



## LETTER OPEN

# Characteristics and significance of peripheral blood T-cell receptor repertoire features in patients with indeterminate lung nodules

Signal Transduction and Targeted Therapy (2022)7:348

; <https://doi.org/10.1038/s41392-022-01169-7>**Dear Editor,**

Lung cancer is one of the most common cancers and a leading cause of cancer-related deaths. T cells are known to play a significant role in the destruction of cancer cells. T cells have therefore become the focus of lung cancer immunotherapy. T-cell receptors (TCRs) can recognize antigenic peptides presented by HLA proteins. TCRs are distinct individually and vary with pathophysiological condition, so T cells can respond to a wide range of antigens. TCR repertoire diversity reflects the potential for cellular immunity, and several studies have demonstrated that complementarity determining region 3 (CDR3 $\beta$ ) diversity is important in cancer therapy and prognosis.<sup>1</sup> Liu et al. have reported that the CDR3 $\beta$  diversity of patients with advanced lung cancer differs significantly from that of healthy individuals.<sup>2</sup>

Here, we have developed a novel model called “TCRnodseek” based on the repertoire properties of TCRs in peripheral blood, which can accurately classify small pulmonary nodules into malignant or benign types. The prospective study was initiated at Sichuan Cancer Hospital (Supplementary Materials). The flow chart for this study is shown in Fig. 1a (Supplementary Fig. S1). In Supplementary Table 1, we describe the main baseline characteristics of the 109 patients. Among them, 99 patients with indeterminate lung nodules are the main research subjects. The number of male and female subjects is comparable (male 52; 52.5%), the mean age is 55.5 years, and the mean size of indeterminate lung nodules is 13.7 mm. There are two independent groups of patients enrolled in this study (a discovery group and a validation group). We extracted DNA from their peripheral blood samples and analyzed their TCR repertoires. Supplementary Table 2 provides detailed information on the quality control of sequencing.

Utilizing Venn diagrams, we identified malignant associated clones (46% of all clones) that only existed in the stage I group (Fig. 1b). We defined the top 30 CDR3 amino acid sequences and the top 3000 sequences that are present in at least two subjects are enriched CDR3 $\beta$  amino acid sequences (aaSeqs, Supplementary Fig. S4). In enriched CDR3 $\beta$  aaSeqs, one common motif was observed, ‘SSGGSSYEQYF’, which is similar to some previously reported high-quality specificity CDR3 $\beta$  aaSeq motif in non-small cell lung cancer<sup>3</sup> (Supplementary Data S1, Data S2). In addition, some CDR3 $\beta$  aaSeqs were annotated by VDJdb, which matched multiple HLA types and antigens. Then, we applied pMTnet to rank various pairs and obtained three candidate CDR3 $\beta$  aaSeqs (Supplementary Data S1).<sup>4</sup>

The correlation between clonal fraction and clinical characteristics was studied. As shown in Fig. 1c, packed circles indicate that some stage I subjects harbor high fraction clones. We classified TCR clones into five types: hyperexpanded, large, medium, small, and rare clone types (Fig. 1d). The benign group harbored significantly more small-type clones, whereas the malignant group

harbored significantly more hyperexpanded-type clones ( $p = 0.0082$  and  $p = 0.011$ , respectively, Fig. 1e). The relationship between TCR features and clinical information was assessed using the spearman correlation. The correlation plot reveals that nodule size (largest diameter) is positively correlated with CloneReads ( $p = 0.01$ ,  $r = 0.24$ , Supplementary Fig. S3).

The Shannon index, evenness index, Simpson index, and clonality index were significantly different between the benign and stage I group ( $p = 0.01$ ,  $p = 0.0071$ ,  $p = 0.01$ ,  $p = 0.01$ , respectively; Fig. 1f, Supplementary Data S3). ROC analysis was used to determine the diagnostic value of the above features, and the AUC of each feature was higher than that of previous nodule diagnosis methods (Fig. 1g). In addition, we have added several important clinical features for the diagnosis of lung nodules to the list of candidate features. Random forest and information gain methods were applied to select the top three vital features (Ground glass nodule, Shannon index, and evenness index) to distinguish benign from malignant lung nodules (Fig. 1i). GGN (Ground glass nodule) is the most important feature. The Shannon index can be used as a useful complementary indicator for non-GGN indeterminate lung nodules ( $p = 0.01$ , Fig. 1h), which is consistent with the previous study.<sup>3</sup>

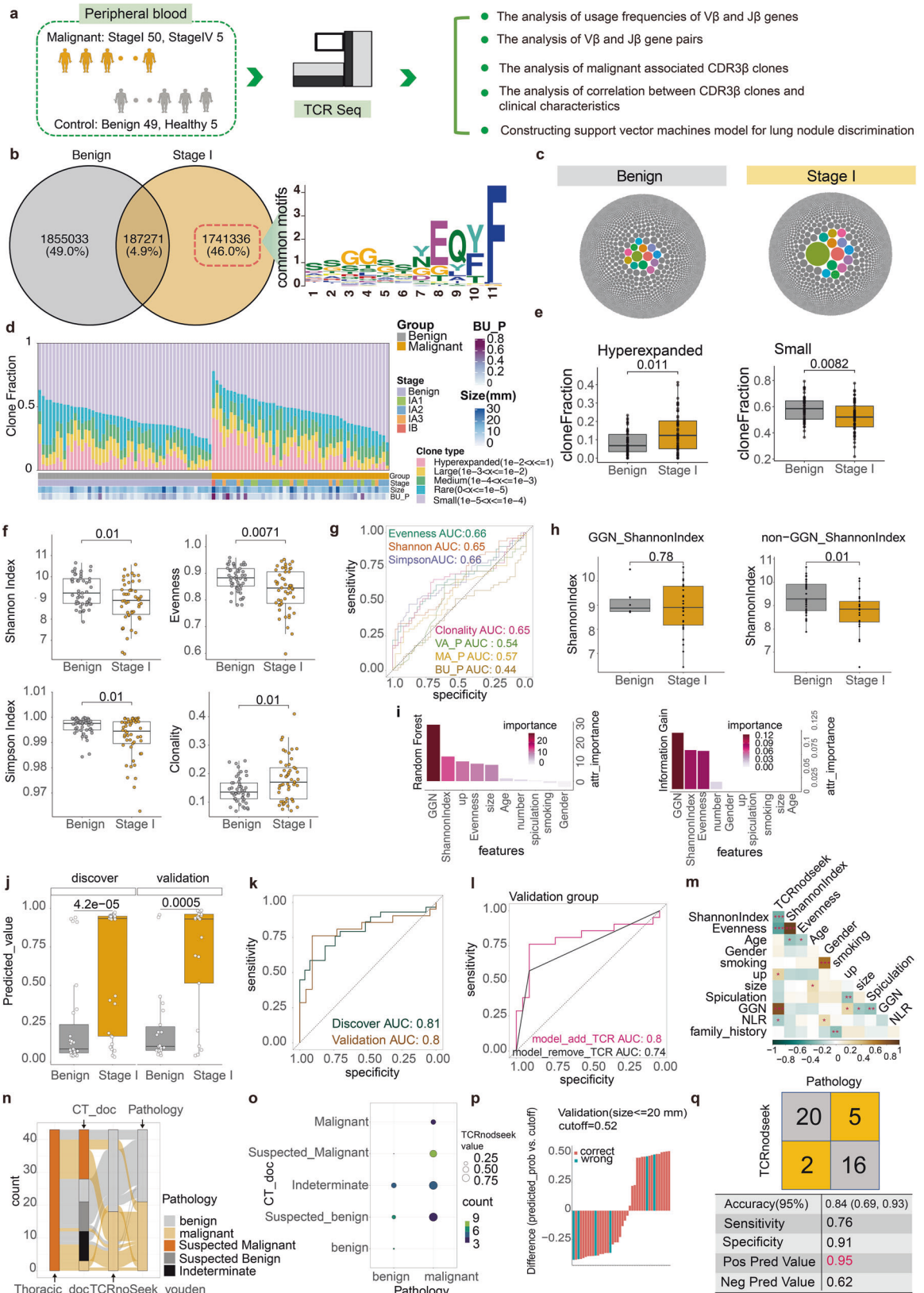
Next, a robust support vector machine model was constructed and its parameters were optimized with genetic algorithm (Supplementary Fig. S5). The model was thereafter called TCRnodseek, which was short for seeking to distinguish malignant nodules from benign ones via a TCR-based model. TCRnodseek performed well in both discovery and validation groups (AUC values of 0.81 and 0.80, respectively; Fig. 1j, k). In the validation group, the AUC value of a model without TCR features decreased to 0.74 (Fig. 1l). The correlation between TCRnodseek predicted value and clinical information was determined using spearman correlation (Fig. 1m).

Furthermore, our model was able to correctly identify not only suspected malignant lung nodules diagnosed by a thoracic surgeon, but also most indeterminate lung nodules determined by an advanced radiologist with more than 10 years of experience (Fig. 1n, o, Supplementary Fig. S4). In addition, TCRnodseek was highly accurate for indeterminate lung nodules less than 20 mm or 10 mm in size (Fig. 1p and Supplementary Fig. S6). TCRnodseek performs exceptionally well in the validation group, with a positive predictive value of 0.95 (Fig. 1q).

Our study found the highest TCR diversity in benign subjects, which is in agreement with results reported in a renal cell carcinoma study.<sup>5</sup> In addition, we have applied bladder cancer and bladder benign data to validate,<sup>6</sup> and we found that the TCR Shannon index of cancer is lower than that of benign subjects (Supplementary Fig. S7). In stage I subjects, we observed a decrease in TCR diversity, which may be related to intratumoral

Received: 3 January 2022 Revised: 21 July 2022 Accepted: 19 August 2022

Published online: 10 October 2022



T cells with clonal expansion that circulated to the peripheral bloodstream.<sup>7</sup> In contrast to stage II and stage III, stage I subjects possess a significantly higher diversity of TCRs<sup>8</sup> (Supplementary Fig. S7). In this way, TCRnodseek may be used easily to detect

malignant lung nodules with stage II and stage III as positive lung nodules.

Currently, regular low-dose CT (LDCT) screening is the most effective way to detect lung cancer early. LDCT screening,

**Fig. 1** The study design and main results. **a** In this prospective cohort study, patients suspected of having indeterminate pulmonary nodules were enrolled and their CT information was recorded. Blood samples were collected and stored prior to surgery. TCR sequencing was performed and analyzed. **b** Venn diagrams were used to identify clones just existing in the malignant group. common motif of enriched CDR3 $\beta$  aaSeqs is displayed. **c** The top 10,000 TCR clones are shown by Packed circles for the benign and stage I groups (one circle represents one clone; circle size represents clonal fraction). **d** The relationship between the clone types and the clinical characteristics of the benign and malignant groups is demonstrated. **e** A boxplot to illustrate the difference between benign and malignant clones for small-type and hyperexpanded-type clone types. **f** Comparison of TCR repertoire diversity between benign and malignant groups using boxplots. **g** Comparison of the diagnostic value of TCR repertoire diversity with previous models using ROC curves. **h** Boxplot showing the difference in Shannon index between benign and malignant groups in the GGN group and the non-GGN group separately. **i** Random forest and information gain methods were used to select significant features for the TCRnodseek model. **j** The predicted value calculated by TCRnodseek in the discovery and validation groups are presented as boxplots and ROC curves (**k**). **l** the ROCs of the models with and without TCR features. **m** the relationship between TCRnodseek predicted value and various clinical characteristics. **n** A Sankey diagram illustrating the misclassification of TCRnodseek and experts in comparison with pathology results (Thoracic\_doc, the results determined by Thoracic doctors; CT\_doc, the results determined by Radiology doctors; TCRnodseek\_youden, the results determined by TCRnodseek with youden cutoff). **o** A bubble plot illustrating the relationship between CT doctors' diagnoses, pathology results, and TCRnodseek predicted value (the size of the bubble represents the mean predicted value of TCRnodseek; the color of the bubble represents the number of subjects in the condition). **p** A waterfall plot showing the TCRnodseek function for nodule sizes  $\leq 20$  mm (the X-axis represents the sample id sorted by the predicted value of TCRnodseek; the Y-axis represents the predicted value of TCRnodseek minus the cutoff value). **q** The diagnostic performance of TCRnodseek in the validation group (the original subjects of (**b**) to (**i**) figures are all subjects; the original subjects of (**l**) to (**q**) figures are independent validation subjects; GGN ground glass nodule)

however, has a false positive rate of 96.4%.<sup>9</sup> Experienced physicians can exclude some benign nodules using image features and clinical characteristics. Nevertheless, 30–50% of patients undergoing surgery are still diagnosed as benign nodules pathologically in the end. In fact, regular screenings and unnecessary surgeries have resulted in unnecessary medical care and psychological stress. Liquid biopsy may help. However, the AUCs of published liquid biopsy methods are  $<0.85$  in validation groups.<sup>10</sup> TCRnodseek provides a promising solution to this issue, with AUC of 0.8. It depends on three features: GGN, Shannon index, and Evenness. The first feature can be obtained from LDCT, and the two left features can be computed from TCR repertoire data of a patient. To obtain TCR repertoire data, only about 1 ml of peripheral blood is needed, which can be very easily got in clinical practice.

A web server, named “Tool Box of Lung Nodule Predictors” (TB-LNPs) (<http://i.uestc.edu.cn/TB-LNPs>), has been developed to allow public access to the TCRnodseek model. Also, we collected several other canonical models, which could be used in academic studies to evaluate indeterminate lung nodules.

In conclusion, we have developed the TCRnodseek model, which integrates TCR diversity and clinical information to distinguish benign from malignant lung nodules more accurately. In addition to providing evidence for diagnosis, information on CDR3 $\beta$  might benefit the development of CAR T-cell Therapy.

#### DATA AVAILABILITY

Raw data of this project have been uploaded to <https://bigd.big.ac.cn/gsa> (HRA001754 and HRA002253).

#### ACKNOWLEDGEMENTS

This study was supported by grants from the Sichuan Medical Association Research project (S20087), Sichuan Cancer Hospital Outstanding Youth Science Fund (YB2021033), the National Natural Science Foundation of China (62071099), and supported by Medico-Engineering Cooperation Funds from University of Electronic Science and Technology of China (No. ZYGX2022YGRH004).

#### ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41392-022-01169-7>.

**Competing interests:** W.L. is one of the Associate Editors of Signal Transduction and Targeted Therapy, but he has not been involved in the process of manuscript handling. The authors declare no conflict of interest.

**Ethics approval:** The study was approved by the medical ethical committee of Sichuan Cancer Hospital (SCCHEC-02-2021-037).

Huaichao Luo<sup>1,2</sup>, Ruiling Zu<sup>2</sup>, Ziru Huang<sup>1</sup>, Yingqiang Li<sup>3</sup>,  
Yulin Liao<sup>2</sup>, Wenxin Luo<sup>4</sup>, Peng Zhou<sup>2</sup>, Dongsheng Wang<sup>2</sup>,  
Shifu Chen<sup>3</sup>✉, Weimin Li<sup>4</sup>✉ and Jian Huang<sup>1</sup>✉

<sup>1</sup>School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, Sichuan, China; <sup>2</sup>Department of Clinical Laboratory & Division of Radiology, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, Sichuan, China; <sup>3</sup>HaploX Biotechnology Co., Ltd, Shenzhen, China and <sup>4</sup>Department of Respiratory and Critical Care Medicine, West China Hospital, West China Medical School, Sichuan University, Chengdu 610041, China

These authors contributed equally: Huaichao Luo, Ruiling Zu, Ziru Huang, Yingqiang Li, Yulin Liao

Correspondence: Shifu Chen ([chen@haplox.com](mailto:chen@haplox.com)) or Weimin Li ([weimi003@scu.edu.cn](mailto:weimi003@scu.edu.cn)) or Jian Huang ([hj@uestc.edu.cn](mailto:hj@uestc.edu.cn))

#### REFERENCES

- Chen, K. et al. Multiomics analysis reveals distinct immunogenomic features of lung cancer with ground-glass opacity. *Am. J. Respir. Crit. Care Med.* **204**, 1180–1192 (2021).
- Liu, Y. Y. et al. 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 (2019).
- Chiou, S.-H. et al. Global analysis of shared T cell specificities in human non-small cell lung cancer enables HLA inference and antigen discovery. *Immunity*. **54**, 586–602 (2021).
- Lu, T. et al. Deep learning-based prediction of the T cell receptor–antigen binding specificity. *Nat. Mach. Intell.* **3**, 864–875 (2021).
- Guo, L. et al. Characteristics, dynamic changes, and prognostic significance of TCR repertoire profiling in patients with renal cell carcinoma. *J. Pathol.* **251**, 26–37 (2020).
- Sankin, A., Chand, D., Schoenberg, M. & Zang, X. Human urothelial bladder cancer generates a clonal immune response: the results of T-cell receptor sequencing. *Urol. Oncol.* **37**, 810.e811–810.e815 (2019).
- Joshi, K. et al. Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer. *Nat. Med.* **25**, 1549–1559 (2019).
- Reuben, A. et al. Comprehensive T cell repertoire characterization of non-small cell lung cancer. *Nat. Commun.* **11**, 603 (2020).
- de Koning, H. J. et al. Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force. *Ann. Intern. Med.* **160**, 311–320 (2014).
- Liang, W. et al. Accurate diagnosis of pulmonary nodules using a noninvasive DNA methylation test. *J. Clin. Invest.* **131**, e145973 (2021).



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022