

LETTER TO THE EDITOR

TCR repertoire and transcriptional signatures of circulating tumour-associated T cells facilitate effective non-invasive cancer detection

The concept of precision medicine in oncology has dramatically transformed the clinical application of tumour screening, which makes the malignancies more curable if diagnosed early. Traditional serological tumour biomarkers like α -fetoprotein, prostate-specific antigens, carcinoembryonic antigens, CA19-9 and CA125 have been widely investigated in clinic, but the specificity has not reached a satisfactory stage for population level.¹⁻³ Novel technologies utilizing tumour-derived signals from blood non-invasively provide a new tumour diagnostic strategy called liquid biopsy over the past decades.³⁻⁵ Several peripheral biomarkers such as cell-free DNA (cfDNA)⁶⁻⁸ especially circulating tumour DNA (ctDNA),⁹ circulating tumour cells (CTCs),^{10,11} circulating micro-RNAs,^{12,13} tumour-derived exosomes¹⁴ and cancer cell metabolites¹⁵ achieved great progress and showed huge prospects in tumour screening. However, these methods are all derived from the modality of tumour and often need predefined panels or biomarkers for diagnosis, which may be non-specific and subjective due to the heterogeneous properties of tumour. The feasibility of using tumour-associated T cell response involved in tumour initiation and development, as a supplementary diagnosis choice has not been explored widely.^{16,17}

Until recently, tumour-infiltrated T lymphocytes (TILs) were considered to be beneficial tumour-specific T cells.¹⁸ But due to the complex interaction of different immune components mediated by chemokines or cytokines within the tumour microenvironment (TME), the majority of passively expanding TILs cannot recognize tumour-specific antigens (TSAs) and are thus believed to be bystander T cells.¹⁹⁻²¹ These bystander T cells may dilute tumour-specific signals and make the identification of TSAs-specific T cells challenging.²²⁻²⁴ Programmed cell death protein 1 (PD-1) is suggested to be a biomarker for tumour-specific CD8⁺ T cells both in TILs and in peripheral blood mononuclear cells (PBMCs),^{22,25-27} but the efficacy needs

to be further validated in practical applications.^{21,28,29} Tracking the general immunophenotype of T lymphocytes when they encounter antigens and enrichment of tumour-associated T cells over a pool of irrelevant signals during tumour development reflects the overall immune status of patients and offers opportunities for cancer prevention and therapy.³⁰

Next generation sequencing (NGS)-based T cell receptor (TCR) repertoire quantification has provided methods for TSA recognition and now is extensively used in the identification of tumour-reactive T lymphocytes.^{19,31-34} The past few years have witnessed a series of studies utilizing T or B-cell repertoire to pinpoint disease-associated signatures, and evidences have demonstrated the diagnostic potential of TCR repertoire in autoimmune diseases,³⁵ infectious diseases^{36,37} and even cancer.³⁸⁻⁴² Sustained neoantigen stimulation during tumour cell development impels TCR to shift towards a tumour-specific distribution and to exhibit different amino acids motifs than those in healthy cells.^{38,43}

Under physiological conditions, naïve T cells maturing in the thymus will flow through peripheral blood or lymphatic vessels and migrate through high endothelial venules into secondary lymphoid organs where they encounter potential tumour antigens.^{44,45} T cell trafficking and circulation theoretically enable these tumour-specific T cells to be detectable both in tumour sites and peripheral blood. T lymphocytes circulating among PBMCs that paired with TILs residing in tumour tissues have been suggested to be highly correlated with T cell-induced cytotoxicity and to indicate enrichment of tumour reactive signals.^{22,46-51} Elucidating the connection between anti-tumour T cells in the periphery and those in the TME^{22,25,44,45,52} may provide clues to design novel approaches for non-invasive tumour screening. Assessing overlapping TCRs between PBMCs and TILs and considering them tumour-specific predictors will not only help

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Clinical and Translational Medicine* published by John Wiley & Sons Australia, Ltd on behalf of Shanghai Institute of Clinical Bioinformatics.

us comprehensively study T cell circulation and migration but will alleviate the deficiencies caused by using the TIL population only, which is enriched with bystander T cells.

In this study, we defined a group of circulating T lymphocytes in PBMCs that shared TCRs with TILs as tumour-associated T cells (TATs). Using the CDR3 sequences of TATs with those in healthy TCRs as input data, we trained a binary model to distinguish TATs from healthy clones. Applying this model on several independent clinical datasets, we acquired the number of TAT sequences in PBMCs for each individual. We then designed a TCR repertoire risk score (TRRS) as the number of TATs our model predicted in the PBMCs divided by the number of detected healthy TCRs from healthy individuals. We demonstrated that the TRRS separated tumour patients from healthy donors effectively. Next, we characterized the transcriptional signatures of TATs in the PBMC populations using multiple single cell RNA sequencing (scRNA-seq) coupled with TCR sequencing datasets and found that T cell activation pathway was significantly up-regulated in TATs. Combining the TCR repertoire and transcriptional signatures of TATs,⁵³ we developed an integrated framework for non-invasive tumour screening using only PBMC samples. Furthermore, we performed bulk TCR and RNA sequencing of PBMC samples from 11 tumour patients and six healthy donors and validated the performance of this tumour screening strategy with these data and another independent cohort. Our study proves the principle of using TATs as an alternative non-invasive tumour screening biomarker and broadens the liquid biopsy application from the view of the immune landscape.

T cells with identical TCR sequences are thought to be derived from a single naïve T cell, which migrates and circulates among different tissue types and may undergo a functional state transition upon antigen stimulation.⁵⁴ Based on the TCR-sharing relationship of TILs and PBMCs, we first divided TCR clonotypes in PBMCs and TILs into four different compartments (Figure 1A). TCR clonotypes of PBMCs that are identical to those of TILs are called *PBMCs_Shared* compartment, while clonotypes of TILs that are identical to those of PBMCs are called *TILs_Shared* compartment. These two compartments have the same TCR clonotypes, but their tissue sources distinguish them. Due to different tissue environment, the frequency of each clonotype and the degree of clonality between the two compartments may differ. Therefore, we specifically named the T cells among PBMCs that share TCRs with TILs as circulating Tumor Associated T cells (cTATs). In contrast, the TCR clonotypes that are unique to TILs are in the *TILs_only* compartment, and the TCR clonotypes observed only in PBMCs are in the *PBMCs_Only* compartment. We believe that the *PBMCs_Only* compartment most likely represents naïve,

effector or memory T cells in periphery that are not related to the tumour immune response. T cells in the *TILs_Only* compartment may largely represent tissue-resident T cells, which are not in the set of T lymphocytes prevalent in circulation. It should be noted that due to the technical limitation of TCR repertoire sequencing, both *TILs_Only* and *PBMCs_Only* compartments actually contain a proportion of overlapped clonotypes that cannot be detected sensitively at present.

This framework allowed us to use TCR sequences as molecular barcodes to track and analyse the function of TATs among TILs and PBMCs. We collected a series of TCR CDR3 β chain sequencing data of paired PBMCs and tumour tissues from the same patient (Table S1) and assigned each T cell clonotype to one of the four compartments using the aforementioned definitions. After removing non-functional TCRs, we found that, among the different compartments, most (91%) of the TCR CDR3 sequences were 12~17 amino acids in length (Figure 1B). In addition, the CDR3 sequence length distribution in the TILs resembles that in the PBMCs, indicating that there is no CDR3 length difference between these cells in the two different tissue types.

Next, because TCR beta chain variable (TCRBV) genes contributed the most diversity to CDR3 sequences, we analysed the TCRBV gene usage in the different compartments. We found that TRBV genes, such as TRBV06/14/15/20/25, are expressed more frequently in the shared compartments than in the non-shared compartments⁴³ (Figure 1C), which suggests that the antigen specificities of TCRs may differ between these compartments. In the following analysis, subsets with CDR3 sequences of 12~17 amino acids were analysed, and TCRs with excessively long or short CDR3 regions were removed.

To estimate the degree of TCR sequence overlap in PBMCs and TILs, we calculated the relative proportion of TCRs shared by both TILs and PBMCs in each sample. We found that the proportion of TCRs in the *TILs_Shared* compartment (approximately 22.56%, 95% confidence interval (CI): 12.82%–37.51%) was significantly higher ($p < 1e^{-4}$) than the proportion in the cTATs of the PBMC population (approximately 3.08%, 95% CI: .64%–5.15%, Figure 1D), which indicated that TILs show higher shared TCR enrichment than PBMCs possibly due to the close interaction of T lymphocytes with TSAs in the TME. This result is consistent both at the TCR clone and TCR sequence level (Figure 1D). The same analysis based on single-cell TCR sequencing (TCR-seq) data showed no significant differences ($p > .05$) in the proportion of shared TCRs at either the clone or sequence level, possibly due to the limited number of cells captured in single cell TCR-seq datasets, and many shared clones might be labelled as not shared (Figure S1A).

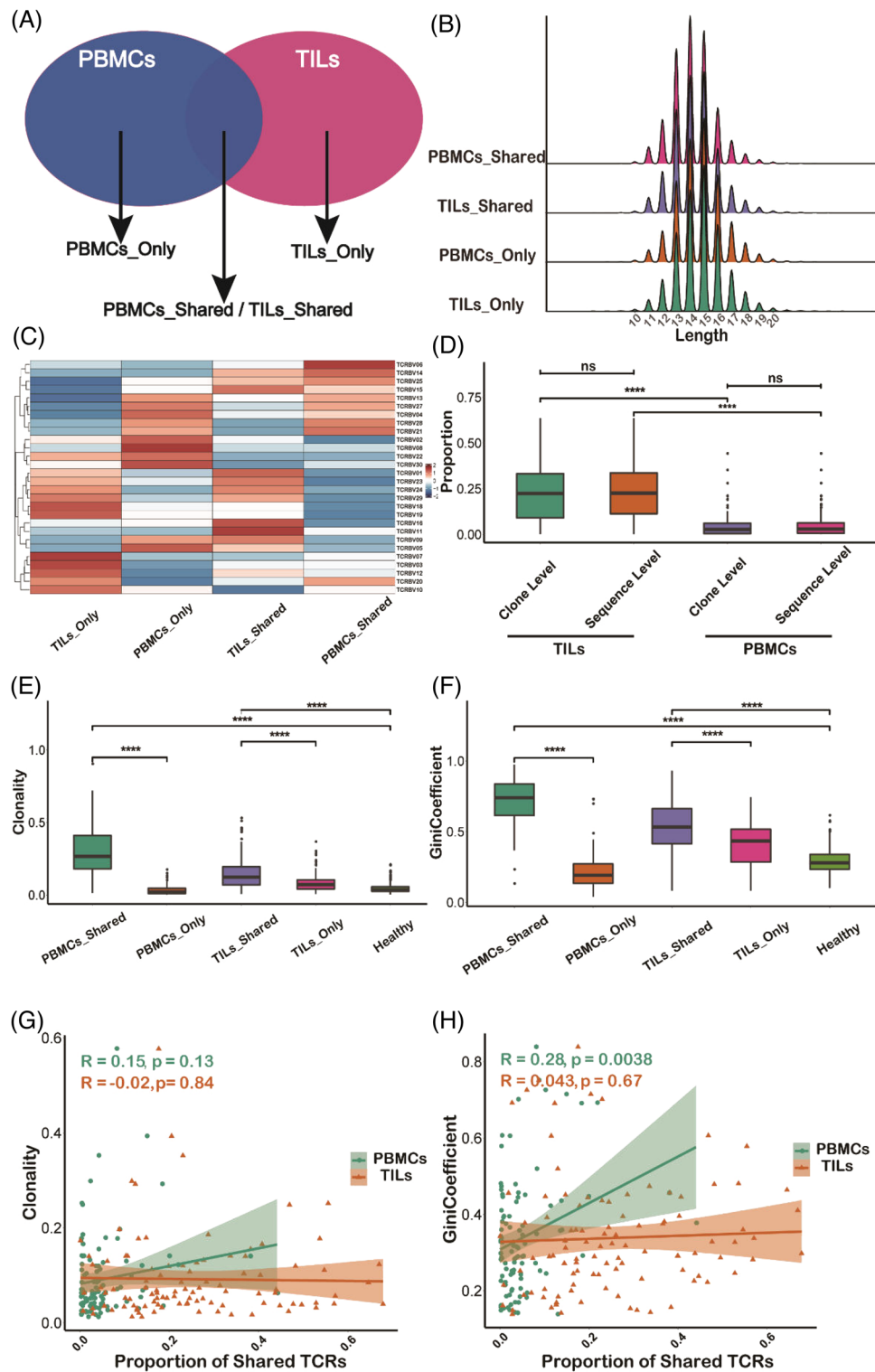


FIGURE 1 T cell receptor repertoire (TCR) sharing relationship between the tumour microenvironment, and the periphery defines four different compartments. (A) Schematic overview of TCR sharing relationship between PBMCs and tumour-infiltrated T lymphocytes (TILs) and the definition of the four TCR compartments. (B) Length distribution of the CDR3 beta chains within the four different TCR compartments. (C) V gene usage of the CDR3 beta chains in different TCR compartments. (D) Proportion of shared TCRs among different tissue types (TILs or PBMCs) at both sequence level and clone level. (E-F) The indices of clonality and Gini coefficient among different TCR compartments and that from healthy donor PBMCs samples. (G-H) Correlation between the overall clonality/Gini coefficient and the proportion of shared TCRs in PBMCs and TILs

TCR sequences with a high degree of similarity and clonal expansion are more likely to recognize TSAs effectively. We found the indices of clonality and Gini coefficient in the *PBMCs_Shared* compartment were higher ($p < 1e^{-4}$) than those in *PBMCs_Only* compartment (Figure 1E,F). The same trend was also observed in the PBMC population when single cell TCR-seq data were analysed (Figure S1B). These results indicate that TATs are more likely to undergo clonal expansion and to represent functional tumour-reactive T cells. Adding a healthy donor cohort PBMC dataset³⁶ as the control (see Methods), we found that the clonality and Gini coefficient of the healthy samples were lower ($p < 1e^{-3}$) than those of the shared compartments and higher than those of the tissue-only compartments (Figure 1E,F), possibly due to the baseline immune activity that developed against common antigens in the surrounding environments, such as influenza virus or human cytomegalovirus (HCMV). Our results suggest T cell clones in the shared compartment are more likely to be tumour reactive and are different from those induced by non-tumour antigens commonly present in healthy individuals.

Moreover, we found that the proportion of TATs in blood, which is observable only when tumour tissue is sequenced, was highly correlated with the overall T cell clonality and Gini coefficient of PBMCs, which were obtained non-invasively; however, such correlation was not observed in TILs (Figure 1G,H). These results highlight the potential of using the T cell clonality and Gini coefficient of PBMCs as indicators of cancer development. TATs among PBMCs are more likely to reflect the clonal expansion of T lymphocytes in periphery, and a higher degree of shared TCR clones among PBMCs may indicate that more neoantigen-specific T cells pre-exist in the PBMCs.

It has been reported that a greater degree of PBMC-TIL TCR repertoire overlap indicates an improved immune response and is associated with better clinical outcome of immunotherapy.^{42,47,49,55} We believe that this compartment largely represents tumour reactive T cells and may serve as a biomarker to distinguish blood samples of cancer patients from those of healthy individuals. In this study, we sought to build a deep learning binary classifier to predict tumour-reactive TCR sequences. To construct a training dataset for the model, we first downloaded a publicly available TCR sequencing data obtained from PBMC samples of healthy individuals³⁶ as the control dataset and only used data from HCMV-negative individuals to exclude potential tumour-irrelevant immune signals. Two datasets of healthy cohorts were included in the analysis, and we named these sets Healthy351 and Healthy69 according to the number of samples after filtering. Since the Healthy351 included more healthy donors and TCRs (more than 30 million), we considered the TCRs in this

cohort to be a healthy TCR pool and used these data to identify the TCR sequences that overlapped with those in the PBMCs from cancer patients. Then, we extracted TAT TCRs in PBMC samples from TIL-PBMC-paired TCR sequencing datasets described above and filtered TCRs that were also detected in Healthy351.

We labelled the remaining TAT TCRs as positive samples and the TCRs in Healthy351 as negative samples. The schematic workflow and experimental design are summarized in Figure S2A. Deep convolutional neural networks (CNNs) generally performed better in TCR pattern recognition studies^{56–58}; therefore, we encoded the CDR3 beta chain using the one-hot encoding method and built a three-layer CNN to distinguish the TCRs of TATs from those of healthy individuals. The output of the CNN is the probability of each input TCR sequence being the TCR of a TAT. Next, we generated a TRRS for each PBMC sample summarizing the number of TATs our model had predicted in PBMC relative to the number of healthy TCRs that had been detected in the Healthy351 dataset. We evaluated the performance of the TRRS for non-invasive cancer detection with several independent PBMC datasets obtained from cancer patients using Healthy69 as negative samples. The detailed illustration of model construction is presented in the Methods.

We first selected the same number of negative TCRs as that of TATs and used five-fold cross validation to test the generalization ability of our CNN model. Considering the heterogeneity of cancer patients, the data were split at the patient level rather than at the TCR sequence level to ensure that the model did not learn sample-specific confounding effects. Both the receiver operating characteristic curve (ROC) curve and precision-recall curve (PRC) (Figure 2A,B) showed the model performed modestly well in differentiating TCRs of TATs from TCRs of healthy samples (ROC: .699–.706, PRC: .446–.787) and are not influenced by human leukocyte antigen (HLA) haplotypes (Methods, Figure S2B,C). Because we randomly split patients into the training and test dataset and high variation in the number of TATs exists in different patients, the PRC shows high variability across the different iterations of random splits. The final model was trained and validated using the entire data, and as the number of training epochs increased to about 60, the loss and accuracy had reached a plateau (Figure 2C). Further inspection of the prediction probability distribution of different CDR3 length indicated a significant difference between the TCRs of TATs and those of healthy samples (Figure 2D).

Then, to test whether the TAT prediction probability of our model can be used as a biomarker for differentiating tumour patients from healthy donors, we obtained seven independent datasets^{55,59–64} containing PBMC TCR sequencing data of patients with different cancer types

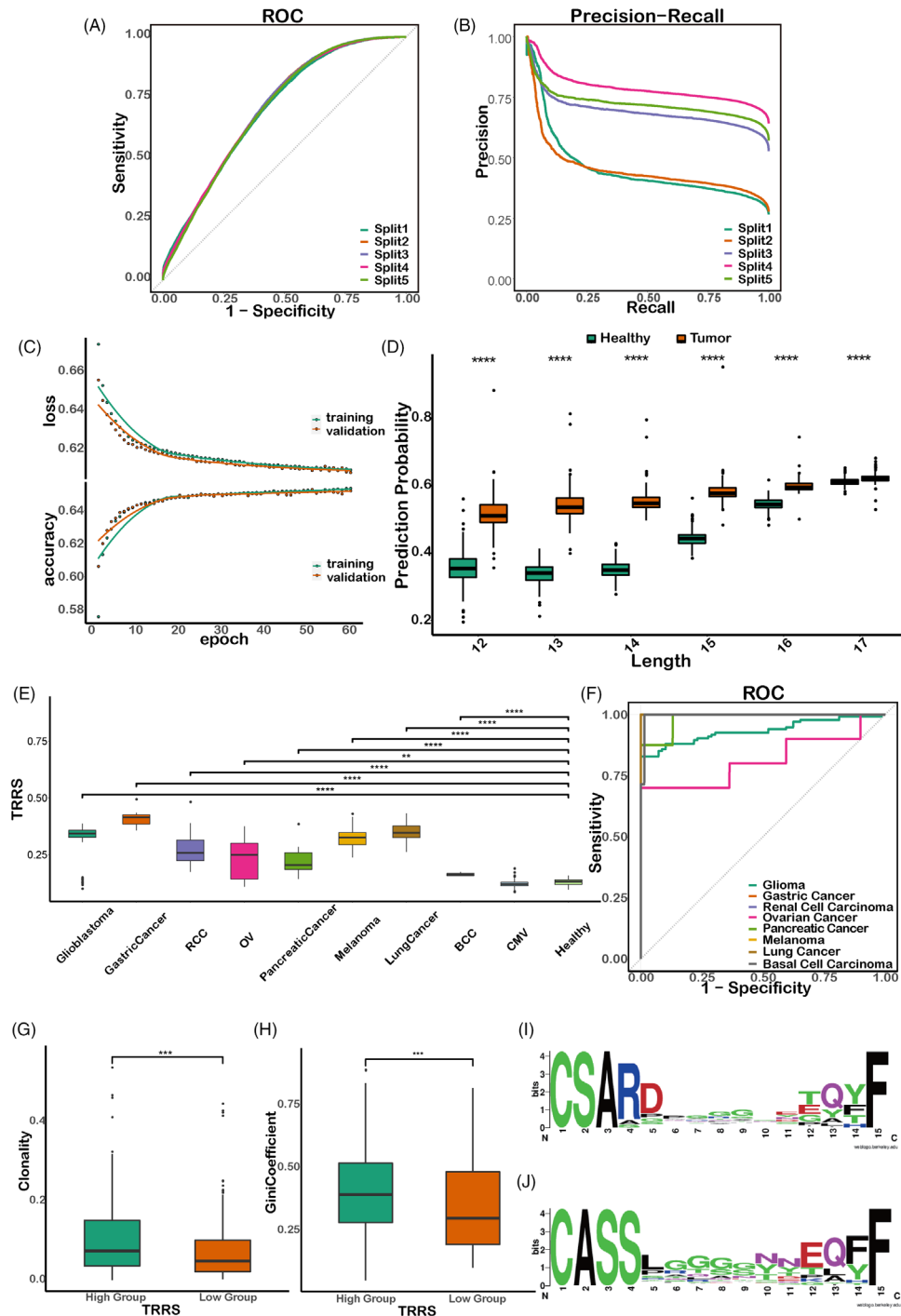


FIGURE 2 Development of binary model to predict tumour-associated T cells (TATs) in the peripheral blood and using the T cell receptor repertoire (TCR) repertoire risk score (TRRS) to distinguish tumour patients from healthy individuals. (A–B) Model ROC and precision-recall curve (PRC) by five-fold cross validation. (C) The model loss/accuracy changes at different epochs. (D) The model prediction probability score for different CDR3 lengths among tumoural and non-tumoural TCRs. (E) The TCR repertoire risk score (TRRS) in PBMC samples among tumour patients from different cancer types, healthy individuals from the Healthy69 cohort and human cytomegalovirus (HCMV) positive samples. (F) The ROC plot of TRRS in differentiating tumour patients from healthy donors at a given threshold of .66. (G–H) The indices of clonality and Gini coefficient between TRRS-high group and TRRS-low group. (I–J) The amino acid motif between TAT probability high group (top panel) and low group (bottom panel)

(Figure S2A). The Healthy69 cohort was used as the negative controls set. We used a TRRS (Methods) to estimate the degree of TAT enrichment in the PBMC population of each sample. Briefly, we counted the number of TATs our model has predicted and then divided it by the number of healthy TCRs in PBMCs that overlapped with the Healthy351 pool.

At a threshold above .66, the TRRS can differentiate PBMC samples of normal individuals from those of patients with various cancer types effectively (Figure 2E,F), indicating the feasibility of using TRRS of TATs for non-invasive tumour screening. We found that our result is robust by setting the threshold at different levels (Figure S3A–F). To provide evidence that the prediction of our model was not simply a generic active cell-mediated immune response, we introduced an experimentally validated HCMV positive cohort with TCR-seq data from their PBMC samples (Table S1). We calculated the TRRS for these individuals and compared them with that of the cancer patients. We show that the TRRS of HCMV positive cohort is lower than that of the cancer patients, indicating our model is able to distinguish individuals with cancer from those with infectious diseases. Based on the TRRS, we divided the samples in the independent validation cohort into high-risk and low-risk group separated at the 50% quantile. We found that the clonality and Gini coefficient in the high-risk group were significantly higher ($p < 1e^{-3}$) than those in the low-risk group (Figure 2G,H), implying that the TCRs in the high-risk group were associated with more clonal expansion and active immune functions. In addition, we analysed sequence motif enrichment in TCRs with the top 25% and bottom 25% probabilities of being a TAT (Figure 2I,J). We found that CDR3 sequences in the high probability group showed an enrichment of serine in the second position, while those in the low probability group tended to have an alanine in this position. Sending publicly available databases of virus/bacterial TCR sequences into the prediction model shows that CDR3 sequences among different lengths tended to have an alanine in the second position, which indicates that features specifically associated with the TCR of clonally expanded TATs are not enriched in virus/bacteria TCRs (Figure S3L). In summary, we used TCR sequences from TATs and healthy donors to build a binary predictive CNN model and designed a TRRS based on the model prediction for effective non-invasive cancer screening with PBMCs.

Single-cell-RNA-multiplexed-TCR-sequencing (scRNA-TCR-seq) technology makes it possible to not only trace the TCR clone sharing relationship between tumour and paired PBMCs samples, but also quantify the transcriptomics patterns at the single cell level.^{43,54,65–69} We performed a comprehensive literature review and obtained 14

high-quality scRNA-TCR-seq datasets that met our criteria (Table S2). We found no significant difference ($p > .05$) in the CD4⁺/CD8⁺ ratio between PBMCs from cancer patients and healthy donors (Figure 3A), implying that the relative ratio of CD4⁺ and CD8⁺ T cells remains unchanged after tumour initiation, in contrast to the ratio in patients with acute infection, which is usually lower than that in uninfected healthy samples.^{70–72} However, the proportion of clonal T cells (clone frequency > 2) was higher ($p < 1e^{-3}$) in tumour patients than in healthy donors (Figure S4A), indicating a higher clonal expansion of T cells in cancer patients. Moreover, we found that the CD4⁺/CD8⁺ ratio was significantly lower in TATs than in non-clonal T cells (clone frequency = 1) in patient PBMCs (Figure 3B), suggesting that the expansion of CD8⁺ T cell is greater than that of the CD4⁺ cells upon tumour antigen stimulation.

Next, to provide further evidence that TATs represent tumour-specific T cells among PBMCs, we performed clustering analysis utilizing both single-cell RNA and TCR information of T cells among tumour and healthy donor PBMCs by the TCR functional landscape estimation supervised with scRNA-seq analysis (TESSA)⁷³ algorithm. We found that tumour-specific clusters had a higher proportion of TATs than normal specific clusters ($p < 1e^{-4}$) in nearly all the 14 datasets (Figure 3C, Methods), indicating that TATs in the blood of patients with various types of cancer are tumour specific and dissimilar to T cells in the blood of healthy donors. To integrate datasets from various sources, we used a label transfer method⁷⁴ by taking one clear cell renal cell carcinoma (ccRCC) dataset⁷⁵ as the reference due to its detailed cell type annotation information. Then we projected the cells from other datasets onto the reference map to transfer the cell type annotation. We found that most T cells in the non-clonal group were CD4⁺ naïve/proliferating/effector T cells, while most TATs were CD8⁺ NK-like/effector T cells (Figure 3D). These results demonstrated that TATs are mostly activated CD8⁺ T cells and may exert cytotoxic functions upon tumour stimulation.

To further explore transcriptional signatures of TATs among PBMCs, we performed differential gene expression analysis between CD8⁺ TATs and non-clonal T cells in each dataset. To prevent potential batch effects caused by using different data sources and guarantee a robust analysis, genes that were differentially expressed in more than 10 of the 14 datasets were selected to generate the TAT signature genes (Figure 3E). Enrichment analysis of TAT signature genes indeed implies a T lymphocytes activation and cytotoxicity function exemplified by pathways like antigen processing and presentation, PD-1 signaling and immune effector process pathways, which are involved in anti-tumour process (Figure 3F).

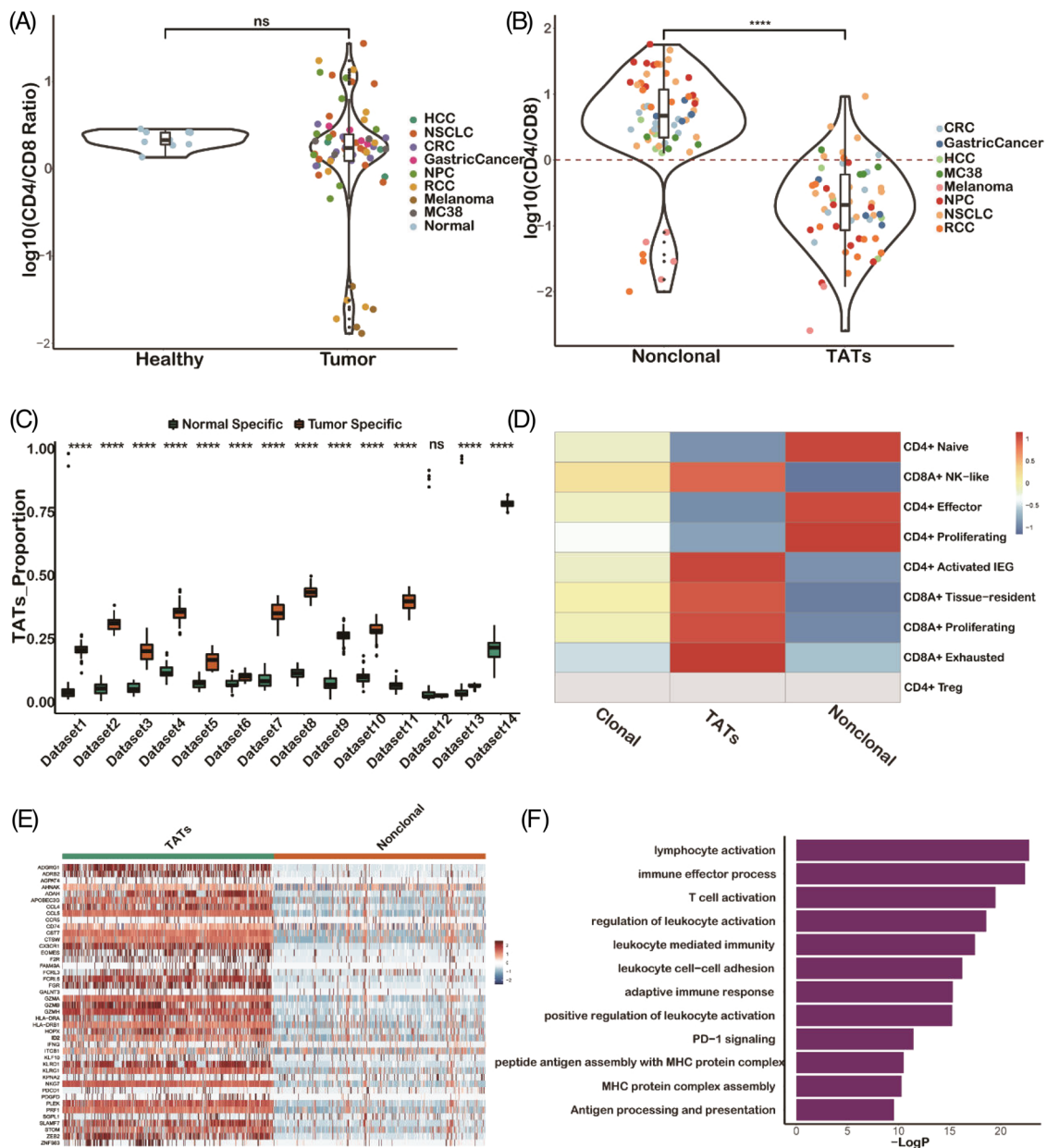


FIGURE 3 Transcriptional signature gene analysis using single cell data shows T cell activation is involved in the tumour-associated T cells (TATs). (A) $\text{CD4}^+/\text{CD8}^+$ T cell relative ratio in PBMCs between tumour and healthy samples calculated by single cell data. (B) $\text{CD4}^+/\text{CD8}^+$ T cell relative ratio between TATs and non-clonal T cells in PBMCs. (C) The proportion of TATs in tumour and healthy-specific clusters (in which tumour patients or healthy donors derived T cells occupy more than 70%, see Methods) by TCR functional landscape estimation supervised with scRNA-seq analysis (TESSA) clustering analysis. (D) T cell subtype distribution in different T cell receptor repertoire (TCR) compartments. (E) Heatmap shows differentially expressed genes between TATs and non-clonal T cells. (F) Gene Ontology (GO) analysis of TAT differentially expressed genes

Similarly, T cell activation was enriched in non-clonal group from patient PBMCs compared with that in T cells from healthy PBMCs (Figure S4B), indicating that non-clonal T cells in tumour patients are generally more active than T cells in healthy donors. The reason might be two-fold: (1) clonal T cells were labelled as non-clonal due to sampling dropout in single-cell experiments; (2) systematic immune response was induced upon tumour stimula-

tion. Moreover, we found that the frequencies of TAT TCRs in tumour tissues were significantly correlated with those in PBMCs (Figure S4C), suggesting that the clone size of the TATs among PBMCs reflects the clone size in tumour sites to some extent.⁴⁸

T cell metabolism is coupled with many immunological signals and facilitates the adaptation of T cells encountering pathogens and tumours.⁷⁶ We mined Kyoto

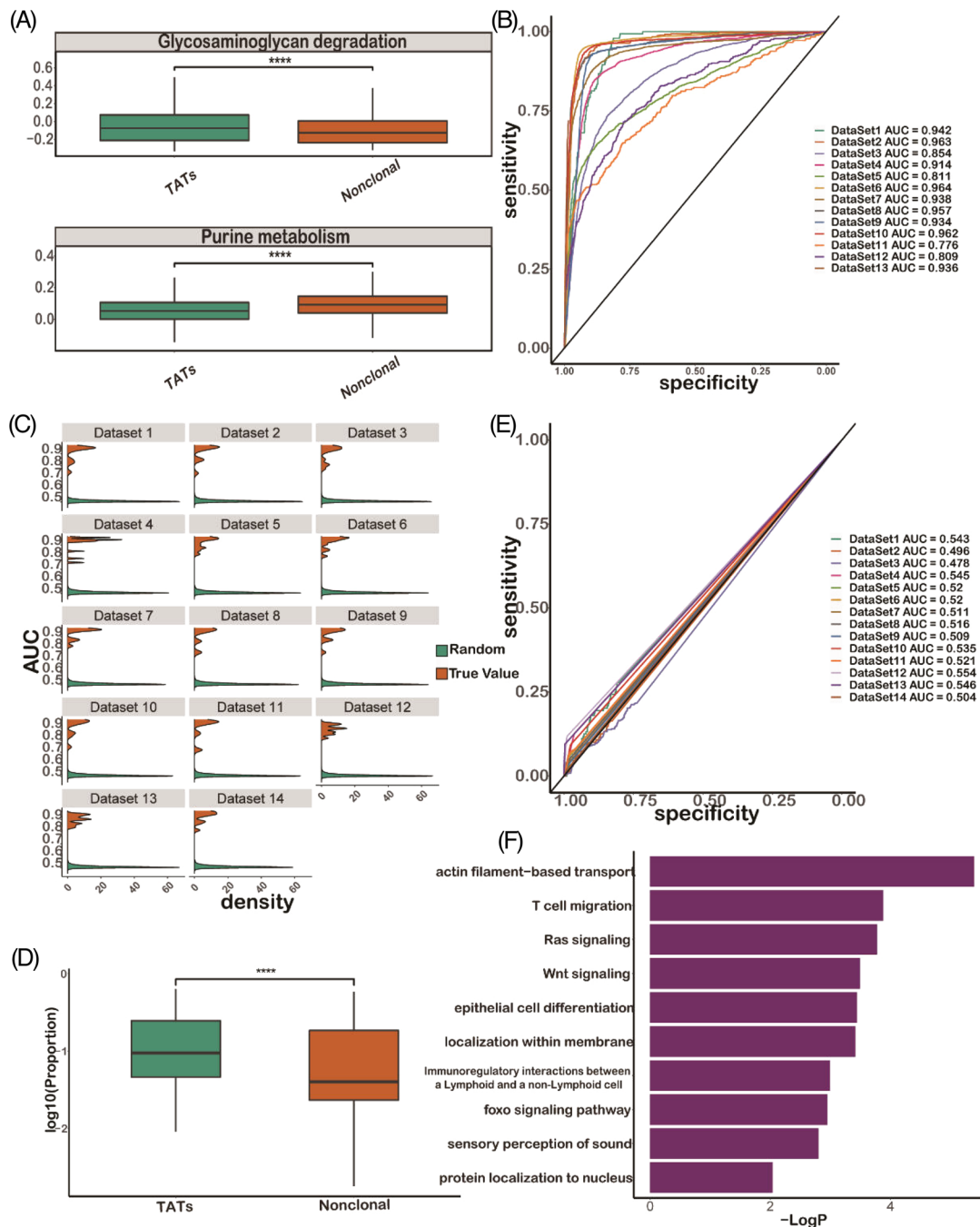


FIGURE 4 Transcriptional signatures are generalizable to predict the tumour-associated T cells (TATs) from other single cell datasets. (A) Glycosaminoglycan degradation pathway is enriched in TATs, while the purine metabolism pathway is depleted in TATs. (B) ROC plot by using a colorectal cancer (CRC) dataset to train the prediction logistic regression model and testing the performance of model on the other datasets. (C) Permutation analysis by randomizing the true labels of TATs 1000 times and comparing the Area Under the Curve (AUC) distribution with the true prediction AUC values on each dataset. (D) PD-1 expression between TATs and non-clonal T cell group. (E) ROC plot using the expression of PD-1 alone to differentiate TATs from non-clonal T cells. (F) Differential expression analysis between TATs in PBMCs and *TILs_Shared* compartment in tumour-infiltrated T lymphocytes (TILs)

Encyclopedia of Genes and Genomes (KEGG) metabolism pathways and analysed metabolic pathway changes in TATs. We found that the glycosaminoglycan degradation pathway was significantly enriched ($p < 1e^{-3}$) in TATs, while the purine metabolism pathway was significantly

depleted ($p < 1e^{-3}$) in TATs (Figure 4A). The glycosaminoglycan degradation pathway is mediated by enzymes produced by activated T cells, and it has been reported to be involved in the immune response and regulation of T cell homeostasis.^{77,78} On the other hand, purine metabolism

especially the adenosine synthesis axis serves as a common path for attenuating T cell activation and can mediate regulatory T cell to suppress immune activity^{79–81}; therefore in the TATs compartment, this metabolic pathway is down-regulated.

Since the TAT data we collected were from samples manifesting different cancer types, with some obtained from different species, we wondered whether the transcriptional signatures of TATs in these datasets can be generalized to predict TATs in other datasets. We therefore performed a leave-one-dataset-out cross-validation experiment and used signatures of TATs explored in each dataset to build a logistic regression binary model. Then, we applied the model to independently test the TAT prediction performance on all the other datasets (Methods). The ROC demonstrated that the gene module can be satisfactorily generalized across different datasets (Figure 4B, Figure S4D–O) satisfactorily, and a permutation analysis demonstrated the robustness of our signatures (Figure 4C).

Our results suggest that using TCR sharing as molecular barcodes, we can characterize the transcriptional signatures of a group of tumour-specific TATs. Previously, PD-1⁺CD8⁺ T cells were believed to be tumour-specific biomarkers in both TILs and PBMC population,^{28,29} and we found that PD-1 was indeed highly-expressed in TATs (Figure 4D) compared with its expression in non-clonal T cells ($p < 1e^{-4}$). However, using the normalized expression of this gene alone was not sufficient to distinguish TATs from non-clonal T cells (Figure 4E).

Because of the differences in TME and blood, TATs circulating in the periphery may acquire immunophenotypes that differ from those in TME, even when the TCR sequences are the same. Using the aforementioned procedure, we performed differential gene expression analysis between cells of TATs among PBMCs and T cells of *TILs_Shared* among TILs. We identified genes that were consistently differentially expressed in more than 10 datasets. We found that compared to T cells of *TILs_Shared* among tumour tissue, TATs among blood showed significantly higher levels of T cell migration and immunoregulatory interactions between lymphoid and non-lymphoid cell (Figure 4F), further demonstrating that circulating TATs have the capacity to migrate across periphery, lymphoid and tumour tissues.

After identification of transcriptional signatures in TATs on the basis of scRNA-TCR-seq data, we sought to validate whether TAT signature genes can be applied to classify PBMC samples with bulk RNA-seq. We collected three independent bulk RNA-seq PBMCs datasets comprising 33 breast cancer PBMC samples, eight hepatocellular carcinoma (HCC) PBMC samples and 12 healthy PBMC samples. We named these datasets validation cohort 1 (Table S3). Principle component analysis showed that

the tumour samples were transcriptionally separated from healthy samples (Figure 5A) in validation cohort 1, indicating salient differences between the peripheral blood of cancer patients and that of healthy individuals. We found that the expression pattern of TAT signature genes was distinct in tumour and healthy samples (Figure 5B), confirming that the signature module derived from the scRNA-TCR-seq data can also be used with bulk PBMCs RNA-seq data to distinguish cancer patients from healthy individuals.

The tumour reactive immune response is usually triggered by immunogenic neoantigens expressed on tumour cells. Therefore, we sought to determine whether the tumour immunogenicity score (TIGS)⁸² is correlated with the expression of TAT signature genes. We utilized the mutation and transcriptional data of 31 cancer types in the The Cancer Genome Atlas (TCGA) database and defined the TIGS as the tumour mutation burden multiplied by the mean expression of a group of antigen-presenting genes for each sample. We found that in nearly all the cancer types, the TIGS was significantly higher in the TAT-signature-high group than in the TAT-signature-low group (Figure 5C), providing evidence supporting our hypothesis that transcriptional activation of TATs is associated with the tumour mutation load.

Both the TCR repertoire and transcriptional signatures of TATs have the capacity to distinguish tumour patients from healthy samples. To determine whether the combination of these two types of features can be leveraged to achieve better performance for non-invasive tumour screening, we performed bulk TCR and RNA sequencing of PBMCs samples from 11 tumour patients and six healthy donors (Table S4). Hereafter, we named this dataset as validation cohort 2. We merged the PBMC samples from validation cohort 1 and 2 to increase the sample number included in the integrated analysis. We found that the clonality and Gini coefficient of the PBMCs in cancer patients were significantly higher than those in healthy donors (Figure 6A,B), which is consistent with the results reported in the previous sections. Next, we predicted the relative proportion of immune cells in each sample by cell type deconvolution of the bulk RNA-seq data using an R package, Immuno-Oncology Biological Research (IOBR).⁸³ We found that the fractions of T cells-related components such as Tregs and T follicular helper cells were significantly different between tumour and healthy PBMCs samples (Figure 6C).

Next, by integrating these two types of features, we designed a novel tumour screening framework. We first used the TAT prediction model to generate the TRRS for each PBMC sample. Then, we calculated the TAT signature score using bulk RNA-seq data obtained from the same sample. The two scores were then used to build a binary logistic regression model and define the final

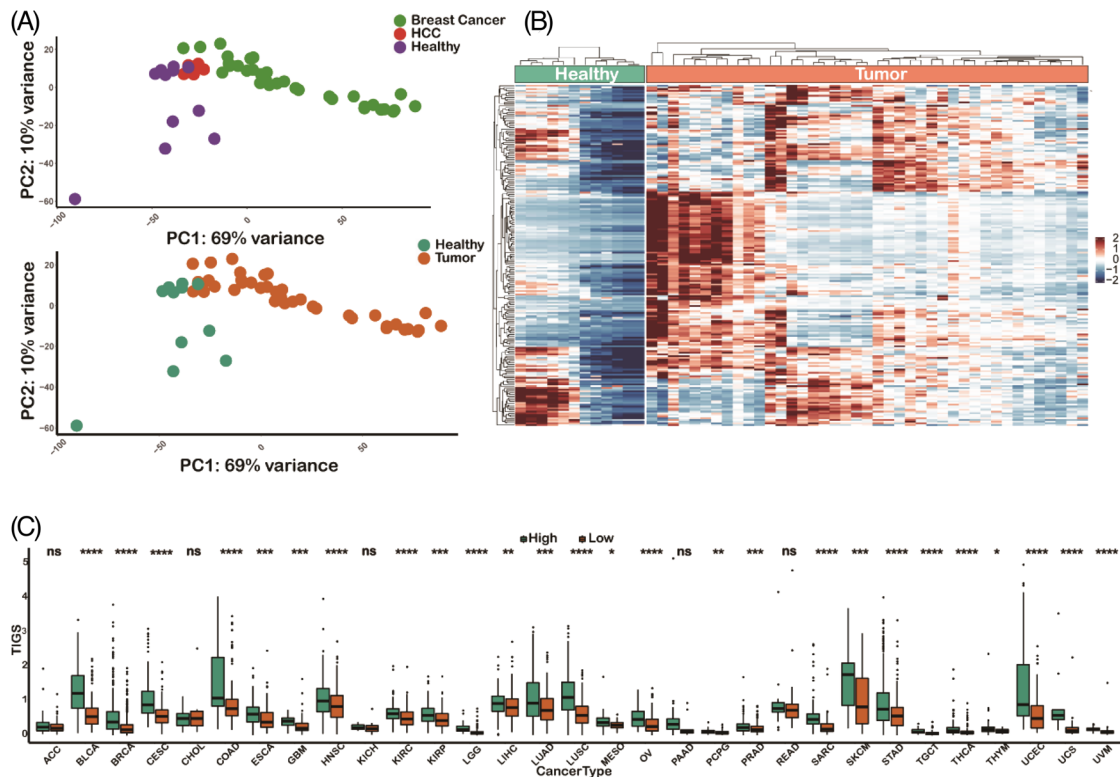


FIGURE 5 Transcriptional signatures acquired from single cell datasets can be used to distinguish tumour patients from healthy individuals of bulk RNA-seq data. (A) Principle component analysis (PCA) plot of PBMC samples from tumour patients and healthy donors. (B) Heatmap of the expression patterns of tumour-associated T cells (TATs) signature genes from tumour and healthy bulk RNA-seq samples. (C) The tumour immunogenicity score (TIGS) in 31 cancer types from The Cancer Genome Atlas (TCGA) database between the TAT-signature-high group and the TAT-signature-low group

cancer risk score as the probability of the model prediction (Figure 6D). We found that either TRRS or TAT signature score has the capacity to distinguish blood samples of cancer patients from those of healthy individuals. Combining these two scores led to better predictive performance (Figure 6E), demonstrating the potential application of the two scores for use in non-invasive tumour screening.

Understanding the tumour-specific T lymphocyte response helps us explore immune signatures of tumorigenesis. In this study, we first separated the T cells into four different compartments based on the TCR-sharing relationship and analysed the clone expansion differences between these compartments. We used a group of TATs in the PBMC population that shared TCRs with TILs and a group of healthy TCRs to build a binary model. We designed a TRRS, which is the number of predicted TATs among PBMCs divided by the number of healthy TCRs in the healthy TCR pool. We found that TRRS can serve as an indicator to distinguish tumour patients from healthy donors in a series of clinical cohorts with different cancer types. In addition, leveraging the scRNA-TCR-seq data of paired tumour tissues and PBMCs, we characterized the transcriptional signatures of TATs and found that T

cell activity and cytotoxicity were increased in TATs. TAT signature genes mined from single cell data were also validated with bulk RNA-seq PBMC data, showing high applicational prospect. Finally, combining the information from the TCR sequence and signature gene level of TATs, we designed an integrated framework for tumour screening and validated it with an independent clinical dataset. This framework considered tumour-associated TCR repertoire and transcriptional patterns of TATs, providing insights into an alternative strategy of liquid biopsy on the basis of the immune cells.

Our method offers a new perspective for tumour screening harnessing the tumour-associated immune response. Imaging-based diagnostic assessments, including breast mammogram and low dose computed tomography scan screening methods, are used in limited cancer types.^{84,85} False positive rate and overdiagnosis remain major concerns.⁸⁶ Serum protein-based markers such as α -fetoprotein, prostate-specific antigens, carcinoembryonic antigen, CA19-9 and CA125 have been investigated in clinical studies, but the specificity is not satisfactory for population level tumour screening.¹⁻³ Liquid biopsy methods utilizing ctDNA usually need predefined biomarkers of

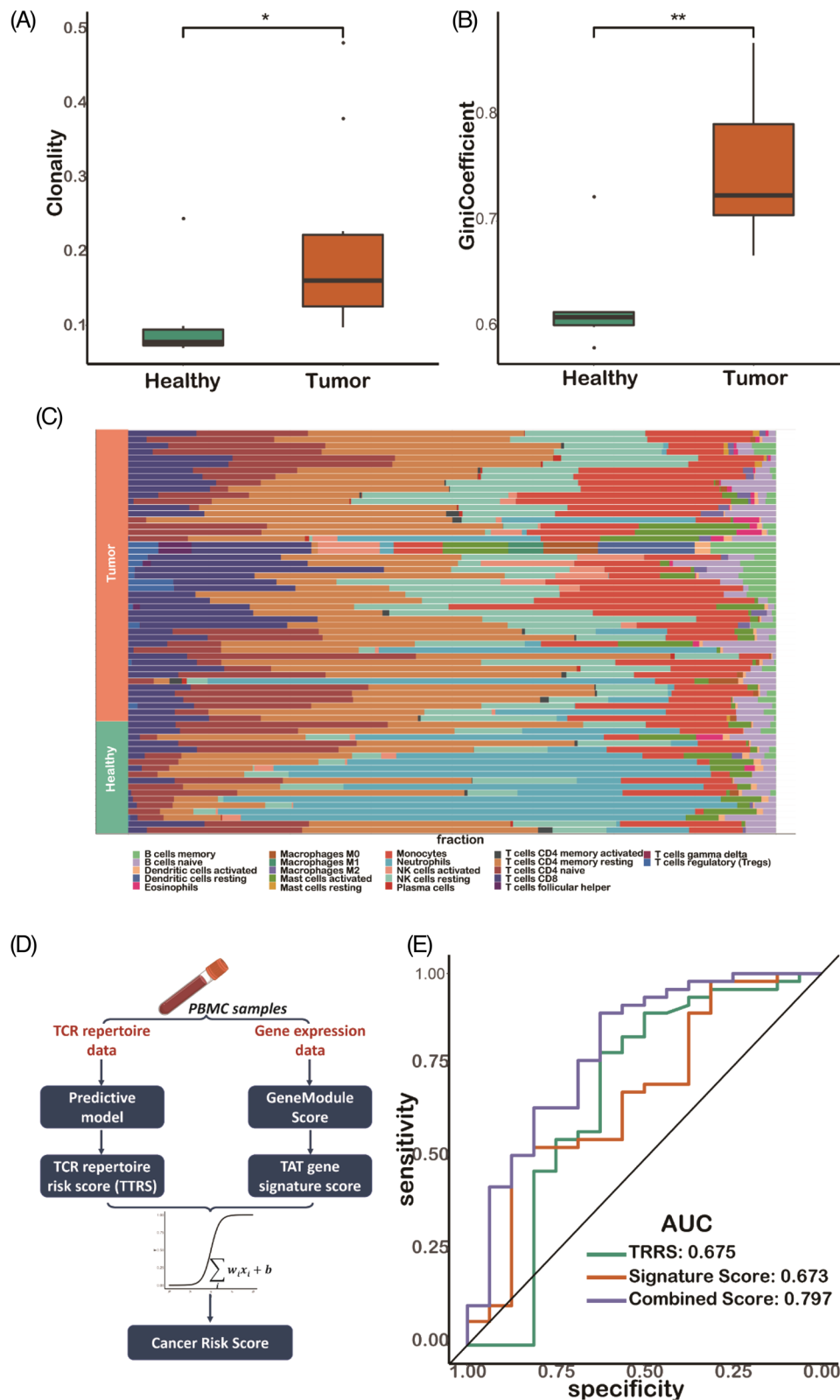


FIGURE 6 Combining both the T cell receptor repertoire (TCR) repertoire information and transcriptional signatures improves the capacity of non-invasive tumour screening. (A–B) The indices of clonality and Gini coefficient among tumour and healthy samples in validation cohort 2 performed by bulk TCR sequencing. (C) Immune cell subtypes deconvoluted from bulk RNA-seq data among PBMCs samples from tumour patients and healthy donors. (D) The schematic framework of combining TCR repertoire risk score (TTRS) and tumour-associated T cells (TATs) signature gene score for tumour screening. (E) The ROC shows the model performance of using TCR repertoire risk score (TTRS) alone, using TAT signature score alone and using both of them for tumour screening

tumour somatic mutations and are highly heterogeneous among and within cancer types. Mutations called from plasma cfDNA often come from white blood cells rather than tumour cells.⁸⁷ Besides, tumour-specific ctDNA and CTC are expected to be rare in the bloodstream, and capturing the tumour-specific signals often needs to perform ultra-deep next-generation sequencing.⁸⁸

Our method provides a feasibility of tumour screening from the view of tumour-associated T cells that are antigen-independent, shifting the diagnostic paradigm from tumour-modality-driven to immune-response-driven.⁸⁹ Since immune responses are usually ahead of any measurable symptoms, this framework may have the potential to detect diseases at early stages. The melanoma cohort⁶⁰ in our study consists of tumour samples of early stage (Stage I, Table S1), indicating the potential capability of our model for identifying early stage tumour patients. Besides, our framework can serve as a cheap, non-redundant and complementary diagnostic paradigm for tumour screening since we use information of both DNA (for TCR repertoire quantification) and RNA (or protein, for immunophenotype quantification) level from peripheral blood. Last but not the least, the framework we provided is not intended to replace the current diagnostic paradigms, and we think it can serve as a complementary method with existing modalities.

There are limitations of our study, which should be considered. Our binary predictive model based on TCR sequences from different patients and cancer types does not currently address HLA restriction information. HLA restriction information is very important to the TCR repertoire distribution in patients since it is related to the pattern of the antigen peptides presented. We believe that, when it is accessible, this information can be incorporated into the model, and adding it will improve model performance. Leveraging deep learning architecture allows us to incorporate many types of features to predict TATs. However, we realize that transcriptional signatures are essential but not determinant factors for distinguishing tumours from healthy samples. The immune reaction induced by TSAs may be similar to that included by common pathogen infections. In our study, we have mined publicly available databases of virus/bacteria TCR sequences and removed all the T cell clones overlapped with the sequences in these databases before developing the model or performing differential gene expression analysis, alleviating this problem to a great extent. Notably, the starting material and library construction methods for TCR sequencing can profoundly influence the TCR repertoire detected. Using gDNA as the starting material for TCR sequencing leads to more stable results and allows for better quantification of each single TCR clone, while employing RNA potentially provides

information on expression levels. Therefore, consolidating TCR sequencing data from different starting materials and library construction warrants further efforts.

ACKNOWLEDGEMENTS

The authors would like to thank the ImmuneAccess, TCGA and GEO databases for the availability of the data. This work was supported by the Tsinghua University-Peking University Jointed Center for Life Science (grant number: 61020100119 to XL) and National Thousand Young Talents Program of China (grant number: 042021011 to XL). They thank American Journal Experts (AJE) for English language editing. We thank the staff members of LC-BIO TECHNOLOGIES (HANGZHOU) CO., LTD. for their support and expertise in NGS for expe.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

Fansen Ji^{1,2} 

Lin Chen^{2,3}

Zhizhuo Chen⁴

Bin Luo³

Yongwang Wang⁵

Xun Lan^{1,2}

¹*Tsinghua-Peking Center for Life Sciences, MOE Key Laboratory of Tsinghua University, Beijing, China*

²*School of Medicine, Tsinghua University, Beijing, China*

³*General Surgery Department, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing, China*

⁴*School of Life Science, Tsinghua University, Beijing, China*

⁵*Department of Anesthesiology, Affiliated Hospital of Guilin Medical University, Guilin, China*

Correspondence

Xun Lan, School of Medicine, Tsinghua University, Haidian District, Beijing 100084, China.

Email: xlan@tsinghua.edu.cn

Bin Luo, Beijing Tsinghua Changgung Hospital, Room 431, Building 3, 168 Litang Road, Changping District, Beijing 102218, China.

Email: luobin@mail.tsinghua.edu.cn

Yongwang Wang, Department of Anesthesiology, Affiliated Hospital of Guilin Medical University, No.15 Lequn Road, Xiufeng District, Guilin 541001, China.

Email: wangyongwang81@126.com

Fansen Ji and Lin Chen contributed equally to this study.

ORCID

Fansen Ji  <https://orcid.org/0000-0002-4220-0786>

REFERENCES

1. Tempero MA, Uchida E, Takasaki H, et al. Relationship of carbohydrate antigen 19-9 and Lewis antigens in pancreatic cancer. *Cancer Res.* 1987;47:5501-5503.
2. Mukama T, Fortner RT, Katzke V, et al. Prospective evaluation of 92 serum protein biomarkers for early detection of ovarian cancer. *Br J Cancer.* 2022;1-9. <https://doi.org/10.1038/s41416-021-01697-z>
3. Underwood JJ, Quadri RS, Kalva SP, et al. Liquid biopsy for cancer: review and implications for the radiologist. *Radiology.* 2020;294:5-17.
4. Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating liquid biopsies into the management of cancer. *Nat Rev Clin Oncol.* 2017;14:531-548.
5. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol.* 2013;10:472-484.
6. Ignatiadis M, Sledge GW, Jeffrey SS. Liquid biopsy enters the clinic - implementation issues and future challenges. *Nat Rev Clin Oncol.* 2021;18:297-312.
7. Luo H, Wei W, Ye Z, et al. Liquid biopsy of methylation biomarkers in cell-free DNA. *Trends Mol Med.* 2021;27:482-500.
8. Liu MC, Oxnard GR, Klein EA, et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol.* 2020;31:745-759.
9. Cescon DW, Bratman SV, Chan SM, Siu LL. Circulating tumor DNA and liquid biopsy in oncology. *Nat Cancer.* 2020;1:276-290.
10. Ye Q, Ling S, Zheng S, Xu X. Liquid biopsy in hepatocellular carcinoma: circulating tumor cells and circulating tumor DNA. *Mol Cancer.* 2019;18:114.
11. Alix-Panabières C, Pantel K. Circulating tumor cells: liquid biopsy of cancer. *Clin Chem.* 2013;59:110-118.
12. de Miguel Pérez D, Rodríguez Martínez A, Ortigosa Palomo A, et al. Extracellular vesicle-miRNAs as liquid biopsy biomarkers for disease identification and prognosis in metastatic colorectal cancer patients. *Sci Rep.* 2020;10:3974.
13. Giannopoulou L, Zavridou M, Kasimir-Bauer S, Lianidou ES. Liquid biopsy in ovarian cancer: the potential of circulating miRNAs and exosomes. *Transl Res J Lab Clin Med.* 2019;205:77-91.
14. Zhou B, Xu K, Zheng X, et al. Application of exosomes as liquid biopsy in clinical diagnosis. *Signal Transduct Target Ther.* 2020;5:144.
15. Li Z, Wang Z, Tang Y, et al. Liquid biopsy-based single-cell metabolic phenotyping of lung cancer patients for informative diagnostics. *Nat Commun.* 2019;10:3856.
16. Oettgen HF. Immunology in the diagnosis of cancer. *Internist.* 1975;16:199-205.
17. Herberman RB. Tumor immunology and its application to the diagnosis of cancer. *Ann Clin Lab Sci.* 1974;4:323-330.
18. Rosenberg SA, Dudley ME. Adoptive cell therapy for the treatment of patients with metastatic melanoma. *Curr Opin Immunol.* 2009;21:233-240.
19. Caushi JX, Zhang J, Ji Z, et al. Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. *Nature.* 2021;596:126-132.
20. Scheper W, Kelderman S, Fanchi LF, et al. Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. *Nat Med.* 2018;25:89-94.
21. Simoni Y, Becht E, Fehlings M, et al. Bystander CD8 + T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature.* 2018;557:575-579.
22. Gros A, Parkhurst MR, Tran E, et al. Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat Med.* 2016;22:433-438.
23. Parkhurst M, Gros A, Pasetto A, et al. Isolation of T-Cell receptors specifically reactive with mutated tumor-associated antigens from tumor-infiltrating lymphocytes based on CD137 expression. *Clin Cancer Res.* 2017;23:2491-2505.
24. Han A, Glanville J, Hansmann L, Davis MM. Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat Biotechnol.* 2014;32:684-692.
25. Oliveira G, Stromhaug K, Klaefer S, et al. Phenotype, specificity and avidity of antitumour CD8+ T cells in melanoma. *Nature.* 2021;596:119-125.
26. Kamphorst AO, Pillai RN, Yang S, et al. Proliferation of PD-1+ CD8 T cells in peripheral blood after PD-1-targeted therapy in lung cancer patients. *Proc Natl Acad Sci.* 2017;114:4993-4998.
27. Fairfax BP, Taylor CA, Watson RA, et al. Peripheral CD8 + T cell characteristics associated with durable responses to immune checkpoint blockade in patients with metastatic melanoma. *Nat Med.* 2020;26:193-199.
28. Duraiswamy J, Ibegbu CC, Masopust D, et al. Phenotype, function, and gene expression profiles of programmed death-1hi CD8 T cells in healthy human adults. *J Immunol.* 2011;186:4200-4212.
29. Sharpe AH, Pauken KE. The diverse functions of the PD1 inhibitory pathway. *Nat Rev Immunol.* 2018;18:153-167.
30. Munson DJ, Egelston CA, Chiotti KE, et al. Identification of shared TCR sequences from T cells in human breast cancer using emulsion RT-PCR. *Proc Natl Acad Sci.* 2016;113:8272-8277.
31. Robins HS, Campregher PV, Srivastava SK, et al. Comprehensive assessment of T-cell receptor β -chain diversity in $\alpha\beta$ T cells. *Blood.* 2009;114:4099-4107.
32. Peng S, Zaretsky JM, Ng AHC, et al. Sensitive detection and analysis of neoantigen-specific T cell populations from tumors and blood. *Cell Rep.* 2019;28:2728-2738.
33. Le DT, Durham JN, Smith KN, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science.* 2017;357:409-413.
34. Creelan BC, Wang C, Teer JK, et al. Tumor-infiltrating lymphocyte treatment for anti-PD-1-resistant metastatic lung cancer: a phase 1 trial. *Nat Med.* 2021;27:1410-1418.
35. Yao Y, Zia A, Neumann RS, et al. T cell receptor repertoire as a potential diagnostic marker for celiac disease. *Clin Immunol.* 2021;222:108621.
36. Emerson RO, DeWitt WS, Vignali M, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet.* 2017;49:659-665.
37. Schultheiß C, Paschold L, Simnica D, et al. Next-generation sequencing of T and B cell receptor repertoires from COVID-19 patients showed signatures associated with severity of disease. *Immunity.* 2020;53:442-455.
38. Ostmeier J, Christley S, Toby IT, Cowell LG. Biophysicochemical motifs in T-cell receptor sequences distinguish repertoires

- from tumor-infiltrating lymphocyte and adjacent healthy tissue. *Cancer Res.* 2019;79:1671-1680.
39. Kirsch IR, Watanabe R, O'Malley JT, et al. TCR sequencing facilitates diagnosis and identifies mature T cells as the cell of origin in CTCL. *Sci Transl Med.* 2015;7:308ra158-308ra158.
 40. Li M, Zhang C, Deng S, et al. Lung cancer-associated T cell repertoire as potential biomarker for early detection of stage I lung cancer. *Lung Cancer Amst Neth.* 2021;162:16-22.
 41. Li N, Yuan J, Tian W, et al. T-cell receptor repertoire analysis for the diagnosis and treatment of solid tumor: a methodology and clinical applications. *Cancer Commun.* 2020;40:473-483.
 42. Beausang JF, Wheeler AJ, Chan NH, et al. T cell receptor sequencing of early-stage breast cancer tumors identifies altered clonal structure of the T cell repertoire. *Proc Natl Acad Sci.* 2017;114:E10409-E10417.
 43. Li B, Li T, Pignon J-C, et al. Landscape of tumor-infiltrating T cell repertoire of human cancers. *Nat Genet.* 2016;48:725-732.
 44. Han J, Khatwani N, Searles TG, et al. Memory CD8+ T cell responses to cancer. *Semin Immunol.* 2020;49:101435.
 45. Masopust D, Schenkel JM. The integration of T cell migration, differentiation and function. *Nat Rev Immunol.* 2013;13:309-320.
 46. Gros A, Tran E, Parkhurst MR, et al. Recognition of human gastrointestinal cancer neoantigens by circulating PD-1+ lymphocytes. *J Clin Invest.* 2019;129:4992-5004.
 47. Aoki H, Shichino S, Matsushima K, Ueha S. Revealing clonal responses of tumor-reactive t-cells through T cell receptor repertoire analysis. *Front Immunol.* 2022;13:807696.
 48. Iwahori K, Shintani Y, Funaki S, et al. Peripheral T cell cytotoxicity predicts T cell function in the tumor microenvironment. *Sci Rep.* 2019;9:2636.
 49. Han J, Yu R, Duan J, et al. Weighting tumor-specific TCR repertoires as a classifier to stratify the immunotherapy delivery in non-small cell lung cancers. *Sci Adv.* 2021;7:eabd6971.
 50. Huang AC, Postow MA, Orlovski RJ, et al. T-cell invigoration to tumour burden ratio associated with anti-PD-1 response. *Nature.* 2017;545:60-65.
 51. Zippelius A, Batard P, Rubio-Godoy V, et al. Effector function of human tumor-specific CD8 T cells in melanoma lesions: a state of local functional tolerance. *Cancer Res.* 2004;64:2865-2873.
 52. du Bois H, Heim TA, Lund AW. Tumor-draining lymph nodes: at the crossroads of metastasis and immunity. *Sci Immunol.* 2021;6:eabg3551.
 53. Poran A, Scherer J, Bushway ME, et al. Combined TCR repertoire profiles and blood cell phenotypes predict melanoma patient response to personalized neoantigen therapy plus Anti-PD-1. *Cell Rep Med.* 2020;1:100141.
 54. Zhang L, Yu X, Zheng L, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature.* 2018;564:268-272.
 55. Hsu M, Sedighim S, Wang T, et al. TCR sequencing can identify and track glioma-infiltrating T cells after DC vaccination. *Cancer Immunol Res.* 2016;4:412-418.
 56. Lu T, Zhang Z, Zhu J, et al. Deep learning-based prediction of the T cell receptor-antigen binding specificity. *Nat Mach Intell.* 2021;3:864-875.
 57. Sidhom J-W, Larman HB, Pardoll DM, Baras AS. DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nat Commun.* 2021;12:1605.
 58. Jokinen E, Heinonen M, Huuhtanen J, et al. TCRGP: determining epitope specificity of T cell receptors. *bioRxiv.* 2019. <https://doi.org/10.1101/542332>
 59. Formenti SC, Rudqvist N-P, Golden E, et al. Radiotherapy induces responses of lung cancer to CTLA-4 blockade. *Nat Med.* 2018;24:1845.
 60. Robert L, Tsoi J, Wang X, et al. CTLA4 blockade broadens the peripheral T-cell receptor repertoire. *Clin Cancer Res.* 2014;20:2424-2432.
 61. Yost KE, Satpathy AT, Wells DK, et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat Med.* 2019;25:1251-1259.
 62. Beshnova D, Ye J, Onabolu O, et al. De novo prediction of cancer-associated T cell receptors for noninvasive cancer detection. *Sci Transl Med.* 2020;12:eaaz3738. <https://doi.org/10.1126/scitranslmed.aaz3738>
 63. Shitara K, Ueha S, Shichino S, et al. First-in-human phase 1 study of IT1208, a defucosylated humanized anti-CD4 depleting antibody, in patients with advanced solid tumors. *J Immunother Cancer.* 2019;7:195.
 64. Platten M, Bunse L, Wick A, et al. A vaccine targeting mutant IDH1 in newly diagnosed glioma. *Nature.* 2021;592:463-468.
 65. Gueguen P, Metoikidou C, Dupic T, et al. Contribution of resident and circulating precursors to tumor-infiltrating CD8+ T cell populations in lung cancer. *Sci Immunol.* 2021;6:eabd5778. <https://doi.org/10.1126/sciimmunol.abd5778>
 66. Guo X, Zhang Y, Zheng L, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med.* 2018;24:978-985.
 67. Borcherdig N, Vishwakarma A, Voigt AP, et al. Mapping the immune environment in clear cell renal carcinoma by single-cell genomics. *Commun Biol.* 2021;4:1-11.
 68. Pauken KE, Shahid O, Lagattuta KA, et al. Single-cell analyses identify circulating anti-tumor CD8 T cells and markers for their enrichment. *J Exp Med.* 2021;218:e20200920.
 69. Zhang Y, Chen H, Mo H, et al. Single-cell analyses reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative breast cancer. *Cancer Cell.* 2021;39:1578-1593.e8.
 70. Scherrenburg J, Piriou ERWAN, Nanlohy NM, van Baarle D. Detailed analysis of Epstein-Barr virus-specific CD4+ and CD8+ T cell responses during infectious mononucleosis. *Clin Exp Immunol.* 2008;153:231-239.
 71. Miles DJC, Sanneh M, Holder B, et al. Cytomegalovirus infection induces T-cell differentiation without impairing antigen-specific responses in Gambian infants. *Immunology.* 2008;124:388-400.
 72. Strindhall J, Ernerudh J, Mörner A, et al. Humoral response to influenza vaccination in relation to pre-vaccination antibody titres, vaccination history, cytomegalovirus serostatus and CD4/CD8 ratio. *Infect Dis.* 2016;48:436-442.
 73. Zhang Z, Xiong D, Wang X, et al. Mapping the functional landscape of T cell receptor repertoires by single-T cell transcriptomics. *Nat Methods.* 2021;18:92-99.
 74. Stuart T, Butler A, Hoffman P, et al. Comprehensive integration of single-cell data. *Cell.* 2019;177:1888-1902.e21.
 75. Krishna C, DiNatale RG, Kuo F, et al. Single-cell sequencing links multiregional immune landscapes and tissue-resident T

- cells in ccRCC to tumor topology and therapy efficacy. *Cancer Cell*. 2021;39:662-677.
76. Reina-Campos M, Scharping NE, Goldrath AW. CD8+ T cell metabolism in infection and cancer. *Nat Rev Immunol*. 2021;21:718-738.
 77. Berdiaki A, Neagu M, Giatagana E-M, et al. Glycosaminoglycans: carriers and targets for tailored anti-cancer therapy. *Biomolecules*. 2021;11:395.
 78. Wrenshall LE, Platt JL. Regulation of T cell homeostasis by heparan sulfate-bound IL-2. *J Immunol*. 1999;163:3793-3800.
 79. Shinohara Y, Tsukimoto M. Adenine nucleotides attenuate murine T cell activation induced by concanavalin A or T cell receptor stimulation. *Front Pharmacol*. 2017;8:986.
 80. de Oliveira Bravo M, Carvalho JL, Saldanha-Araujo F. Adenosine production: a common path for mesenchymal stem-cell and regulatory T-cell-mediated immunosuppression. *Purinergic Signal*. 2016;12:595-609.
 81. Deaglio S, Dwyer KM, Gao W, et al. Adenosine generation catalyzed by CD39 and CD73 expressed on regulatory T cells mediates immune suppression. *J Exp Med*. 2007;204:1257-1265.
 82. Wang S, He Z, Wang X, et al. Antigen presentation and tumor immunogenicity in cancer immunotherapy response prediction. *eLife*. 2019;8:e49020.
 83. Zeng D, Ye Z, Shen R, et al. IOBR: multi-omics immunology biological research to decode tumor microenvironment and signatures. *Front Immunol*. 2021;12:2547.
 84. Fass L. Imaging and cancer: a review. *Mol Oncol*. 2008;2:115-152.
 85. Hackshaw A, Clarke CA, Hartman A-R. New genomic technologies for multi-cancer early detection: rethinking the scope of cancer screening. *Cancer Cell*. 2022;40:109-113.
 86. Patz EF, Pinsky P, Gatsonis C, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med*. 2014;174:269-274.
 87. Razavi P, Li BT, Brown DN, et al. High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. *Nat Med*. 2019;25:1928-1937.
 88. Freitas C, Sousa C, Machado F, et al. The role of liquid biopsy in early diagnosis of lung cancer. *Front Oncol*. 2021;11:634316.
 89. Zhang H, Zhan X, Li B. GIANA allows computationally-efficient TCR clustering and multi-disease repertoire classification by isometric transformation. *Nat Commun*. 2021;12:4699.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.