

# The evolution of T-cell receptor repertoire in different stages of non-small cell lung cancer

**Ziqi Jia**

Peking Union Medical College Hospital

**Yadong Wang**

Peking Union Medical College Hospital

**Xiaoying Yang**

Peking Union Medical College Hospital

**Pancheng Wu**

Peking Union Medical College Hospital

**Yanyu Wang**

Peking Union Medical College Hospital

**Yang Song**

Peking Union Medical College Hospital

**Huihui Xu**

Peking Union Medical College Hospital

**Dejian Gu**

Geneplus-Beijing Ltd

**Rongrong Chen**

Geneplus-Beijing Ltd.

**Jin Li**

Geneplus-Beijing Ltd

**Xuefeng Xia**

Geneplus-Beijing Ltd

**Zhongxing Bing**

Peking Union Medical College Hospital

**Lei Cao**

Peking Union Medical College Hospital

**Zhili Cao**

Peking Union Medical College Hospital

**Peng Liu**

Peking Union Medical College Hospital

**Huaxia Yang**

Peking Union Medical College Hospital

**Shanqing Li**

Peking Union Medical College Hospital

**Naixin Liang** (✉ [PUMCH\\_GPLiangNaix@163.com](mailto:PUMCH_GPLiangNaix@163.com))

Chinese Academy of Medical Sciences and Peking Union Medical College

---

## Research

**Keywords:** T-cell receptor repertoire, tumor microenvironment, NSCLC, EGFR

**Posted Date:** May 7th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-26421/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background** The intricate relationship between the tumor and host was not well understood, and antigen-specific T cell is fundamental in understanding the interaction. TCR repertoire analysis which described TCR clonotypes and TCR numbers has shown that TCRs with high frequency was tumor-specific T cells, while others might be 'bystander' T cells within tumors. However, how these "expanded" tumor-specific T cells was selected during the tumor development was not clear.

**Methods** We retrospectively analyzed TCR sequencing and mutation sequencing results from 144 non-small cell lung cancer (NSCLC) patients.

**Results** A rich TCR repertoire comprising thousands of different TCR sequences was identified in all stages of NSCLC, with most TCR clonotypes presented at low frequency. Interestingly, Stage IV NSCLC tumors contain more expanded TCRs as compared to earlier stages, however, lymph node metastasis or tumor size had little impact on expanded TCRs. Moreover, accumulation of mutations did not significantly change the number of TCR clonotypes, however, EGFR mutant patients had significantly lower while KRAS mutant patients had significantly higher number of TCR clonotypes especially in terms of those "expanded" TCRs.

**Conclusions** In summary, T cells in the tumor microenvironment were gradually activated with tumor development. Critical events such as distal metastases and generation of EGFR or KRAS mutations might be the major factors affecting the changing of tumor-specific T cells in the tumor microenvironment.

## Background

Tumor is characterized by the accumulation of mutations, the relationship between tumor and the evolving tumor microenvironment is still not well-understood (1). As a major component of adaptive immunity, T cells are recruited and resided in the tumor microenvironment and interacted with tumor cells during tumor development and in the cancer treatments (2–4). Because the T cell receptor (TCR) confers unique antigen specificity (5, 6), and the diversity and specificity of TCR are determined by the highly variable complementarity determining region 3 (CDR3) (7), we use TCR V $\beta$  CDR3 sequencing (TCRseq) to track TCR clonotypes of non-small cell lung cancer (NSCLC) patients at different stage to explore evolution of TCR repertoire and potential factors associated with the remodeling process.

TCR clonotypes and TCR numbers, which evaluated the Richness of T cell clonotypes and the density of T cell expansion, were widely used to characterize the anti-tumor changes in the tumor microenvironment (8–10). As reported (10, 11), most T cells in the tumor microenvironment were "bystander" T cells, such as memory T cells; while tumor-specific T cells were characterized by the TCRs within the top 1% of clonotypes. These expanded tumor-specific TCRs were always with the TCR frequency higher than 2/1,000 (0.002).

It was reported that tumor patients showed a significantly lower TCR clonotypes, but similar TCR numbers comparing with healthy people (8, 10). This prompted us to study the changes rules of TCR repertoire during tumor development. Tumor, node, metastasis (TNM) staging system is an internationally accepted system used to determine the disease stage. It is based upon an evaluation of the primary tumor, the regional lymph nodes and lymphatic drainage, and the presence or absence of distant metastases to determine prognosis and guide management. We employed the TNM system to show the difference of TCR in different stage, and also investigated the different effects of T, N and M on TCR evolution. We also analyzed the correlation between TCR numbers/clonotypes and genetic mutations to observe the influence of tumor development to TCR repertoire in this study.

## Methods

### Subject cohorts

We collected TCR sequencing results from 144 lung cancer patients, including 95 stage I patients, 14 stage II patients, 15 stage III patients and 20 stage IV patients (Additional file 1: Table S1). All the patients were treatment naïve and received previous anti-cancer therapy at the Cancer Center, Peking Union Medical College Hospital (Beijing, China). Clinical information was collected from the hospital's case management system and confirmed with relevant doctors.

### NGS base somatic mutation detection

All patients have been tested for EGFR/KRAS mutation status, and 40 of them had received 1021-gene panel sequencing by next-generating sequencing (NGS). Genetic analysis was conducted as previously described (12, 13) (Additional file 2: Table S2). Briefly, serial sections from formalin-fixed paraffin-embedded tumor tissues were used for genomic tumor DNA extraction using the QIAamp DNA mini kit (Qiagen, Valencia, CA). DNA from leukocytes was extracted using the DNeasy Blood Kit (Qiagen, Valencia, CA). Sequencing libraries were prepared from ctDNA using KAPA DNA Library Preparation Kits (Kapa Biosystems, Wilmington, MA, USA), and genomic DNA sequencing libraries were prepared with Illumina TruSeq DNA Library Preparation Kits (Illumina, San Diego, CA). Libraries were hybridized to custom-designed biotinylated oligonucleotide probes (Roche NimbleGen, Madison, WI, USA) targeting 1021 genes. Prepared libraries were sequenced on a NextSeq CN 500 (Illumina, San Diego, CA).

Sequencing data was analyzed using default parameters. Adaptor sequences and low-quality reads were removed. The clean reads were aligned to the reference human genome (hg19) with Burrows-Wheeler Aligner (BWA; version 0.7.12-r1039). Variants were called with GATK (version 3.4-46-gbc02625) and MuTect (version 1.1.4). Contra (v2.0.8) was used to detect copy number variants, and NCsv (in house) was used to detect structural variants. The final candidate variants were all manually verified using Integrative Genomics Viewer. Variants were filtered to exclude synonymous variants, known germline variants in dbSNP, and variants that occur at a population frequency of > 1% in the Exome Sequencing Project.

# High-throughput sequencing of T-cell receptor $\beta$ genes

TCR repertoire was sequenced using surgical tissues of patients without distant metastasis and needle biopsy tissues of patients with advanced lung cancer. TCR sequencing and TCR repertoire quantification was performed as previously research (8, 14, 15). CDR3 region of TCR  $\beta$  chain was amplified by multiplex PCR including PCR1 and PCR2. For as much amplification as possible V(D)J combinations, a set of 32 V forward and 13 J reverse primers was adopted to perform multiplex PCR1 and assays. Universal primers were performed in PCR2. TCR $\beta$  CDR3 region was sequenced using the Illumina Novaseq System, and reads of length 151-bp were obtained. Then, the CDR3 sequences were identified and assigned by the MiXCR software package (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) (16). Sequences were collapsed and filtered in order to identify and quantitate the absolute abundance of each unique TCR $\beta$  CDR3 region for further analysis, as previously described(15, 17, 18)

## Statistical analysis

Mann-Whitney test was used to compare differences between groups. Correlations between variables were analyzed using Spearman's rank test. All statistical analyses were performed using GraphPad Prism 8.0. In this study, all tests were two-sided and p values < 0.05 were considered statistically significant.

## Results

### 1. Elderly patients had more expanded TCRs

A rich TCR repertoire comprising thousands of different TCR sequences was identified in all stages of NSCLC. To investigate the correlation between clinical characteristics and TCR repertoire, we first performed different thresholds of TCR analysis to assess the association between TCR repertoire and age, gender and smoking. We found TCR numbers showed a positive correlation with ages with TCR frequency cutoff value at 0.002, which is corresponding to TCRs within the top 1% TCRs as reported (Fig 1A)(10). Though we had more patients in stage I than other stages, the age of different stages were comparable ( $p=0.17$ , Fig 1B). We also found neither gender nor smoking status affect TCR clonotypes or numbers (19, 20) (Supplementary Fig S1, Fig S2).

### 2. Stage IV NSCLC tumors contain more expanded TCRs as compared to earlier stages

Most TCR clonotypes presented at low frequency which may be tumor irrelevant bystander T cells as reported (10). When focusing on the TCR sequences at high frequency within the tumor (threshold of 0.002), stage IV samples showed a significantly higher TCR numbers than stage I, II and III samples (Fig 2A). There was no difference in TCR clonotypes among different stages (Fig 2B). As all the stage IV patients had distal metastasis, we speculated that distal metastasis might associate with the TCR clonal

expansion (recruitment and proliferation). To test our hypothesis, we further studied the correlation between lymph node metastasis (N) or tumor size (T) with TCR.

### 3. Lymph node metastasis or tumor size had little impact on expanded TCRs

To study the correlation between lymph node metastasis (N) or tumor size (T) with TCR, we excluded patients with distal metastases and only analyzed stage II/III patients. These patients were divided into lymph node metastasis group (LM) and non-lymph node metastasis group (non-LM). The number of expanded TCRs accounted for almost 20% of the total TCRs in each group, and no difference was found in TCR numbers among all frequency (Fig 3A). LM group showed decreased TCR clonotypes only at a threshold of 0.002 and 0.003 (Fig 3B).

The effect of tumor size on TCR repertoire was analyzed in stage I patients. The patients were divided into three groups according to subgroup division of T1 (T: diameter of tumor;  $T \leq 1\text{cm}$ ;  $T \leq 2\text{cm}$ ;  $T > 2\text{cm}$ ). Similar to lymph node metastasis, no difference in TCR numbers was noticed and difference in TCR clonotypes was found in the some low frequency group which were threshold of 0.0005, 0.001, and 0.002, and (Fig 3C and Fig 3D). These results indicated that lymph node metastasis or tumor size may influence the TCR clonotypes with low frequency, but had little impact on expanded TCRs.

### 4. EGFR but not accumulation of mutations was associated with expanded TCRs

It was reported that numbers of expanded TCRs was correlated with the number of nonsynonymous mutations in early stage NSCLC tumors (10). We therefore tested whether the increased number of expanded TCRs was due to accumulation of nonsynonymous mutation in advanced NSCLC. Out of our surprise, neither TCR numbers nor TCR clonotypes changed significantly following the accumulation of non-synonymous mutations at most thresholds (Fig 4A). We then focused on the driver mutations of NSCLC. Epidermal growth factor receptor (*EGFR*) and Kirsten rat sarcoma viral oncogene homolog (*KRAS*) were the most common mutated oncogenes in NSCLC (21, 22). We found there were no difference between the *EGFR* mutant and *EGFR* wildtype group in TCR numbers, but the *EGFR* wildtype group showed significant more TCR clonotypes of expanded TCRs (Fig 4B, 4C). This indicated that *EGFR* may play more significant roles in TCR clonal expansion compared with accumulation of other mutations. Given that *KRAS* also was a driver mutation of NSCLC, we further analyzed the relationship between *KRAS* mutation and expanded TCRs. As depicted in Supplementary Fig. S3, *KRAS*-mutant patients showed more clonotypes and numbers in highly expanded TCRs than non-*KRAS* patients. These results indicated that driver mutations might play crucial roles in the evolution of TCR repertoire.

## Discussion

In this study, we used two direct and effective indexes of TCR repertoire: TCR clonotypes and TCR numbers to study the evolution of TCR repertoire during tumor development. To the best of our knowledge, it was the first comprehensive analysis of TCR repertoire within non-small cell lung cancer patients at different stages of disease progression (Supplementary Fig. S4). And it went beyond the reported difference of TCR repertoire between tumor patients and healthy peoples (8, 10).

Our study demonstrated that TCR numbers of expanded TCRs was positive correlated with age as others (23, 24). Immunosenescence is characterized by a progressive and global remodeling of the immune functions during aging and it involves both innate and adaptive immunity. Senescence leads to an extensive decrease of naïve CD8 + T cells(25). A reduction of apoptotic molecules such as CD95, suggest that T cells were less prone to spontaneous apoptosis and may accumulate in elderly individuals (26). Moreover, senescent T cells also secrete high levels of inflammatory cytokines(27, 28) and previously has described a positive correlation between age, number of CD8 + CD28- T cells and cytotoxicity capacity of freshly isolated T lymphocytes, suggesting that an activation within the cytotoxic arm of the immune system may occur during aging (29). All this results revealed that senescence leads to change of T cells clonotypes, but not means drastic decrease in density of T cells. This was consistent with recent studies showing that elderly patients can benefit from immunotherapy (30).

Our study showed that lymph node metastasis or tumor size had little impact on expanded TCRs. It was consistent with previous studies which have revealed that tumor-infiltrating lymphocytes (TILs) in resected NSCLC were not significantly different among I-III stages (31, 32). Remarkably, numbers of vast majority of expanded TCRs were increased in stage IV samples compared with earlier stages. Increased numbers of expanded TCRs may due to more diversity (increased clonotypes) or enhanced proliferation or recruitment of T cells of each clonotypes or both. Though we could not delineate the exact mechanism in our study, the significant difference of TCR clonotypes between Stage IV and earlier stages indicated that predictive markers of immunoadjuvant therapy and immunotherapy for advanced patients should be different.

As a parallel, we also analyzed whether accumulation of non-synonymous mutation could induce TCR expansion through increased neoantigen production(33). The accumulation of mutation will lead to increase in intratumoral heterogeneity, which will affect the tumor microenvironment (9, 13). The spatial heterogeneity of TCRs has been reported and numbers of expanded ubiquitous TCRs (present in all regions of tumor) correlated with the number of ubiquitous non-synonymous mutations (present in all regions of tumor), but not with the number of regional non-synonymous mutations (10). In our study, numerical but not significant correlation between TCR numbers or clonotypes and number of non-synonymous mutations was found probably due to failure to distinguish ubiquitous and regional mutations(9, 10). Interestingly, we found *EGFR* mutant patients showed a significantly lower TCR clonotypes in expanded TCRs. On the contrary, *KRAS* mutant patients had higher TCR clonotypes and numbers in expanded TCRs. Several preclinical and clinical findings suggest that the tumor microenvironment of NSCLC patients with *EGFR* mutations may have less CD8 + TILs, but more regulatory T cells (Tregs) accumulated in the tumor environment (34, 35). A recent study showed that

*KRAS* mutant NSCLC was positive related with tumor immunity-associated features, including PD-L1 expression, CD8 + TILs and tumor mutational burden (22). In line with these, EGFR/*KRAS* which were known as common ubiquitous driver mutations in NSCLC, played more significant roles than other passage mutations in the evolution of TCR repertoire.

Our study has several limitations. First, as a retrospective study, we were unable to connect our findings with clinical outcomes. Secondly, in spite of the large cohort of patients, there were more stage I patients than all other stages. Finally, due to the limited availability of tissue samples, we were unable to perform other study such as T cells staining to further delineate the mechanism.

## Conclusions

Despite the gradual accumulation of mutation, and continuous development of NSCLC, critical events such as distal metastases and generation of EGFR or *KRAS* mutations might be the major factors affecting the changing of tumor-specific T cells in the tumor microenvironment. The significant difference of TCRs between advanced and early stage warranted further study considering that cancer immunotherapy is more and more widely used in NSCLC patients of earlier stage.

## Abbreviations

TCR, T cell receptor

NGS, Next generation sequencing

CDR3, complementarity determining region 3

NSCLC, non-small cell lung cancer

## Declarations

Ethics approval and consent to participate

We have obtained written informed consent from the participants before genetic testing for further scientific analysis. This study was approved by the Ethic Committee of Peking Union Medical College Hospital.

Consent for publication

Written informed consent was obtained from the patient. The manuscript does not contain any personally identifiable information or images.

Availability of data and materials

The datasets used and analyzed during this study are not publicly available to protect patient privacy, which will be available from the corresponding author on reasonable request.

Conflict of interest:

Dejian Gu, Rongrong Chen, Jin Li, and Xuefeng Xia are employees of Geneplus-Beijing Ltd. No potential conflicts of interest were disclosed of other authors.

Funding

This study was supported by Chinese Academy of Medical Sciences Young Medical Talent Award Fund (no.2018RC320005), Beijing Natural Science Foundation (7182132), Beijing Students' platform for innovation and entrepreneurship training program(2019zlgc0629), and The National Key Research and Development Program of China (No. 2016YFC0901500).

Acknowledgements

Not applicable

Authors' contributions

NL, SL, HY, and ZJ were involved in the conception and design of the study. ZJ and XY were major contributors in developing the manuscript. ZJ, XY, YW, PW, and YW collected and analyzed the clinical information. DG, RC, and XX were in charge of the statistical analysis of data of TCR repertoire. All authors read and approved the final manuscript.

## References

1. Horton BL, Fessenden TB, Spranger S. Tissue Site and the Cancer Immunity Cycle. *Trends in cancer*. 2019;5(10):593–603.
2. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature*. 2014;515(7528):568–71.
3. Twyman-Saint Victor C, Rech AJ, Maity A, Rengan R, Pauken KE, Stelekati E, et al. Radiation and dual checkpoint blockade activate non-redundant immune mechanisms in cancer. *Nature*. 2015;520(7547):373–7.
4. Pfirschke C, Engblom C, Rickelt S, Cortez-Retamozo V, Garris C, Pucci F, et al. Immunogenic Chemotherapy Sensitizes Tumors to Checkpoint Blockade Therapy. *Immunity*. 2016;44(2):343–54.
5. Chen Y, Xu Y, Zhao M, Liu Y, Gong M, Xie C, et al. High-throughput T cell receptor sequencing reveals distinct repertoires between tumor and adjacent non-tumor tissues in HBV-associated HCC. *Oncoimmunology*. 2016;5(10):e1219010.
6. Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and coreceptors. *Annu Rev Immunol*. 2006;24:419–66.

7. Rubtsova K, Scott-Browne JP, Crawford F, Dai S, Marrack P, Kappler JW. Many different Vbeta CDR3s can reveal the inherent MHC reactivity of germline-encoded TCR V regions. *Proc Natl Acad Sci USA*. 2009;106(19):7951–6.
8. Liu YY, Yang QF, Yang JS, Cao RB, Liang JY, Liu YT, et al. Characteristics and prognostic significance of profiling the peripheral blood T-cell receptor repertoire in patients with advanced lung cancer. *International journal of cancer*. 2019.
9. Jia Q, Wu W, Wang Y, Alexander PB, Sun C, Gong Z, et al. Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung cancer. *Nature communications*. 2018;9(1):5361.
10. Joshi K, Robert de Massy M. Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer. *Nature medicine*. 2019.
11. Simoni Y, Becht E, Fehlings M, Loh CY, Koo SL, Teng KWW, et al. Bystander CD8(+) T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature*. 2018;557(7706):575–9.
12. Nong J, Gong Y, Guan Y, Yi X, Yi Y, Chang L, et al. Circulating tumor DNA analysis depicts subclonal architecture and genomic evolution of small cell lung cancer. *Nature communications*. 2018;9(1):3114.
13. Zhang Y, Chang L, Yang Y, Fang W, Guan Y, Wu A, et al. The correlations of tumor mutational burden among single-region tissue, multi-region tissues and blood in non-small cell lung cancer. *J immunother Cancer*. 2019;7(1):98.
14. Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood*. 2009;114(19):4099–107.
15. Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung MW, Parsons JM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. *Nature communications*. 2013;4:2680.
16. Bolotin D, Poslavsky S, Mitrophanov I, Shugay M, Mamedov I, Putintseva E, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods*. 2015;12(5):380–1.
17. Kirsch I, Vignali M, Robins H. T-cell receptor profiling in cancer. *Molecular oncology*. 2015;9(10):2063–70.
18. Reuben A, Gittelman R, Gao J, Zhang J, Yusko EC, Wu CJ, et al. TCR Repertoire Intratumor Heterogeneity in Localized Lung Adenocarcinomas: An Association with Predicted Neoantigen Heterogeneity and Postsurgical Recurrence. *Cancer discovery*. 2017;7(10):1088–97.
19. Wang GZ, Zhang L, Zhao XC, Gao SH, Qu LW, Yu H, et al. The Aryl hydrocarbon receptor mediates tobacco-induced PD-L1 expression and is associated with response to immunotherapy. *Nature communications*. 2019;10(1):1125.
20. Calles A, Liao X, Sholl LM, Rodig SJ, Freeman GJ, Butaney M, et al. Expression of PD-1 and Its Ligands, PD-L1 and PD-L2, in Smokers and Never Smokers with KRAS-Mutant Lung Cancer. *Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer*. 2015;10(12):1726–35.

21. Li Y, Li X, Li H. Genomic characterisation of pulmonary subsolid nodules: mutational landscape and radiological features. *The European respiratory journal*. 2019.
22. Liu C, Zheng S, Jin R, Wang X, Wang F, Zang R, et al. The superior efficacy of anti-PD-1/PD-L1 immunotherapy in KRAS-mutant non-small cell lung cancer that correlates with an inflammatory phenotype and increased immunogenicity. *Cancer letters*. 2020;470:95–105.
23. Naylor K, Li G, Vallejo AN, Lee WW, Koetz K, Bryl E, et al. The influence of age on T cell generation and TCR diversity. *Journal of immunology (Baltimore, Md: 1950)*. 2005;174(11):7446-52.
24. Salam N, Rane S, Das R, Faulkner M, Gund R, Kandpal U, et al. T cell ageing: effects of age on development, survival & function. *Indian J Med Res*. 2013;138(5):595–608.
25. Adibzadeh M, Mariani E, Bartoloni C, Beckman I, Ligthart G, Remarque E, et al. Lifespans of T lymphocytes. *Mech Ageing Dev*. 1996;91(2):145–54.
26. Posnett DN, Edinger JW, Manavalan JS, Irwin C, Marodon G. Differentiation of human CD8 T cells: implications for in vivo persistence of CD8 + CD28- cytotoxic effector clones. *Int Immunol*. 1999;11(2):229–41.
27. Akbar AN, Henson SM, Lanna A. Senescence of T Lymphocytes: Implications for Enhancing Human Immunity. *Trends Immunol*. 2016;37(12):866–76.
28. Franceschi C, Bonafe M, Valensin S, Olivieri F, De Luca M, Ottaviani E, et al. Inflamm-aging. An evolutionary perspective on immunosenescence. *Ann N Y Acad Sci*. 2000;908:244–54.
29. Franceschi C, Valensin S, Fagnoni F, Barbi C, Bonafe M. Biomarkers of immunosenescence within an evolutionary perspective: the challenge of heterogeneity and the role of antigenic load. *Experimental gerontology*. 1999;34(8):911–21.
30. Ferrara R, Mezquita L, Auclin E, Chaput N, Besse B. Immunosenescence and immunecheckpoint inhibitors in non-small cell lung cancer patients: Does age really matter? *Cancer treatment reviews*. 2017;60:60–8.
31. Donnem T, Hald SM, Paulsen EE, Richardsen E, Al-Saad S, Kilvaer TK, et al. Stromal CD8 + T-cell Density-A Promising Supplement to TNM Staging in Non-Small Cell Lung Cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2015;21(11):2635–43.
32. Donnem T, Kilvaer TK, Andersen S, Richardsen E, Paulsen EE, Hald SM, et al. Strategies for clinical implementation of TNM-Immunescore in resected nonsmall-cell lung cancer. *Annals of oncology: official journal of the European Society for Medical Oncology*. 2016;27(2):225–32.
33. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science*. 2015;348(6230):69–74.
34. Dong ZY, Zhang JT, Liu SY, Su J, Zhang C, Xie Z, et al. EGFR mutation correlates with uninflamed phenotype and weak immunogenicity, causing impaired response to PD-1 blockade in non-small cell lung cancer. *Oncoimmunology*. 2017;6(11):e1356145.
35. Lin A, Wei T, Meng H, Luo P, Zhang J. Role of the dynamic tumor microenvironment in controversies regarding immune checkpoint inhibitors for the treatment of non-small cell lung cancer (NSCLC) with EGFR mutations. *Mol Cancer*. 2019;18(1):139.

## Additional Files

**Additional file 1:** Table S1. Clinical characteristic of 144 patients.

**Additional file 2:** Table S2. List of target regions of the pan-cancer 1021-gene panel.

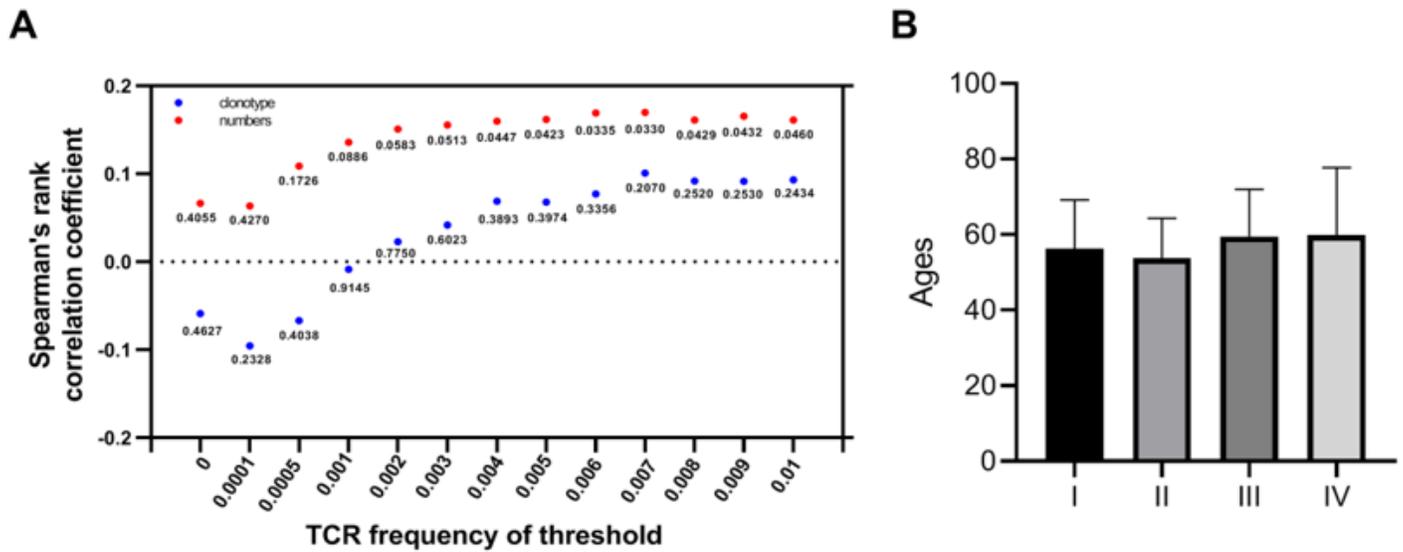
**Additional file 3:** Figure S1. The difference of TCR numbers (A) and clonotypes (B) in different gender at different frequencies (ranging from all TCRs (threshold of zero) up to those found at a frequency of  $\geq 0.01$ ). The minimum and maximum are indicated by the extreme points of the box plot; the median is indicated by the thick horizontal line; and the first and third quartiles are indicated by box edges. Statistical analysis is performed using the Mann-Whitney test and Spearman's rank test. \* $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\* $p < 0.001$ .

Figure S2. The difference of TCR numbers (A) and clonotypes (B) in smoking and never smoking patients at different frequencies (ranging from all TCRs (threshold of zero) up to those found at a frequency of  $\geq 0.01$ ). The minimum and maximum are indicated by the extreme points of the box plot; the median is indicated by the thick horizontal line; and the first and third quartiles are indicated by box edges. Statistical analysis is performed using the Mann-Whitney test and Spearman's rank test. \* $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\* $p < 0.001$ .

Figure S3. The impact of KRAS mutation to TCR numbers (A) and clonotypes (B) at different frequencies (ranging from all TCRs (threshold of zero) up to those found at a frequency of  $\geq 0.01$ ). The minimum and maximum are indicated by the extreme points of the box plot; the median is indicated by the thick horizontal line; and the first and third quartiles are indicated by box edges. Statistical analysis is performed using the Mann-Whitney test and Spearman's rank test. \* $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\* $p < 0.001$ .

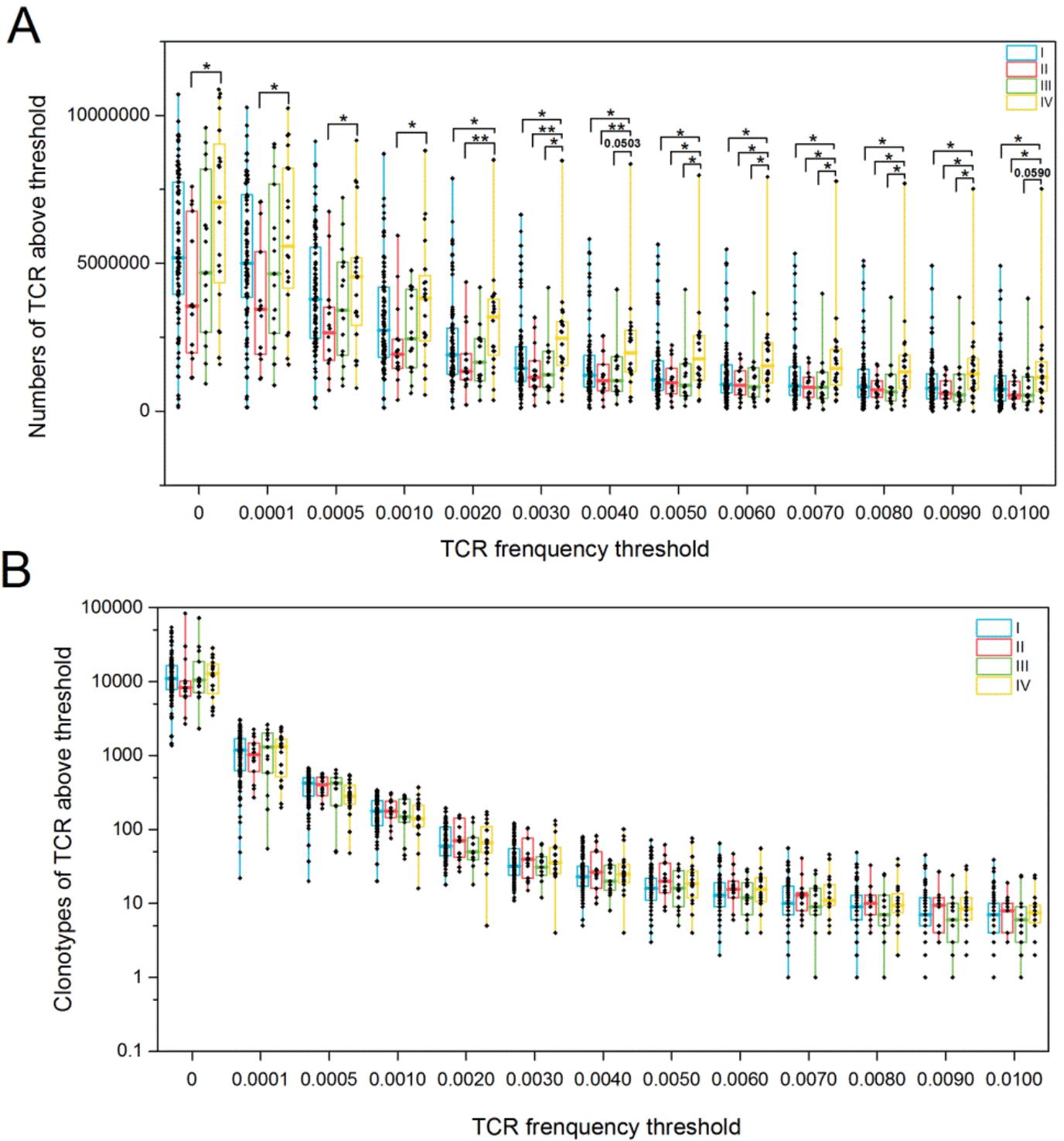
Figure S4. **A schematic diagram of our study.** The bar chart of TCR numbers was shown at different frequencies and the line in bars represent TCR clonotypes; blue bars represent low frequency TCRs, red bars represent expanded TCRs. The colored boxes indicated different frequencies TCRs associated with tumor development. The bars of TCR numbers were bisected by blue and beige backgrounds to represent the non-EGFR group and EGFR group, suggesting that similar TCR numbers and markedly different TCR clonotypes were found in two groups.

## Figures



**Figure 1**

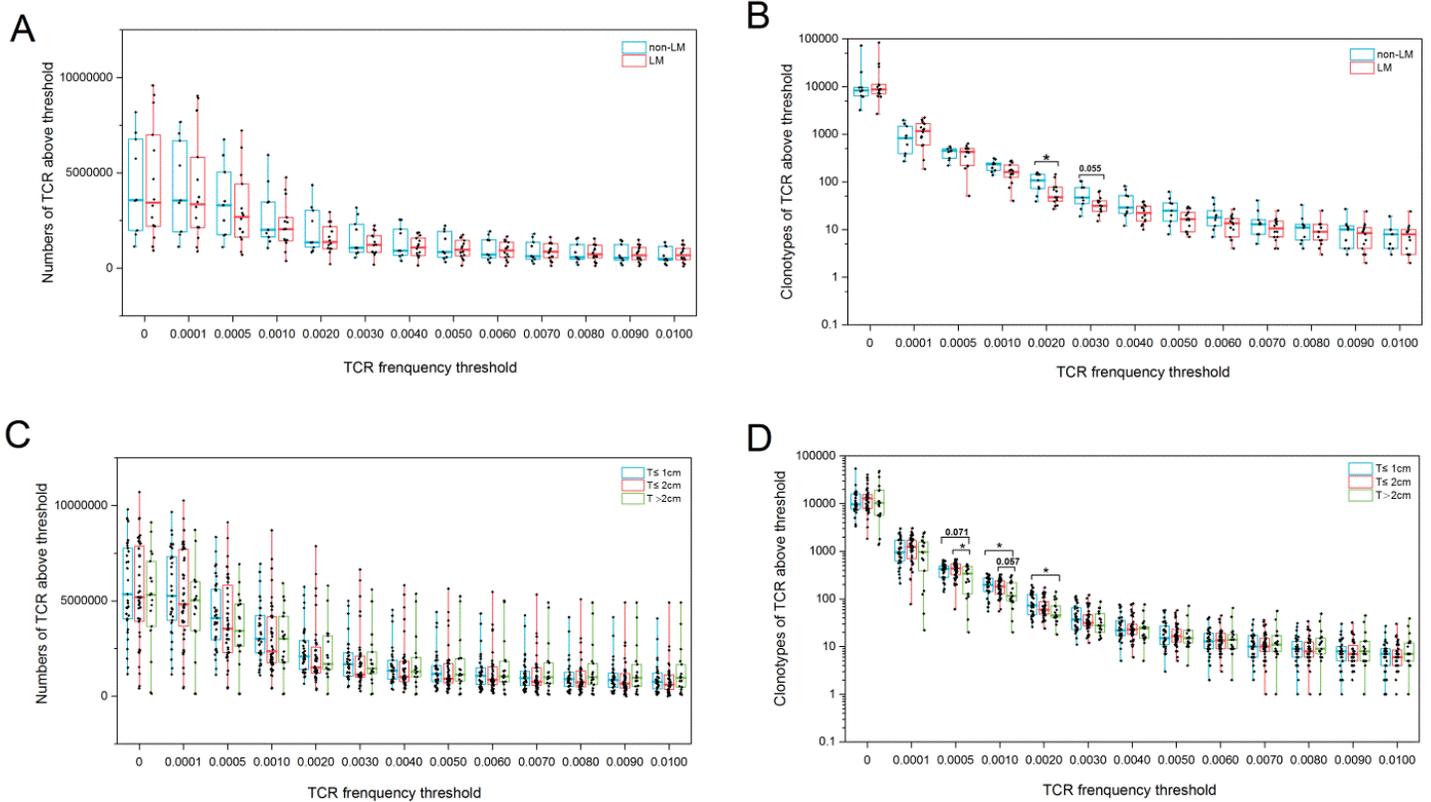
A. The Spearman's rank correlation coefficient and p value (shown bottom each point of red and blue) for the relationship between the age and TCR clonotypes (blue point) or TCR numbers (red point) at different frequencies (ranging from all TCRs (threshold of zero) up to those found at a frequency of  $\geq 0.01$ ). B. Distribution of age in all stages. The minimum and maximum are indicated by the extreme points of the box plot; the median is indicated by the thick horizontal line; and the first and third quartiles are indicated by box edges.



**Figure 2**

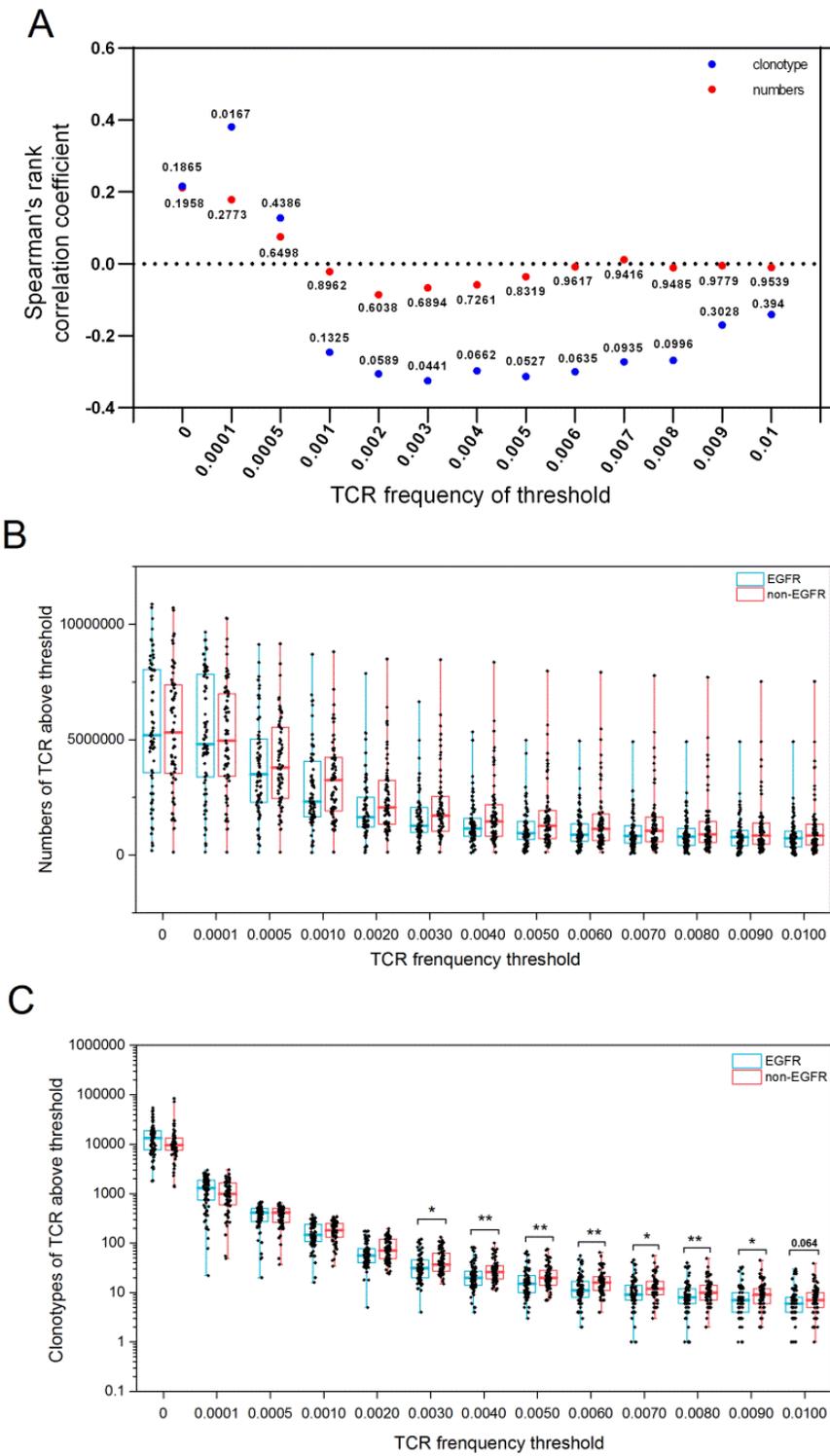
The correlation between TCR abundance and pathological stage. The numbers (A) and clonotypes (B) of TCR sequences detected above a given frequency threshold are shown for stage I, stage II, stage III, and stage IV. The illustration of figure B is an enlargement version of the red box. The minimum and maximum are indicated by the extreme points of the box plot; the median is indicated by the thick horizontal line;

and the first and third quartiles are indicated by box edges. Statistical analysis is performed using the Mann-Whitney test. \* $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\* $p < 0.001$ .



**Figure 3**

The changes of TCR abundance with lymphatic metastasis and tumor size enlargement. The difference of TCR numbers (A) and clonotypes (B) in lymphatic metastasis group (LM) and non-lymphatic metastasis group (non-LM) at different frequencies (ranging from all TCRs (threshold of zero) up to those found at a frequency of  $\geq 0.01$ ). The difference of TCR numbers (C) and clonotypes (D) and in different tumor size groups at different frequencies. The patients were divided into three groups (T: diameter of tumor;  $T \leq 1\text{cm}$ ;  $T \leq 2\text{cm}$ ;  $T > 2\text{cm}$ ) according to the size of the tumor diameter. Top of figure B and C are the enlargement version of the red box of bottom. The minimum and maximum are indicated by the extreme points of the box plot; the median is indicated by the thick horizontal line; and the first and third quartiles are indicated by box edges. Statistical analysis is performed using the Mann-Whitney test. \* $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\* $p < 0.001$ .



**Figure 4**

The influence of mutation accumulation and EGFR-driver mutation to TCR abundance. A. The Spearman's rank correlation coefficient and p value (shown above each point of blue and shown bottom each point of red) for the relationship between the number of nonsynonymous mutations and TCR clonotypes (blue point) or TCR numbers (red point) at different frequencies (ranging from all TCRs (threshold of zero) up to those found at a frequency of  $\geq 0.01$ ). The impact of EGFR-driver mutation to TCR numbers (B) and

clonotypes (C) at different frequencies. Right of figure C is a enlargement version of the red box of left. The minimum and maximum are indicated by the extreme points of the box plot; the median is indicated by the thick horizontal line; and the first and third quartiles are indicated by box edges. Statistical analysis is performed using the Mann-Whitney test and Spearman's rank test. \*p < 0.05; \*\* p <0.01;\*\*\*p <0.001.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarymaterialTableS2.xlsx](#)
- [supplementarymaterialFigure.docx](#)
- [supplementarymaterialTableS1.docx](#)
- [supplementarymaterialTableS2.xlsx](#)
- [supplementarymaterialTableS1.docx](#)
- [supplementarymaterialFigure.docx](#)