mutated in tumors ($p = 0.022$, Figure S8F).

Since specific antigen recognition by TCRs is a critical process in the adaptive immune response, exome-wide association analysis between somatic mutations and TCRβ V-J combinations was performed to explore the unique relationship between them in this study. The results showed there were a certain number of V-J combinations and gene mutations pairs could be observed with significant associations ($p < 0.01$, Figure S9). To explore the potential recognition relationship of these TCRβ type-peptide pairs, we analyzed binding affinities with structures for TCRβ-peptide complexes. TCRβ-peptide pairs with observed association (e.g., TRBV6.5-CFAP74 [S1572Y] peptide) and those with no association (e.g., TRBV10.3-HSP90B1 [V160L] peptide), were respectively selected to analyze further. Specifically, the 3D crystal structure of TRBV6.5 family (PDB: 1BD2) in Protein Data Bank,[28] and CFAP74 peptide with WT or S1572Y mutation were used for docking analysis (Figures 5A and 5B). The result showed that total score of TRBV6.5-CFAP74 (S1572Y) peptide was observed lower than that of TRBV6.5-CFAP74 (WT) peptide (Figure 5C), suggesting the more likely native structure occurred by TRBV6.5-CFAP74 (S1572Y), instead of TRBV6.5-CFAP74 (WT) peptide. Meanwhile, the docking results of TRBV10.3 family 3D crystal structure (PDB: 3QEQ)-CFAP74 (WT) pair, and TRBV10.3 family 3D crystal structure-CFAP74 (V160L) pair (Figures 5D and 5E) showed that the very similar total scores were observed between two pairs (Figure 5F).

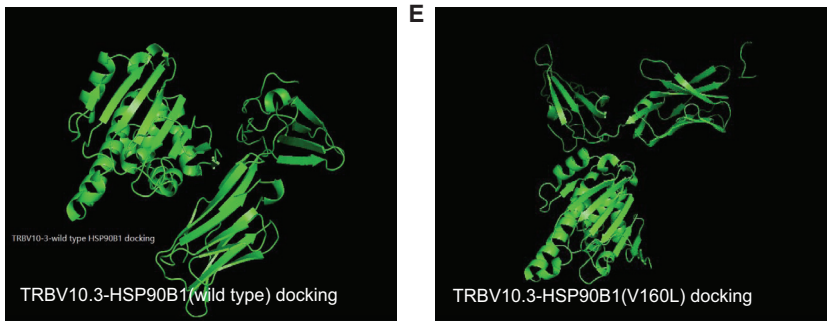
Prognostic Analysis of TCRβ V-J Pairing Usage in Postoperative Recurrence

As numerous TCRβ V-J genes showed differential usages among individuals, we continue to investigate whether the usage of TCRβ V-J genes have potential prognostic roles in postoperative recurrence of patients. The Gini Index was used to select the most important features to construct classifier model. As Figure 6A showed, top seven important factors including *TRBV7.7-TRBJ2.5*, *TRBV7.4-TRBJ2.5*, *TRBV4.1-TRBJ2.7*, *TRBV20.1-TRBJ1.3*, *TRBV5.8-TRBJ1.3*,



C

TCR type-gene pairs	Total score	rms
TRBV6.5-CFAP74 (wild type) docking	-574.25	66.90
TRBV6.5-CFAP74 (S1572Y) docking	-632.80	30.80



F

TCR type-gene pairs	Total score	rms
TRBV10.3-HSP90B1(wild type) docking	-400.92	84.35
TRBV10.3-HSP90B1(V160L) docking	-399.85	105.56

Figure 5. Binding Affinities Analysis of Structures for TCRβ-WT and TCRβ-MT Peptide Pairs

(A) TRBV6.5-CFAP74 (WT) docking 3D image. (B) TRBV6.5-CFAP74 (S1572Y) docking 3D image. (C) Docking scores and RMS of TRBV6.5-CFAP74 peptide pairs. (D) TRBV10.3-HSP90B1 (WT) docking 3D image. (E) TRBV10.3-HSP90B1 (V160L) docking 3D image. (F) Docking scores and RMS of TRBV10.3-HSP90B1 peptide pairs.

TRBV10.2-TRBJ2.5, and *TRBV7.6-TRBJ2.2* were featured to construct the classifier model. As a result, the patients with postoperative recurrence were classified from those without recurrence (AUC = 0.891) using these featured TCRβ V-J pairings usage (Figure 6B), while the accuracy of prediction was slightly increased (AUC = 0.929) when these featured TCRβ V-J pairings were combined with the status of somatic mutations of genes such as *EGFR* and *ANKRD36C* (Figure 6C).

DISCUSSION

TCR repertoires of infiltrating T cells are now emerging as potential important biomarkers in the development of cancers with a comprehensive impact on clinical outcomes.[10] In current study, we comprehensively characterized TCR repertoires in tumors and adjacent non-tumor tissues from patients with resectable NSCLC, and hypothesized that TCR repertoires, specifically certain V-J pairings usage, could be potentially used for prognosis analysis for NSCLC patients who received curative surgery.

By high-throughput sequencing of TCRβ in tumors and matched adjacent non-tumor tissues, it showed that T cell clonotype number[23] in tumors was higher and the ratio of highly expanded clones (clones whose frequency >0.1%) in tumors was lower, than those in adjacent non-tumor tissues, respectively (Figure 1). These results were consistent with previous findings that T cells infiltrating in tumor lesions comprise high numbers of clonal expansion[29] and that the proliferation of oligoclonal T cells in tumor lesions would be massively induced in response to tumor-associated antigenic peptides.[30] Our study showed the diversity of TCRs, which was calculated by Shannon diversity index[31] had no difference compared with adjacent non-tumor tissues in NSCLC (Figure 1). However, a recent study that presented data from 15 lung cancer patients showed that significantly higher TCR diversity was observed in tumor tissues than in normal lung tissues,[32] the increases or decreases of the diversity of TCRs in tumors relative to paired normal tissues are inconsistent in different cancer types[33–35] or even same type in different studies.[16,17] For example, compared with adjacent non-tumor tissues, TCR diversity of tumors is decreased in colorectal cancer[33] and clear cell renal cell carcinomas,[34] increased or not changed in hepatocellular carcinoma with inconsistent results in two studies,[16,17] and not changed in esophageal squamous cell carcinoma.[35] These inconsistencies probably resulted from the intrinsic difference of T cell infiltration in paired tissues from distinct individuals and need to be confirmed in the future study. Interestingly, the intratumoral TCR clonotypes numbers in early-stage tumor were suggested to be higher those in late-stage (Figure S3), consistent with previous results that clonotypes were less when the tumor size was larger or the number of metastatic organs was greater in advanced lung cancer.[19] This potential association between the TNM stage and TCR clonotypes suggests that the TCR repertoires' dynamic change along with malignant progression and the defective anti-tumor immunity in NSCLC.

Furthermore, the observed V-J pairings with different usage between tumors and matched adjacent non-tumor tissues (Figure 2), implying a proportion of T cells infiltrating tumors were tumor-specific, which recognize tumor antigens through their native receptors. Some V-J pairings were observed associated with several clinical characteristics (Figure S2), suggesting TCRβ chain V-J rearrangement could reflect

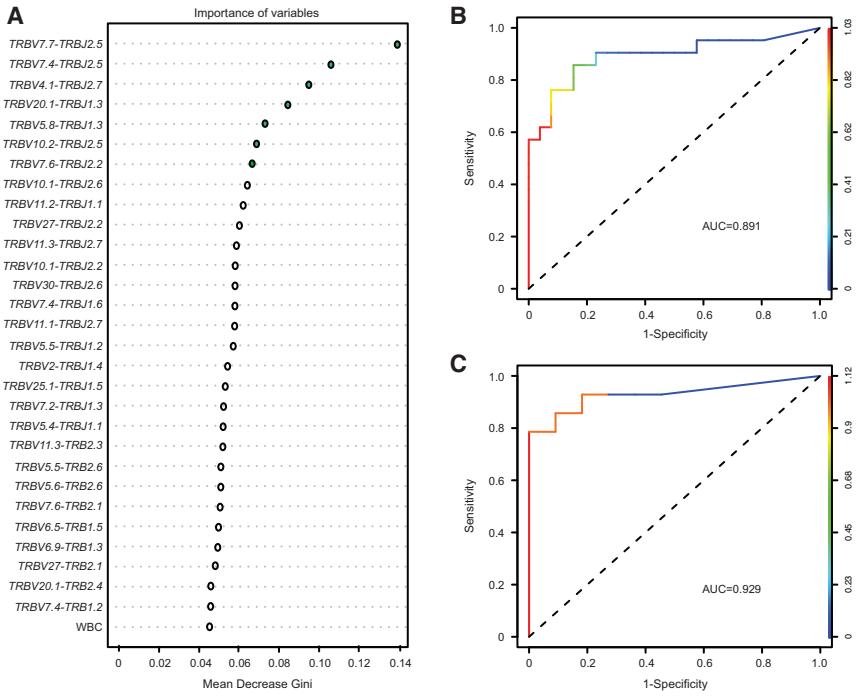


Figure 6. The Performances of Two Models in Predicting the Recurrence of Resectable Non-Small Cell Lung Cancer Patients

(A) Important features selected by the Gini Index. (B) The model using V-J genes alone. (C) The model using the combination of V-J genes and somatic mutations.

the prognosis of patients. Correspondingly, previous studies have shown that the results of TCR gene rearrangements could be prognosis markers in acute lymphoblastic leukemia (ALL).[36] Not only that, we observed a specific module of V-J genes using WGCNA method,[26] the relationship between highly correlated V-J genes in the module and clinical features were investigated (Figure 3). As far as we know, this is the first study to identify the module of V-J genes usage contributing to the TCR repertoires associated with clinical pathogenic characteristics. The observed association suggested the contributing roles of modules in divergent tumor immunity among individuals.

Specific TCR gene rearrangements recognize tumor antigens generated by somatic mutations, as the mutation in cancer could create a neoepitope.[37] Through whole-exome sequencing of these paired samples, top frequently somatic-mutated genes such as *EGFR*, *TP53*, and *TTN*, etc., whose mutations were well-known in lung cancer, were further studied with TCR repertoires. Although it has been previously reported that *EGFR* mutant tumors had a higher TCRβ diversity index than WT tumors,[38] difference of TCR repertoires in tumors with and without *EGFR* mutation was not observed in our study (Figures 4A–4D). Instead, *TTN* gene somatic mutation was shown to affect the profiling of intratumoral TCR repertoires (Figures 4I–4L), this results could be one of reasons of the correlation of *TTN* mutations with favorable prognosis in lung squamous cell carcinoma.[39] Finally, seven TCRβ V-J pairings including *TRBV7.7-TRBJ2.5*, *TRBV7.4-TRBJ2.5*, *TRBV4.1-TRBJ2.7*, *TRBV20.1-TRBJ1.3*, *TRBV5.8-TRBJ1.3*, *TRBV10.2-TRBJ2.5*, and *TRBV7.6-TRBJ2.2* were featured to construct a classifier model to predict postoperative recurrence. These featured V-J pairings involved *TRBV4.1*, *TRBV5.8*, *TRBV7.4*, *TRBV7.6*, *TRBV7.7*, *TRBV10.2*, and *TRBV20.1* for V gene segments, and *TRBJ1.3*, *TRBJ2.2*, *TRBJ2.5*, and *TRBJ2.7* for J gene segments. Specific types or combinations of TCR seem to perform different biological functions. T cells from DQ5+MuSK-MG patients showed that significant difference was observed in frequencies of *TRBJ2.5* and other V genes.[40] Certain V-J combinations, including *TRBJ2.5* and other *TRBV*s, were highly shared in the nonobese diabetic mice.[41] The common usage of *TRBV7* was identified in eight of nine LHA B57 subjects to clarify the roles of TCR diversity controlling HIV-1 viremia; the data demonstrated that these clonotypes, such as *TRBV7*, were selectively recruited over the course of chronic HIV-1 infection.[42] A study of the correlation between immune repertoire and CRC revealed that the usage of *TRBV7.8*, *TRBV7.9*, *TRBJ2.2*, and *TRBJ2.5* had significant difference between the patients and healthy control groups.[43] The result that these featured TCR parings could be used as a prognostic classifier (Figure 6), suggests that they may have important roles in activating certain immunity responses and further affecting the recurrence of lung cancer. Certainly, more evidences are needed to support the assumption. Moreover, the classifier using TCRβ V-J genes combined with somatic mutations only showed a slightly better predictive performance than that using V-J genes alone (Figure 6), implying the pivotal roles of TCR repertoires in prognosis prediction.

It is worth noting that there are still limitations for our study. The sample size was not big and all patients were enrolled from a single center, and these may result in challenges to have more concrete conclusions. Therefore, a multicenter study with a larger cohort would be required to further validate these results. Nevertheless, our study first comprehensively characterized the tumor-related TCR repertoires in patients receiving curative surgery for NSCLC and hypothesized the potential roles of the repertoires in predicting prognosis after surgical operation. The preliminary results of this study provide clues and basis for future more in-depth research.

MATERIALS AND METHODS

Patients and Samples Collection

This study enrolled 39 patients, who were diagnosed as NSCLC and received surgical resection at Zhejiang Cancer Hospital. None of them had other immune-related diseases, such as infectious diseases and autoimmunity diseases, or other tumors. Resected tumor samples and paired adjacent normal tissues were confirmed independently by

3 independent pathologists with extensive clinical experience. Samples were then formalin fixed and paraffin embedded at the time of biopsy. This study was conducted in accordance with the declaration of Helsinki and approved by the Ethics Committee of Zhejiang Cancer Hospital.

High-Throughput TCR Sequencing

Genomic DNA from each tumor sample or paired adjacent normal tissue was extracted by CWE2100 FFPE DNA Kit (Cwbiotech, Beijing, China) according to the manufacturer's protocol. The quantity and purity of extracted DNA were determined using NanoDrop 2000 Spectrophotometer. The CDR3 regions of TCRβ chain (TRB) were amplified from genomic DNA using the multiplex PCR method. Specifically, to enrich all possible V(D)J combinations, a panel of 51 V-forward and 14 J-reverse primers was used in amplification with Platinum Multiplex PCR Master Mix (Invitrogen). The first PCR reaction is as follows: 95°C for 3 min; 30 cycles of 95°C for 30 s, 64.3°C for 1 min, 72°C for 30 s; 72°C for 5 min; 4°C hold. The PCR products were purified by Ampure beads and ligated with Abclone adapters (RK20284, RK20295, ABclonal Technology). The second PCR step is as follows: 98°C for 45 s; 8 cycles of 98°C for 10 s, 60°C for 30 s, 72°C for 30 s; 72°C for 1 min; 12°C hold.

Libraries were loaded onto the Illumina XTen System, and reads of 151-bp fragments were sequenced. The CDR3 sequence was defined as the amino acids between the second cysteine of the V region and the conserved phenylalanine of the J region, according to the ImMunoGeneTics (IMGT) V, D, and J gene references. The CDR3 sequences were identified and assigned using the MiXCR software package.[44]

Immunohistochemical Detection of CD8-Positive Tumor-Infiltrating Lymphocytes, and PD-L1 Protein Expression

Tumor tissue sections of 4∼5 μm thick were deparaffinized and hydrated. Monoclonal antibody (4B11, RTU, Leica Biosystems, Buffalo Grove, IL, USA) was used to detect CD8-positive tumor-infiltrating lymphocytes (TILs) by an automated stainer (BOND RX, Leica Microsystems). In addition, primary antibodies against PD-L1 (SP263; 1:2000; Roche VENTANA, Tucson, AZ, USA) was used to perform immunohistochemical analysis. The PD-L1 immunohistochemistry results were evaluated based on the degree and intensity of cell membrane staining, in which a tumor proportion score (TPS) >25% was defined as high expression. Based on the extent of positive 25 lymphocytes infiltrating within tumor cells, CD8-positive TILs were evaluated semi-quantitatively on a scale of 0–3. Each score was calculated as the fraction of tumor cells on top of which CD8-positive T cells were present. Specifically, 0, none or rare; 1, <5%; 2, >5% to <25%; 3, >25%. A cell is CD8-positive if its score is 2 or 3 and negative otherwise.

Statistical Analysis

The associations among clinical pathological characteristics, somatic mutations and TCR repertoire indices were analyzed by the paired or unpaired t test where applicable and the p value was adjusted by Benjamini-Hochberg method for multiple testing.[24] An association was considered as statistically significant if its testing p value was less than 0.05. Recurrence-free survival rates were estimated using the Kaplan-Meier method, and the associations between the time to recurrence and various features studied (such as basic clinical features, biochemical indices, somatic mutations, and TCR repertoire indices) were evaluated using the Cox proportional hazards regression model. All the statistical analyses were performed using the Graphpad Prism (version 8.0) software. For conjoint analysis between the clinical characteristics, TCRβ combination frequencies and somatic mutations, an algorithm adjusting Matrix eQTL[45] was developed to calculate and adjust the p value by the Benjamini-Hochberg method.

Computational Model to Predict the Risk of Postoperative Recurrence

TCR frequency alone and TCR frequency in conjunction with somatic mutation were used to establish prediction models for the risk of postoperative recurrence of patients. Since the numbers of TCRs and somatic mutations are much larger than the number of samples, a random forest method was first used to select features. Based on the reduced features, a logistic regression model was established for predicting the risk of postoperative recurrence. The receiver operating characteristic (ROC) curve was used to evaluate the sensitivity and specificity of the models.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.omtm.2020.05.020.

AUTHOR CONTRIBUTIONS

Z.S., X.C., J.Y., and G.C. designed the experimental work and wrote the manuscript; Z.S., Y.S., R.H., and W.W. conducted the experiments; Z.S., X.C., K.Z., S.L., M.W., G.T., J.Y., and G.C. evaluated the results; Z.S., G.C. received funds.

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

This work was supported by grants from Science and Technology Program of Fujian Province, China (grant numbers 2018Y2003, 2019L3018, and 2019YZ016006) and National Natural Science Foundation of China (grant number 81802276).

REFERENCES

1. Mountain, C.F. (1997). Revisions in the International System for Staging Lung Cancer. Chest *111*, 1710–1717.

2. Naruke, T., Tsuchiya, R., Kondo, H., and Asamura, H. (2001). Prognosis and survival after resection for bronchogenic carcinoma based on the 1997 TNM-staging classification: the Japanese experience. Ann. Thorac. Surg. *71*, 1759–1764.

3. Howington, J.A., Blum, M.G., Chang, A.C., Balekian, A.A., and Murthy, S.C. (2013). Treatment of stage I and II non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest *143*, e278S–e313S.

4. Bindea, G., Mlecnik, B., Fridman, W.H., Pagès, F., and Galon, J. (2010). Natural immunity to cancer in humans. Curr. Opin. Immunol. *22*, 215–222.

5. Chen, K.J., Zhou, L., Xie, H.Y., Ahmed, T.E., Feng, X.W., and Zheng, S.S. (2012). Intratumoral regulatory T cells alone or in combination with cytotoxic T cells

predict prognosis of hepatocellular carcinoma after resection. Med. Oncol. *29*, 1817–1826.

6. Goeppert, B., Frauenschuh, L., Zucknick, M., Stenzinger, A., Andrulis, M., Klauschen, F., Joehrens, K., Warth, A., Renner, M., Mehrabi, A., et al. (2013). Prognostic impact of tumour-infiltrating immune cells on biliary tract cancer. Br. J. Cancer *109*, 2665–2674.

7. Stanton, S.E., Adams, S., and Disis, M.L. (2016). Variation in the Incidence and Magnitude of Tumor-Infiltrating Lymphocytes in Breast Cancer Subtypes: A Systematic Review. JAMA Oncol. *2*, 1354–1360.

8. Galon, J., Mlecnik, B., Bindea, G., Angell, H.K., Berger, A., Lagorce, C., Lugli, A., Zlobec, I., Hartmann, A., Bifulco, C., et al. (2014). Towards the introduction of the 'Immunoscore' in the classification of malignant tumours. J. Pathol. *232*, 199–209.

9. Li, B., Li, T., Pignon, J.C., Wang, B., Wang, J., Shukla, S.A., Dou, R., Chen, Q., Hodi, F.S., Choueiri, T.K., et al. (2016). Landscape of tumor-infiltrating T cell repertoire of human cancers. Nat. Genet. *48*, 725–732.

10. Jiang, N., Schonnesen, A.A., and Ma, K.Y. (2019). Ushering in Integrated T Cell Repertoire Profiling in Cancer. Trends Cancer *5*, 85–94.

11. Nikolich-Zugich, J., Slifka, M.K., and Messaoudi, I. (2004). The many important facets of T-cell repertoire diversity. Nat. Rev. Immunol. *4*, 123–132.

12. Matos, T.R., de Rie, M.A., and Teunissen, M.B.M. (2017). Research Techniques Made Simple: High-Throughput Sequencing of the T-Cell Receptor. J. Invest. Dermatol. *137*, e131–e138.

13. Manuel, M., Tredan, O., Bachelot, T., Clapisson, G., Courtier, A., Parmentier, G., Rabeony, T., Grives, A., Perez, S., Mouret, J.F., et al. (2012). Lymphopenia combined with low TCR diversity (divpenia) predicts poor overall survival in metastatic breast cancer patients. OncoImmunology *1*, 432–440.

14. Wang, T., Wang, C., Wu, J., He, C., Zhang, W., Liu, J., Zhang, R., Lv, Y., Li, Y., Zeng, X., et al. (2017). The Different T-cell Receptor Repertoires in Breast Cancer Tumors, Draining Lymph Nodes, and Adjacent Tissues. Cancer Immunol. Res. *5*, 148–156.

15. Beausang, J.F., Wheeler, A.J., Chan, N.H., Hanft, V.R., Dirbas, F.M., Jeffrey, S.S., and Quake, S.R. (2017). T cell receptor sequencing of early-stage breast cancer tumors identifies altered clonal structure of the T cell repertoire. Proc. Natl. Acad. Sci. USA *114*, E10409–E10417.

16. Chen, Y., Xu, Y., Zhao, M., Liu, Y., Gong, M., Xie, C., Wu, H., and Wang, Z. (2016). High-throughput T cell receptor sequencing reveals distinct repertoires between tumor and adjacent non-tumor tissues in HBV-associated HCC. OncoImmunology *5*, e1219010.

17. Lin, K.R., Deng, F.W., Jin, Y.B., Chen, X.P., Pan, Y.M., Cui, J.H., You, Z.X., Chen, H.W., and Luo, W. (2018). T cell receptor repertoire profiling predicts the prognosis of HBV-associated hepatocellular carcinoma. Cancer Med. *7*, 3755–3762.

18. Hopkins, A.C., Yarchoan, M., Durham, J.N., Yusko, E.C., Rytlewski, J.A., Robins, H.S., Laheru, D.A., Le, D.T., Lutz, E.R., and Jaffee, E.M. (2018). T cell receptor repertoire features associated with survival in immunotherapy-treated pancreatic ductal adenocarcinoma. JCI Insight *3*, e122092.

19. Liu, Y.Y., Yang, Q.F., Yang, J.S., Cao, R.B., Liang, J.Y., Liu, Y.T., Zeng, Y.L., Chen, S., Xia, X.F., Zhang, K., and Liu, L. (2019). Characteristics and prognostic significance of profiling the peripheral blood T-cell receptor repertoire in patients with advanced lung cancer. Int. J. Cancer *145*, 1423–1431.

20. Reuben, A., Gittelman, R., Gao, J., Zhang, J., Yusko, E.C., Wu, C.J., Emerson, R., Zhang, J., Tipton, C., Li, J., et al. (2017). TCR Repertoire Intratumor Heterogeneity in Localized Lung Adenocarcinomas: An Association with Predicted Neoantigen Heterogeneity and Postsurgical Recurrence. Cancer Discov. *7*, 1088–1097.

21. Kuang, M., Cheng, J., Zhang, C., Feng, L., Xu, X., Zhang, Y., Zu, M., Cui, J., Yu, H., Zhang, K., et al. (2017). A novel signature for stratifying the molecular heterogeneity of the tissue-infiltrating T-cell receptor repertoire reflects gastric cancer prognosis. Sci. Rep. *7*, 7762.

22. Cui, J.H., Lin, K.R., Yuan, S.H., Jin, Y.B., Chen, X.P., Su, X.K., Jiang, J., Pan, Y.M., Mao, S.L., Mao, X.F., and Luo, W. (2018). TCR Repertoire as a Novel Indicator for Immune Monitoring and Prognosis Assessment of Patients With Cervical Cancer. Front. Immunol. *9*, 2729.

23. Liaskou, E., Klemsdal Henriksen, E.K., Holm, K., Kaveh, F., Hamm, D., Fear, J., Viken, M.K., Hov, J.R., Melum, E., Robins, H., et al. (2016). High-throughput T-cell receptor sequencing across chronic liver diseases reveals distinct disease-associated repertoires. Hepatology *63*, 1608–1619.

24. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. B *57*, 289–300.

25. Alves Sousa, A.P., Johnson, K.R., Ohayon, J., Zhu, J., Muraro, P.A., and Jacobson, S. (2019). Comprehensive Analysis of TCR-β Repertoire in Patients with Neurological Immune-mediated Disorders. Sci. Rep. *9*, 344.

26. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

27. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

28. Rose, P.W., Beran, B., Bi, C., Bluhm, W.F., Dimitropoulos, D., Goodsell, D.S., Prlic, A., Quesada, M., Quinn, G.B., Westbrook, J.D., et al. (2011). The RCSB Protein Data Bank: redesigned web site and web services. Nucleic Acids Res. *39*, D392–D401.

29. Thor Straten, P., Schrama, D., Andersen, M.H., and Becker, J.C. (2004). T-cell clonotypes in cancer. J. Transl. Med. *2*, 11.

30. Sun, Y., Möller, P., Berking, C., Schlüpen, E.M., Volkenandt, M., and Schadendorf, D. (1999). In vivo selective expansion of a tumour-specific cytotoxic T-cell clone derived from peripheral blood of a melanoma patient after vaccination with gene-modified autologous tumour cells. Immunology *98*, 535–540.

31. Zhang, W., Du, Y., Su, Z., Wang, C., Zeng, X., Zhang, R., Hong, X., Nie, C., Wu, J., Cao, H., et al. (2015). IMonitor: A Robust Pipeline for TCR and BCR Repertoire Analysis. Genetics *201*, 459–472.

32. Wang, X., Zhang, B., Yang, Y., Zhu, J., Cheng, S., Mao, Y., Feng, L., and Xiao, T. (2019). Characterization of Distinct T Cell Receptor Repertoires in Tumor and Distant Non-tumor Tissues from Lung Cancer Patients. Genomics Proteomics Bioinformatics *17*, 287–296.

33. Sherwood, A.M., Emerson, R.O., Scherer, D., Habermann, N., Buck, K., Staffa, J., Desmarais, C., Halama, N., Jaeger, D., Schirmacher, P., et al. (2013). Tumor-infiltrating lymphocytes in colorectal tumors display a diversity of T cell receptor sequences that differ from the T cells in adjacent mucosal tissue. Cancer Immunol. Immunother. *62*, 1453–1461.

34. Zhang, Q., Jia, Q., Deng, T., Song, B., and Li, L. (2015). Heterogeneous expansion of CD4+ tumor-infiltrating T-lymphocytes in clear cell renal cell carcinomas. Biochem. Biophys. Res. Commun. *458*, 70–76.

35. Chen, Z., Zhang, C., Pan, Y., Xu, R., Xu, C., Chen, Z., Lu, Z., and Ke, Y. (2016). T cell receptor β-chain repertoire analysis reveals intratumour heterogeneity of tumour-infiltrating lymphocytes in oesophageal squamous cell carcinoma. J. Pathol. *239*, 450–458.

36. Kavianpur, M., Shahjahani, M., Jaseb, K., Kasar, S., and Saki, N. (2017). Prognostic value and clinical significance of TCR rearrangements for MRD monitoring in ALL patients. Comp. Clin. Pathol. *26*, 269–276.

37. Schumacher, T.N., and Schreiber, R.D. (2015). Neoantigens in cancer immunotherapy. Science *348*, 69–74.

38. Miyauchi, E., Matsuda, T., Kiyotani, K., Low, S.-K., Hsu, Y.-W., Tsukita, Y., Ichinose, M., Sakurada, A., Okada, Y., Saito, R., and Nakamura, Y. (2019). Significant differences in T cell receptor repertoires in lung adenocarcinomas with and without epidermal growth factor receptor mutations. Cancer Sci. *110*, 867–874.

39. Cheng, X., Yin, H., Fu, J., Chen, C., An, J., Guan, J., Duan, R., Li, H., and Shen, H. (2019). Aggregate analysis based on TCGA: *TTN* missense mutation correlates with favorable prognosis in lung squamous cell carcinoma. J. Cancer Res. Clin. Oncol. *145*, 1027–1035.

40. Marino, M., Maiuri, M.T., Di Sante, G., Scuderi, F., La Carpia, F., Trakas, N., Provenzano, C., Zisimopoulou, P., Ria, F., Tzartos, S.J., et al. (2014). T cell repertoire in DQ5-positive MuSK-positive myasthenia gravis patients. J. Autoimmun. *52*, 113–121.

41. Toivonen, R., Arstila, T.P., and Hänninen, A. (2015). Islet-associated T-cell receptor-β CDR sequence repertoire in prediabetic NOD mice reveals antigen-driven T-cell expansion and shared usage of VβJβ TCR chains. Mol. Immunol. *64*, 127–135.

42. Farrow, S.C., Hagel, J.M., Beaudoin, G.A., Burns, D.C., and Facchini, P.J. (2015). Stereochemical inversion of (S)-reticuline by a cytochrome P450 fusion in opium poppy. Nat. Chem. Biol *11*, 728–732.

43. Liu, X., Cui, Y., Zhang, Y., Liu, Z., Zhang, Q., Wu, W., Zheng, Z., Li, S., Zhang, Z., and Li, Y. (2019). A comprehensive study of immunology repertoires in both preoperative stage and postoperative stage in patients with colorectal cancer. Mol. Genet. Genomic Med. *7*, e504.

44. Bolotin, D.A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I.Z., Putintseva, E.V., and Chudakov, D.M. (2015). MiXCR: software for comprehensive adaptive immunity profiling. Nat. Methods *12*, 380–381.

45. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358.