

Peer Review File

Article Information: <http://dx.doi.org/10.21037/tlcr-21-96>

Reviewer #A

Comment 1: Line 92. Can this be rephrased to communicate the idea better.

Reply 1: We are grateful for your thoughtful suggestion. We apologize for the misleading writing, and we have made appropriate adjustments (**see Page 5, Line 94**).

Changes in the text: What is more, when multiple target proteins co-localize on the same cell, the cross-color interference caused by overlapping signals poses a huge challenge for mIHC. (see Page 5, Line 94)

Comment 2: Line 153 Is this supposed to be brightness in place of RandomBrightnessm.

Reply 2: Thank you for your kind reminder. We are sorry for the spelling mistake and modified the text as advised (**see Page 9, Line 171**).

Changes in the text: There was no other preprocessing except image cropping since many diverse samples decreased the impact of color variability with the application of data augmentation, including flip HueSaturationValue, RandomBrightness, and RandomContrast, to improve the adaptability. (**see Page 9, Line 171**)

Comment 3: Line 287. NSCLC in place of NCLC for consistency.

Reply 3: Thank you for your careful reading. We apologized for the spelling mistake and modified the text as advised (**see Page 15, Line 319**).

Changes in the text: The internal cohort included 1859 TMA images of 121 NSCLC patients. In this cohort, 96 (79.3%) patients were males, and 117 (96.7%) patients were smokers. Most patients (n=92, 76.0%) were under 70 years. Eighty (66.1%) patients had stage-IA NSCLC, while the rest had stage-IB NSCLC. (**see Page 15, Line 319**)

Reviewer #B

The article describes the prognostic models for non-small cell lung cancer, based on the IHC-based metrics, capturing the abundance of cancer cells and immune cells (positive to checkpoint markers) and distances between such cells.

Comment 1: General comment. The paper is well-written linguistically, however, the Result part is very technical and too much shifted to the description of the methods, techniques, statistic etc – i.e. information which could be presented in M&M part. The main line of the clinical importance it therefore partially lost. At the same time, many technical issues remain obscure.

Reply 1: Thank you for spending time on reviewing this paper. We are sorry for the inadequate statement of the clinical value of this study in the Results section. Thus, we added several statements on our potential clinical benefit in the Results section of this version (**see Page 17 Line 354, Page 18 Line 369, Page 24 Line 512, and Page 27 Line 568**).

Changes in the text:

(1) Thus, this model presented a labor-saving way to automatically identify four types of cells with a comparable accuracy with manual recognition, which may promote the clinical routine test of multiple immune checkpoints. (see Page 17, Line 354)

(2) In conclusion, the interaction among the above three pathways (MHC-II/LAG-3, OX40/OX40L, and KIR2D) revealed the great potential of combining immune checkpoint inhibitors, which could provide new ideas for clinical combinational immunotherapy. (see Page 18, Line 369)

(3) The solid prognostic value of the integrated score provided an approach to a convenient risk-stratification of the patient by inputting all patients' relevant immune-checkpoint-staining IHC images into the model. (see Page 24, Line 512)

(4) In short, the high performance of the Res-Score in the external cohort validated its generalization ability in various populations, which provided the great potential to assist clinical decisions in various institutes. (see Page 27, Line 568)

Comment 2: NLR – which data is it? Is it blood test (with I assume) or authors assess these cell ration on the tissue level?

Reply 2: We are very sorry for our negligence of highlighting this issue in our manuscript. In the previous version, we described the source of NLR values in the Methods section as followed: *NLR was calculated as the ratio of neutrophils and lymphocytes count in the pre-operate blood routine results.* In this version, we extra highlighted the pre-operate blood routine as the source of NLR in the Results section again (see Page 26, Line 550).

Changes in the text: Surprisingly, the combination of preoperative NLR from the blood routine and the PD-1/PD-L1 signature was a robust prognostic index for RFS ($P < 0.0001$ for RFS; Fig. 8F) but OS (Fig. 8E), although NLR was not a significant feature for OS and RFS in this population (Figs. 8A and 8B). (see Page 26, Line 550)

Comment 3: I miss better description of the EfficientUnet.

Reply 3: Thank you for your kind advice. We added the detailed introduction of the EfficientUnet and its two components, EfficientNet and UNet, which could better explain why we chose this segmentation algorithm (see). We hope it could meet your expectation. (see Page 8, Line 148)

Changes in the text:

This study performed the EfficientUnet model to segment the TCs and TILs, which was a combination of EfficientNet and UNet (32,33). UNet is a symmetric U-shaped fully convolutional neural network (CNN) developed initially for biomedical image segmentation, which processes a contraction path and an expansion path for encoder and decoder, respectively(32). EfficientNet is an adjusted CNN model which could scale the depth, width, and resolution of networks by a fixed set of scaling factors(33). Considering the better performance of low-level feature maps from the encoder in the complicated spatial analysis, Bhakti Baheti et al. originally applied EfficientNet (with intermediate low-level feature map) as the encoder of UNet (with intermediate high-

level feature map) to replace the previous convolution layers(34). Moreover, the performance of EfficientUnet was much better than the other segmentation algorithms, including Dilated ResNet, ERFNet, DeepLab with ResNet18 Encoder, and the combination of UNet with ResNet or InceptionResNet(34).

As the EfficientNet has eight variants, from EfficientNet-B0 to EfficientNet-B7. According to the preliminary experiment, EfficientNet-B3 has comparable performance and the fewest parameters compared with EfficientNet-B4 to B7. (see Page 8, Line 148)

Comment 4: Authors need to describe how did they deal with artifacts, necroses etc. on the stained and digitalized images during image processing.

Reply 4: Thank you for your careful reading. We feel sorry about the negligence of the exclusion criterion of the impurity in IHC staining. The staining colors of impurity like artifacts and necroses were different from actual positive cells, whose color was close to red and black. Thus, the impurity could be excluded by detecting an accurate range of HSV values of positive cells. Moreover, we also selected five to ten impure dots to define the cutoff values of HSV values of these impurities. The detailed explanation was supplemented (see Page 10, Line 208).

Changes in the text: Meanwhile, we also detected the cutoff of the impurity (such as artifacts and necroses) from 5 to 10 impure false-positive staining of each IHC image according to the same procedure as positive and negative cells. All the impure staining of each slide was excluded based on the cutoff of HSV values. (see Page 10, Line 208)

Comment 5: Although it is briefly mentioned and MnM, I don't really understand how authors consider the collinearity between cell density and proximity. Were the proximity scores in any way normalized to overall density of cell of interest in each individual sample.

Reply 5: Many thanks for your careful reading. We are sorry about our misleading illustration. We want to express that the density of cells (cell/mm²) used in other IHC research is an index with both spatial and quantitative information, which is calculated as the ratio of the number of positive cells and the size of the tissue. However, the IHC images of the training cohort and internal testing cohort were tissue microarrays (TMAs), where all the tissues are cut into similar-sized circles. Thus, the comparative relationship between the densities of different cells depends on the number of positive cells, and the density would be collinear with the number of each kind of positive cells. To explore the quantity and spatial characteristics of immune checkpoints and their roles in NSCLC prognosis separately, we chose to analyze the proximity distance of cells as the spatial feature rather than the density. The corresponding illustration has been added in this version (see Page 11, Line 222).

Changes in the text: The density of cells (cell/mm²) used in IHC research is an index with both spatial and quantitative information, which is calculated as the ratio of the number of positive cells and the size of the tissue. Since the TMAs used in the training and internal testing group were similar-sized circles, the comparative relationship between densities of different cells largely depends on the number of positive cells. To

avoid the multicollinearity between the density and the number of positive cells, we detected the proximity distance between cells as spatial analysis (see Page 11, Line 222).

Comment 6: Authors mentioned manual annotation of 30 cells per class to define the cutoff of positivity/negativity. 1) does it technically mean that the cutoff definition was based on ONE annotated cell from these 30, which had minimal intensity? 2) Did (and how) authors consider heterogeneity of the staining intensity between different samples in the cohort?

Reply 6: We are grateful for the suggestion, and we feel sorry to make you confused. 1) We manually select 30 cells per class from one IHC image and detected the HSV of each cell. The cutoff of each class in this image was the range of HSV values of the 30 cells. Next, we repeated this procedure for each slide to define the cutoff of positive and negative cells one by one. The corresponding illustration was added in this version (see Page 10, Line 205).

2) As mentioned above, the cutoff value of positive cells in each slide was detected manually one by one. Thus, the heterogeneity of the staining intensity could be solved by manual adjustment. The corresponding illustration was added in this version (see Page 10, Line 205 and Line 212).

Changes in the text:

- 1) Here, we manually determined the threshold of positive cells on each slice by measuring the HSV of 30 cells per class with variable shades. The cutoff of each class was the range of the HSV values of the 30 cells of each class. (see Page 10, Line 205)
- 2) Next, we repeated this procedure for each slide to define the cutoff of positive cells one by one, which could manually solve the heterogeneity of the staining intensity among different samples. (see Page 10, Line 212)

Comment 7: Fig 3 and related: as said before, I don't see if authors consider cell density in the proximity calculations. Authors may argue that distance plays primary role independent of either it is driven by overall cell density or by neighboring tendencies, and I would agree on that. But I would still be interested to see how (if) these two types of metrics are independent/dependent.

Reply 7: We are grateful for your insightful comments. As mentioned in Reply 5, we did not calculate the density of cells since the IHC images used in the training and internal testing group were similar-sized circles. Thus, we focused on the number of positive cells and the proximity distance between cells rather than the density. Considering the potential collinearity between the number of positive cells and the density, we consider that the study of these two indexes would be a better quantitative and spatial feature, respectively. The corresponding statement was added in this version as mentioned in Reply 5 (see Page 11, Line 222).

Changes in the text: The density of cells (cell/mm²) used in IHC research is an index with both spatial and quantitative information, which is calculated as the ratio of the number of positive cells and the size of the tissue. Since the TMAs used in the training and internal testing group were similar-sized circles, the comparative relationship

between densities of different cells largely depends on the number of positive cells. To avoid the multicollinearity between the density and the number of positive cells, we detected the proximity distance between cells as spatial analysis (see Page 11, Line 222).

Comment 8: Many figures can be moved to Supp materials, for example Fig 2 (or biggest part of it), Fig 3...

Reply 8: Thank you for your kind advice. We have moved the biggest part of Fig 2 and Fig 3 to the Supp materials according to your comment (see Page 43 of Manuscript and Page 13 and 20 of Supplementary Material).

Changes in the text:

(1) See Page 43:

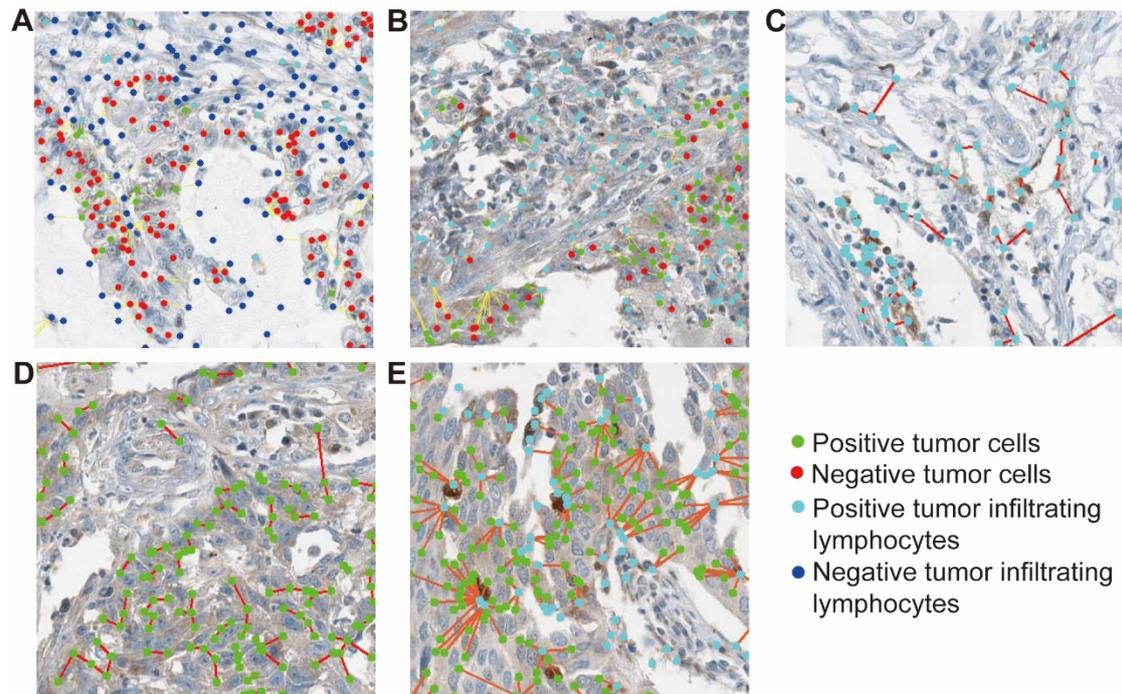


Figure 2. The representative images of segmentation and spatial analysis of the internal cohort. The local magnified images of the distance between all TCs and all TILs (A), all TCs and positive TILs (B), positive TILs and positive TILs (C), positive TCs and positive TCs (D), positive TCs and positive TILs (E). Figs A to C were in $\times 6.2$ magnification, and Figs D to G were 150×150 px. Green dots represented positive TCs; red dots represented negative TCs; light blue represented positive TILs; dark blue represented negative TILs; and the red or yellow lines between cells were straight line distance between two cells.

Abbreviations: TC, tumor cell; TIL, tumor-infiltrating lymphocyte.

(2) See Supplementary Material Page 13:

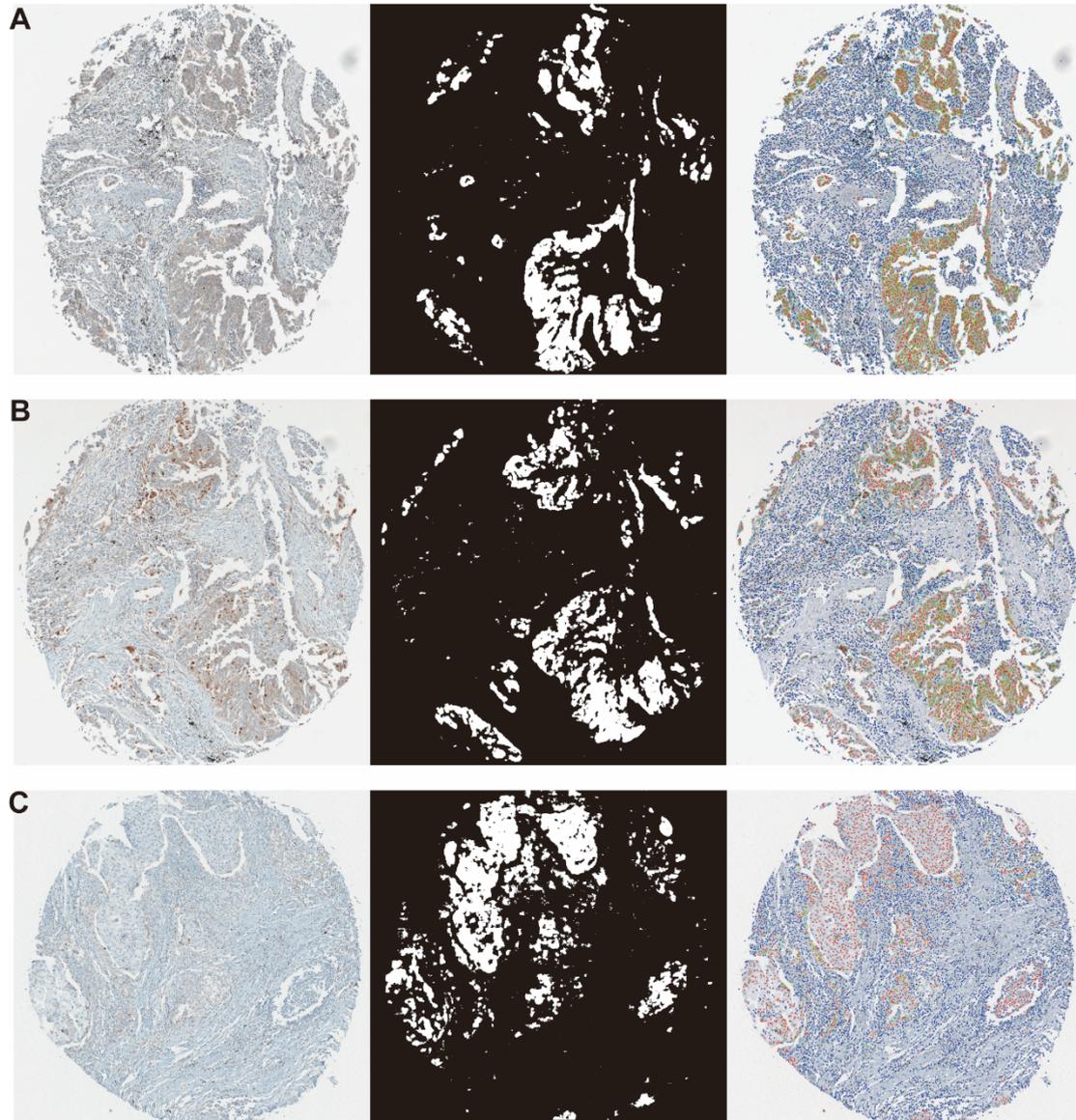


Figure S1. The representative images of segmentation and spatial analysis of the internal cohort. The original IHC image (left), tumor region segmentation mask (middle), and the four classifications of cells (right) of KIR2D (A), galectin-9 (B), and TIM-3 (C).

Abbreviations: IHC, immunohistochemistry; KIR2D, killer cell immunoglobulin-like receptor-2D; TIM-3, T cell immunoglobulin-3.

(3) [See Supplementary Material Page 20:](#)

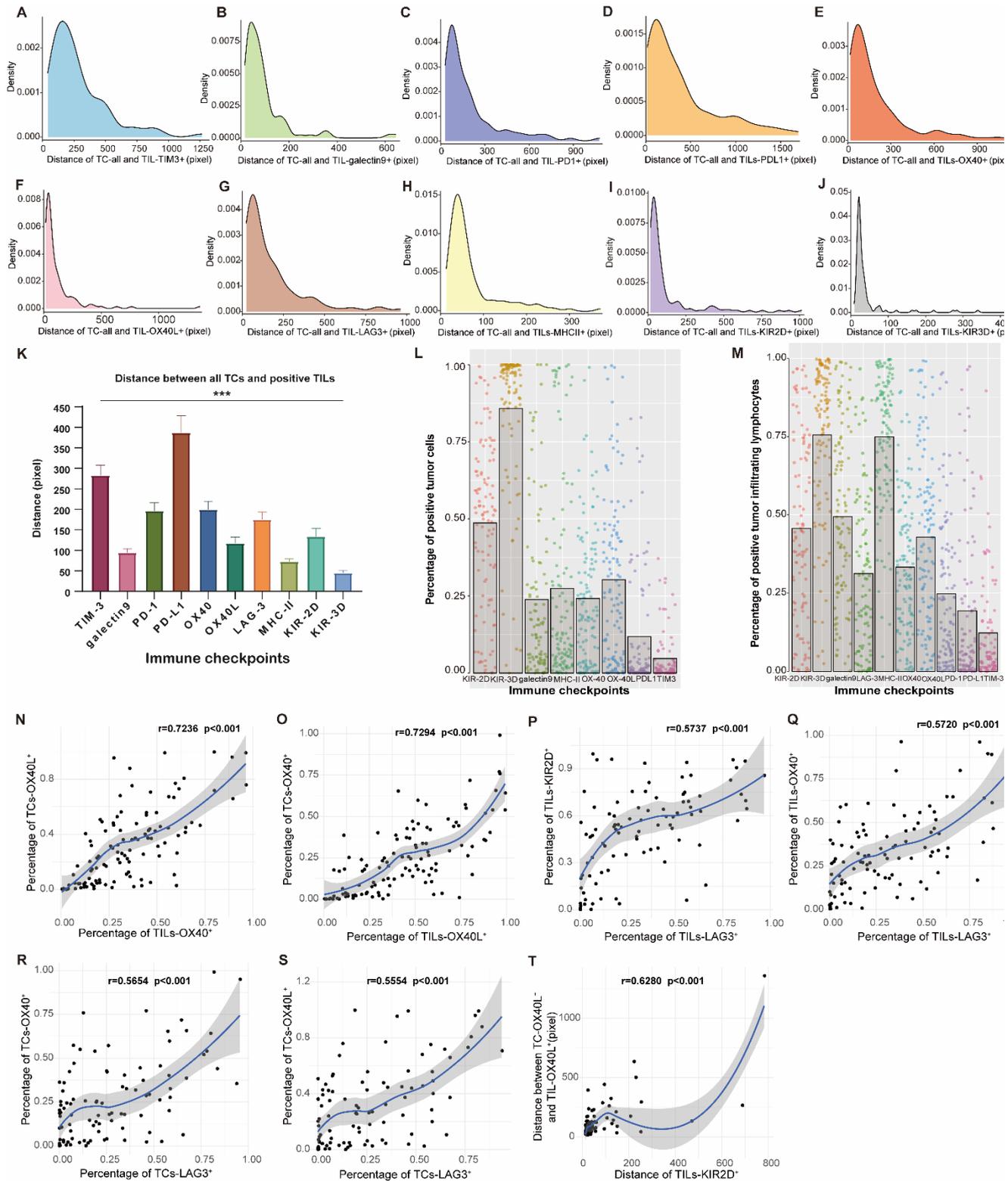


Figure S5. The distribution and correlation of quantitative and spatial data.

Density curves of the distance of TC_{all}-TIL_{TIM3+} (A), TC_{all}-TIL_{galectin-9+} (B), TC_{all}-TIL_{PD-1+} (C), TC_{all}-TIL_{PD-L1+} (D), TC_{all}-TIL_{OX40+} (E), TC_{all}-TIL_{OX40L+} (F), TC_{all}-TIL_{LAG-3+} (G), TC_{all}-TIL_{MHC-II+} (H), TC_{all}-TIL_{KIR2D+} (I), TC_{all}-TIL_{KIR3D+} (J). (K) The column chart of mean distances of all TCs and positive TILs, and the error bars showed the standard error of the mean (SEM).; $P < 0.001$. The distribution jitter blot and mean values column

chart of the percentage of positive TCs (L) and positive TILs (M). Correlation curves of the percentage of TIL_{OX40+} and the percentage of TC_{OX40L+} (N), the percentage of TIL_{OX40L+} and the percentage of TC_{OX40+} (O), the percentage of TIL_{LAG3+} and the percentage of TIL_{KIR2D+} (P), the percentage of TIL_{LAG3+} and the percentage of TIL_{OX40+} (Q), the percentage of TC_{LAG3+} and the percentage of TC_{OX40+} (R), the percentage of TC_{LAG3+} and the percentage of TC_{OX40L+} (S), the distance of TILs-KIR2D+ and the distance between TC_{OX40L+} and TIL_{OX40L+} (T).

Abbreviations: TC, tumor cell; TIL, tumor-infiltrating lymphocyte; KIR2D, killer cell immunoglobulin-like receptor-2D; KIR-3D, killer cell immunoglobulin-like receptor-3D; TIM-3, T cell immunoglobulin-3; LAG-3, lymphocyte activation gene-3; PD-1, programmed cell death receptor-1; PD-L1, programmed cell death ligand-1; MHC-II, major histocompatibility complex class II; OX40L, OX40-ligand.