RESEARCH ARTICLE

Comprehensive characterization of the molecular feature of T cells in laryngeal cancer: evidence from integrated single-cell and bulk RNA sequencing data using multiple machine learning approaches

Jie Cui[a], Yangpeng Ou[b], Kai Yue[a], Yansheng Wu[a], Yuansheng Duan[a], Genglong Liu[c,d], Zhen Chen[e], Minghui Wei[f] and Xudong Wang[a]

[a]Department of Maxillofacial and Otorhinolaryngological Oncology, Tianjin Medical University Cancer Institute and Hospital, Key Laboratory of Basic and Translational Medicine on Head & Neck Cancer, Tianjin, Key Laboratory of Cancer Prevention and Therapy, Tianjin Cancer Institute, National Clinical Research Center of Cancer, Tianjin, PR China; [b]Department of Oncology, Huizhou Third People's Hospital, Guangzhou Medical University, Huizhou, PR China; [c]School of Medicine, Southern Medical University, Foshan, PR China; [d]Editor Office, iMeta, Shenzhen, PR China; [e]Department of Intensive Care Unit, Shunde Hospital, Southern Medical University (the First People's Hospital of Shunde), Foshan, PR China; [f]Department of Head and Neck Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, PR China

ABSTRACT

Background: The clinical relevance of T cell-related molecules at single-cell resolution in laryngeal cancer (LC) has not been clarified.

Materials and methods: Three LC tissues and matching adjoining normal tissues from the hospital were used to perform 10X single-cell RNA sequencing. Hub T cell-related genes (TCRGs) were detected by applying ten machine learning (ML) techniques based TCGA and GEO databases, which were also utilized to create a prediction model (TCRG classifier) and a multicenter validation model. Lastly, we conducted a comprehensive analysis of the TCRG's correlation with immunological properties.

Results: The analysis of single-cell RNA-seq data revealed that T cells are the primary components of the tumor microenvironment (TME), are significantly involved in cell differentiation pathways, and play a considerable role in intercellular communication. Based on 10 ML approaches, TCRG classifier were identified to develop and validate. The TCRG classifier exhibited excellent prognostic values with a mean C-index of 0.66 in six cohorts, serving as an independent risk factor ($p < 0.01$). Additionally, the TCRG exhibited a significant relationship with immune score, immune cell infiltration, immune-associated pathways, immune checkpoint inhibitors, human leukocyte antigen, and immunogenicity. Lastly, IPS, TCIC, TIDE, and IMvigor210 cohort analysis illustrated that the immunotherapy response may be accurately predicted using TCRG.

Conclusion: A TCRG classifier is an excellent resource for predicting a patient's prognosis, potentially guiding the preservation of laryngeal function, and identifying patients who may have a positive response to immunotherapy, which might have profound effects on therapeutic practice.

Introduction

Laryngeal cancer (LC) is a prevalent malignancy that affects the head and neck, with squamous cell carcinoma representing approximately 90% of all LC cases [1]. Based on data from the American Cancer Society, the estimated annual incidence and mortality rates for the disease were 4.0 per 100,000 cases (13,150 new diagnoses) and 1.1 per 100,000 fatalities (3700 deaths), respectively [2]. Due to the imperceptibility of the initial symptoms of LC, over 50% of the patients happen to be in the intermediate or late stages of tumour progression with significant infiltration or lymphatic metastasis when they are diagnosed [3]. Currently, there is a lack of approved targeted therapies for LC, except for cetuximab in 2006 and immune checkpoint inhibitors (ICIs) in 2016[4]. Response rates of cetuximab are moderate, ranging from 10% to 15% when used as a monotherapy, and no established biological markers of response have been identified [5]. Recently, immunotherapy, specifically ICIs, has made significant strides in treating cancer and has progressively

emerged as the primary approach for managing solid tumours; this strategy was also implemented in the LC clinical trial [6]. Patients' responses to ICI therapy were regrettably hindered as a result of the complexity of the tumour microenvironment (TME) and intra-tumour heterogeneity. Recent reports indicate that patients with LC continue to have a dire prognosis for long-term survival, with five-year mortality rates ranging between 40% and 50% [7]. Therefore, to treat and manage LC, it is necessary to develop more precise biological markers and models that consider the TME and address tumour heterogeneity.

The TME refers to a complex, heterogeneous composition of infiltrating immune and resident host cells (T cells), secreted factors and extracellular matrix [8]. Numerous findings since the TME's inception point to its pivotal involvement in the formation, advancement and therapy of tumours [8,9]. T cells, as an essential part of the TME, can recognize the specific antigens and morph into diverse effector T cells to participate in various immune activities. Immune mechanisms based on T cells in the TME have produced different forms of immunotherapy by enhancing immune system responses (especially T cells) and suppressing ICIs [10]. Given the crucial function of T cells in the TME, several recent research works have recognized and validated the use of T cell marker gene signature as a predictor of prognosis and effectiveness of immunotherapy in lung [11], breast [12] and colorectal [13] cancers. There are differences in TME among different cancer types, even among different patients with the same cancer, and the function of the majority of T cell-related molecules in the TME is still not well understood in LC at present. Thus, it is imperative to research unique T cell-related molecules of TME in LC to identify novel biomarkers/models for targeted therapeutics, improve prognostic stratification and screen patients who will respond to these immunotherapies.

Additionally, due to the limitations of biotechnology approaches, it is challenging to adequately measure intra- and inter-tumour heterogeneity [14]. Intriguingly, researchers may examine the molecular constituents of the TME and objectively evaluate diverse immune cells at the cellular level *via* single-cell RNA sequencing (RNA-seq) [15]. Currently, previous studies have not examined any particular molecular characteristic of T cells in the LC TME on a single-cell basis. Comprehending changes in transcriptomes throughout the genome also makes use of bulk RNA-seq, another traditional population-based RNA-seq method [16]. The organic combination of single-cell and bulk RNA-seq has not been applied in T-cell-related molecules in LC.

Importantly, improving the prognosis and immunotherapy effectiveness of LC patients requires combining novel biomarkers using modern machine learning (ML) algorithms, which is especially important in this age of precision medicine.

Here, firstly, we collected three LC samples and related adjoining normal samples in patients with LC who received curative-intend surgery to perform 10X Genomics single-cell RNA-seq accompanied by comprehensive bioinformatic analyses. We highlighted the role of T cells/T cells-related molecules in LC TME and identified potential therapeutic targets. Secondly, we systematically reviewed available bulk RNA-seq datasets in LC, combined single-cell and bulk RNA-seq to screen hub T cells-related genes (TCRGs) and then constructed and multicentre validated a TCRG classifier through the incorporation of ten ML algorithms. Thirdly, we comprehensively evaluated the predictive, prognostic performance and clinical value of the TCRG classifier in multiple transcriptome cohorts. Lastly, to assess the TCRG classifier's value in directing immunotherapy, we methodically correlated it with immunological properties from various perspectives.

Materials and methods

Tissue samples

A total of six samples (three tumour tissues and matching adjoining normal tissues) from three LC patients who received curative resection were collected at Tianjin Medical University Cancer Institute and Hospital. Informed was obtained from all patients and the study obtained approval from the Ethics Committee of Tianjin Medical University Cancer Institute and Hospital following a comprehensive evaluation of the research protocols based on human samples (No. Ek2022226). At least two pathologists with extensive training affirmed the pathological classification of each tissue, and the corresponding clinical features are detailed in Table 1. A comprehensive outline of the processes entailed in collecting and preparing samples for single-cell RNA-seq analysis is presented in Figure 1.

Preparation of single-cell suspensions

Once the tissue samples were surgically removed, they were immediately sectioned into smaller segments (3×3 mm) after immersion in a digestion solution containing Type II and Type IV collagenase (Gibco, Cat#17101015, 17104019), then placed in a 50 ml centrifuge tube with 5 ml of the digestion

solution and left to incubate in a water bath at 37 °C for half an hour. A pre-wet cell sieve (70 µM, BD, cat# 431751) was employed to filter the digestion fluid following digestion, and a fresh 50 ml centrifuge tube was used to collect the filtrate, which was centrifuged at 900 × g. The cells were resuspended by adding pre-cooled erythrocyte lysate (3 ml) to the cell pellet, which was subsequently left to incubate at the ambient temperature for 3 mins. To resuspend the cells and achieve a final concentration of over 10,000 cells/ml, an appropriate volume of DPBS containing 2% FBS (20 µl/ml) was introduced following a 10 min centrifugation (300 × g at ambient temperature). A single-cell solution was determined to be free of cell debris, particulate materials and small cell clusters (defined as having > 2 cells adhering to each particulate matter) to be eligible for further single-cell RNA-seq investigations, and the percentage of viable cells was more than 80%.

Preparing and sequencing single-cell libraries

The single-cell RNA-seq experiment was conducted utilizing the 10X Genomics Chromium Single-Cell Platform. Subsequently, the Chromium Single Cell 30 Library and Gel Bead Kit v2 (10X Genomics, PN-120237) and the Chromium TM Single Cell A Chip Kit (10X Genomics, PN-120236) were employed to develop single-cell RNA-seq libraries, in compliance with the instructions provided by the manufacturer. Thereafter, a Qubit 4 Fluorometer (Thermo Fisher, Cat#Q33238) was employed to measure the cDNA concentration, and an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) was utilized to assess the sizes of the fragments. All samples were sequenced at the multiplex.

Processing and quality control of single-cell data

The Seurat R package was applied for processing single-cell RNA-seq data [17]. The original data contained 64,242 genes and 21,235 cells. The cells were filtered according to these parameters: (1) cells with < 200 and > 10,000 genes; (2) mitochondrial percentage per cell < 5. Subsequently, the dataset was normalized using the NormalizeData and ScaleData functions. FindVariableFeatures function screened the top 2000 highly variable genes (HVGs) into the principal component analysis (PCA) with the RunPCA function. The ElbowPlot function was employed to select significant PCs. The 'FindClusters' functions (resolution = 0.5) and 'FindNeighbors' (dims = 1:20) were utilized to cluster the cells. A dimensionality reduction analysis and cell cluster visualization were performed utilizing the uniform manifold approximation and projection (UMAP) technique.

Cell clustering analysis and identification of specific genes for cell subtypes

The singleR package (HumanPrimaryCellAtlasData), the CellMarker database (http://xteam.xbio.top/CellMarker/) and ScType (Fully-automated cell type identification of single-cell RNA-seq profiles, https://sctype.app/) was utilized to annotate cell clusters. Subsequently, the FindAllMarkers function was employed to screen specific genes per cell subset, setting parameters as logFC threshold = 0.5 and min.pct = 0.2.

Analyses of pseudotime trajectories and cell communication

To examine the associations of distinct clusters with cell types, pseudotime analysis was executed *via* the R package Monocle2 (v.2.28.0) with the parameters of mean expression > 0.1 and num_cells_expressed > = 10). Additional analysis utilizing the Monocle2 pseudotime function unveiled the critical role of T cell-specific genes in LC cellular differentiation. Furthermore, an analysis of intercellular communication networks was conducted utilizing CellChat (version 1.6.1) to gain a deeper understanding of the potential interactions that may occur between T cells and other cells in LC, based on receptor-ligand interactions. The *p* value associated with the probability of a specific cell type possessing a corresponding receptor-ligand complex is computed by employing 1000 permutation tests and the default ligand-receptor interaction data to ascertain the association of cell receptors with ligands.

Dataset collection, sample selection and preprocessing

LC RNA-seq data (FPKM values), somatic mutation data and corresponding clinicopathological characteristics OS (overall survival) and RFS (recurrence-free survival)) were sourced from The Cancer Genome Atlas (TCGA) database (https://gdc.cancer.gov/). Furthermore, from February 2001 through September 2023, the Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo) database was systematically searched to find LC-relevant transcriptome datasets. The transcriptomic profiling datasets needed to fulfil the following criteria: (a) organism: Homo sapiens; (b) expression profiling through high-throughput

Table 1. Summarized detailed basic information for patients included.

ID	Age (years)	Sex	Tumour site	Smoking	TNM stage	Histology grade	Haemoglobin (g/L)	Albumin (g/L)	Surgical procedure	Hospital LOS (days)	Principal diagnosis	Pathological diagnosis	Margin status
1831807	68	Male	Supraglottic	Yes	T3N0M0	Moderate	140	40.3	Total laryngectomy	21	LSCC	Moderately differentiated squamous cell carcinoma; Infiltration depth about 1 cm, cancer tissue invaded bilateral vocal, ventricular cords and epiglottis cartilage, no nerve invasion or intravascular cancer embolus was seen. IHC: PD-L1 (CPS: 5). Ki67 (75%+), p53 (50% moderate+, wild-type), p40 (+), p16 (+), EGFR (+). Mismatch repair proteins: PMS2 (+), MSH2 (+), MSH6 (+), MLHI (+)	Negative
1703556		Male	Glottic and Subglottic	Yes	T4N0M0	Moderate	136	43.5	Total laryngectomy	17	LSCC	Moderately differentiated squamous cell carcinoma; cancer tissue invaded peripheral rhabdomyosarcoma, epiglottis cartilage and peripheral minor salivary glands. IHC: p40(+), PD-L1(CPS :0),p53(1%+,wild-type),Ki-67(ho tspot 90%+),EGFR(++). Mismatch repair protein: PMS2(+)	Negative
1832687		Male	Supraglottic	Yes	T3N0M0	Well and moderate	113	44.1	Total laryngectomy	17	LSCC	Highly-moderately differentiated squamous cell carcinoma, with no clear vascular or neural invasion, invasion of epiglottis cartilage and no carcinoma invasion of the rest of the cartilage tissue. IHC: CK5/6(+), p40(+), Ki-67(30%+), p53(+),p16(++), CK(+), EGFR(+), PD-1(−), PD-1(CPS:30),C-erbB-2(−).	Negative

Note: LOS: length of stay; LSCC: laryngeal squamous cell carcinoma; IHC: immunohistochemistry; CPS: combined positive score.
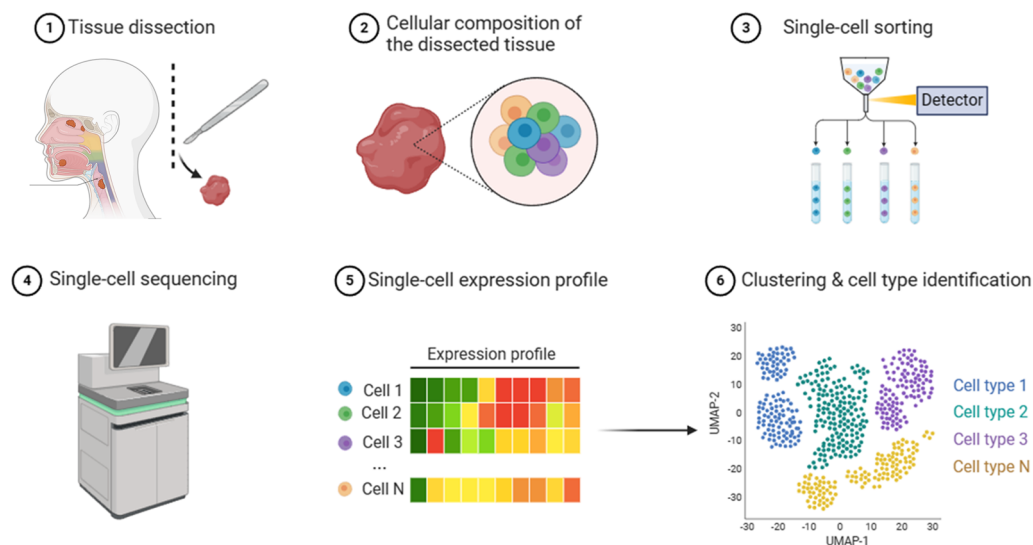
Figure 1. Flow chart described the present work. Samples are dissociated into single cells and captured in 10×genomic platform for library construction and RNA sequencing. The sequencing results were then undergoing bioinformatics analysis.

sequencing or array. Lastly, quantitative and qualitative analyses were conducted on nine GEO datasets (GSE59102, GSE84957, GSE117005, GSE127165, GSE142083, GSE143224, GSE25727, GSE27020 and GSE65858). The GSE25727 (DFS (disease-free survival)), GSE27020 (DFS) and GSE65858 (OS and RFS) datasets contain survival information. For GEO datasets, a robust multi-array average analysis was implemented, including quantile normalization, background correction and summarization.

Identification of T cell-specific genes and functional enrichment analysis

Standardized data including TCGA, GSE59102, GSE84957, GSE117005, GSE127165, GSE142083 and GSE143224 datasets between normal and LC samples were subjected to a variance analysis utilizing the NetworkAnalyst online-Gene Expression Table (https://www.networkanalyst.ca/). The overlapping differentially expressed genes (DEGs) were obtained from the intersection of seven datasets and visualized with the UpSetR package. Then, we take the intersection of common DEGs and T cell marker genes from single-cell analysis (FindAllMarkers), defined as T cell-specific genes in LC. The T cell-specific genes were subjected to GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analyses utilizing the clusterProfiler R package.

Identification and construction of TCRG classifier through multiple machine learning approaches

To construct a consensus prognosis model for LC, we performed the following workflow. (1) We began by combining 10 classical algorithms: survival support vector machine (Survival-SVM), random survival forest (RSF), supervised principal components (SuperPC), gradient boosting machine (GBM), least absolute shrinkage and selection operator (LASSO), ridge regression, elastic network (Enet), partial least squares regression for Cox (plsRcox), Stepwise Cox and CoxBoost. Among them, LASSO, RSF, Stepwise Cox and CoxBoost have feature selection capabilities. In addition, 101 algorithm combinations were constructed as prediction models utilizing the leave-one-out cross-validation (LOOCV) framework with 10-fold cross-validation. (2) Next, we utilized the TCGA with OS in LC as the training cohort and employed these 101 combinations to generate classifiers independently using T cell-specific genes. (3) Lastly, in the five testing cohorts (TCGA with RFS, GSE25727 with DSS, GSE27020 with DSS, GSE65858 with OS and GSE65858 with RFS), the TCRG score was calculated for each cohort by employing the model that was acquired in the training cohort. The best consensus prognostic model for LC was ultimately determined by averaging the C-indices of the six cohorts.

Risk stratification, prognostic value and clinical usefulness of TCRG classifier

Each case was assigned to a high or low TCRG group in the GEO and TCGA cohorts by calculating the median TCRG score. Heatmap based on the Chi-square test shows the relationship between the TCRG and clinicopathologic features (age, sex, histologic grade, perineural invasion, HPV status, T stage, TNM stage, among others) in GEO and TCGA datasets.

The log-rank test was employed in conjunction with Kaplan-Meier (KM) survival analysis to determine the variations in survival rates between the high TCRG and low TCRG groups. Prognostic performance was evaluated through the utilization of time-dependent receiver operating characteristic (ROC) curves. Additionally, univariate and multivariate Cox regression analyses were executed to ascertain the degree of independence of TCRG from other clinicopathological characteristics.

Moreover, a multiple ROC analysis was implemented to assess and contrast the TCRG's discriminatory capability with that of other clinical features (TNM stage). The clinical usefulness and net benefits of the TCRG were assessed via decision curve analysis (DCA), which was followed by a comparison with the clinical features of the TNM stage.

Association of TCRG with immune infiltration

We incorporated GEO datasets into subsequent immune microenvironment analyses to improve their statistical power. Various dataset's batch effects were modified utilizing ComBat from the 'sva' R package.

The 'ESTIMATE' package was utilized to calculate the immune, stromal, estimate and tumour purity scores and compare the differences between different TCRG subgroups [18]. CIBERSORTx and the single-sample gene set enrichment analysis (ssGSEA) algorithm were adopted to deduce the relative levels of infiltrating immunological cells [19]. An analysis was conducted to determine whether distinct TCRG subgroups have a distinct immune cell composition. In addition, Spearman correlation analyses were conducted to examine the association between TCRG classifier/hub TCRGs and immunological cells.

Comprehensive analysis of molecular characteristics in different TCRG subgroups

ssGSEA was performed on several representative gene sets (immune-related pathways) [20]. Then, the gene set variation analysis (GSVA) [21] was carried out to ascertain the signalling pathways based on the KEGG gene sets (c2.cp.kegg.v7.4.symbols.gmt) from the Molecular Signatures Database. We computed the differential enrichment score among various TCRG subgroups as part of the signalling pathway analysis. Afterward, a correlation analysis was executed to offer additional insights into the association of TCRG classifier/hub TCRGs with pivotal biological pathways.

TCRG correlation with immunological properties of the TME in LC

We gathered data on 25 ICI-related genes and 28 HLA-related genes to clarify the relationship between TCRG and critical immunologic molecules of the TME in LC [21]. The expression variation in genes linked to ICIs and HLA was determined across various subgroups of TCRG. The expression variations of genes associated with ICIs and HLA were compared across various TCRG subgroups. The association of TCRG with genes associated with ICIs and HLA was additionally examined via Spearman correlation analyses.

Additionally, the performance of TCRG in predicting possible immunotherapy responses was assessed using the following biological markers: neoantigen (the number of clonal, the number of subclonal, fraction subclonal, fraction subclonal neoantigens, the number of clonal neoantigens, the number of subclonal neoantigens, ploidy and purity, KS of neoantigens, area under curve (AUC) of neoantigens), tumour mutation burden (TMB) and microsatellite instability (MSI). The detailed neoantigen information of TCGA-derived LC patients were sourced from The Cancer Immunome Atlas (TCIA) (https://tcia.at/home). Previous studies [15] were screened to obtain TMB and MSI scores. The distinctions in the aforementioned biological markers between distinct TCRG subgroups are then identified. Finally, correlation analysis was conducted between the TCRG and the above critical biomarkers.

Role of TCRG classifier in immunotherapy

Sensitivity to CTLA4 and PD-1 inhibitors was predicted utilizing immunophenoscore (IPS). IPS data were compiled from the TCIA database and were utilized for comparing the IPS scores of various TCRG subgroups.

Additionally, we calculated the relative activity of the cancer immunity cycle (TCIC) via ssGSEA. Furthermore, we conducted a comparative analysis of the relative recruitment activities of CD8 T cells and T cells across various subgroups of the TCRG. Then, the immunotherapy response rate of patients across various TCRG

subgroups was analysed utilizing the TIDE (tumour immune dysfunction and exclusion) method.

Lastly, we obtained the IMvigor210 cohort which consisted of 348 cancer patients who received a PD-L1-targeting antibody. Thereafter, we validated the prognostic and immunotherapy efficacy prediction capabilities of the TCRG classifier in the IMvigor210 cohorts.

Statistical analysis

Both SPSS 22 and R (4.3.1) were employed for statistical analysis. The comparative analysis of continuous variables was conducted utilizing the Student's t-tests, Mann–Whitney, Wilcoxon, or Kruskal–Wallis tests with two-tailed analyses based on variable distribution. Categorical variables were tested using the Chi-square or Fisher's exact. $p < .05$ was determined to be the criterion for significance unless when specified otherwise.

Results

Single-cell atlas and cellular heterogeneity in LC

After quality control (Figure S1A and S1B), the single-cell transcriptomic data of 50,283 genes and 15,286 cells from 3 patients, including three cancer tissues and three adjoining normal tissues, were found. We identified 20 putative clusters based on the top 2000 HVGs (Figure S1C), using PCA (Figure 2(A,B), Figure S1D). Through the utilization of UMAP for dimensional reduction clustering, a total of 18 distinct subclusters were initially identified (Supplementary Material 1). Then, we used the SingleR, the CellMarker database and ScType tools to annotate subclusters and successfully identified 8 cell types in LC and normal tissues, including, T cells, granulocytes, epithelial cells, dendritic cells (DCs), cancer cells, fibroblast, natural killer (NK) cells, B cells (Figure 2(C)). Supplementary Material 1 summarized detailed annotation information for three different methods. The heatmap (Figure 2(D)) shows cellular variation in the expression of the top three genes.

Although nearly all cell types were presented in all samples, there was an uneven distribution of the proportion of each cell type among the specimens (Figure 2(E)). The findings illustrated that the levels of granulocytes and epithelial cells were quite low in LC tissues relative to normal tissues, while the composition of immune cells (especially T cells) was quite high in LC tissues, with cells exhibiting significant heterogeneity among various tissue types (Figure 2(E)). Interestingly, T cells are the most abundant of all cells in LC samples, which strongly motivated me to further explore the involvement of T cells

and their related molecules in LC TME. Furthermore, using the FindAllMarkers and Wilcoxon test, we identified cell type-specific DEGs (Supplementary Material 2), such as 2096T cell type-specific DEGs.

Pseudotime trajectory analysis and cell communication analysis

To ascertain how each cell type might differentiate in the TME, Monocle 2 was applied to the trajectories of all cell types. The results proved that epithelial cells were in their earliest stages of development (State 1), T cells and B cells were undergoing a transitional phase (State 2), and cancer cells and fibroblasts were in their ultimate stage (State 3) (Figure 3(A)). Additionally, the 'CellChat' algorithm was used to evaluate the weights and probabilities of intercellular signalling pathways in the TME of LC (Figure 3(B)). A complex intercellular communication network appeared to exist among all major cell types in the LC TME (Figure 3(B)). In the TME of LC, numerous ligand-receptor pairs were identified on malignant cells, DCs, T cells and fibroblasts, implying that any two of these distinct cell types could potentially interact (Figure 3(B,C)). T cells were able to influence epithelial cells, dendritic cells, cancer cells, fibroblast and NK cell functions through the IFNG, CCL5, ANXA1, NAMPT and SPP1 signalling pathways. These highlight that T cells play a considerable function in the intercellular communication of LC TME (Figure 3(D)).

Identification of T cell-specific genes and functional enrichment analysis

The following DEGs were detected utilizing the NetworkAnalyst online-Gene Expression Table: 2308 DEGs in GSE57065, 1125 in GSE84957, 1829 in GSE117005, 1741 in GSE127165, 1795 in GSE142083, 1512 in GSE143224 and 1512 in TCGA datasets, respectively (Supplementary Material 3).

We identified 346 DEGs that were shared by at least five results by analysing the intersection of the aforementioned datasets (Figure 4(A)). Then, we further take the intersection of the results of common 346 DEGs and T cell type-specific DEGs, and a total of 55 genes were determined to be T cell-specific genes in LC (Figure 4(B)). Subsequently, the GO term enrichment analysis for T cell-specific genes shows the top five significantly enriched clusters in cellular components (CCs), molecular functions (MFs) and biological processes (BPs) (Figure 4(C)), which include extracellular matrix (ECM) organization, collagen–containing ECM, external encapsulating structure organization,
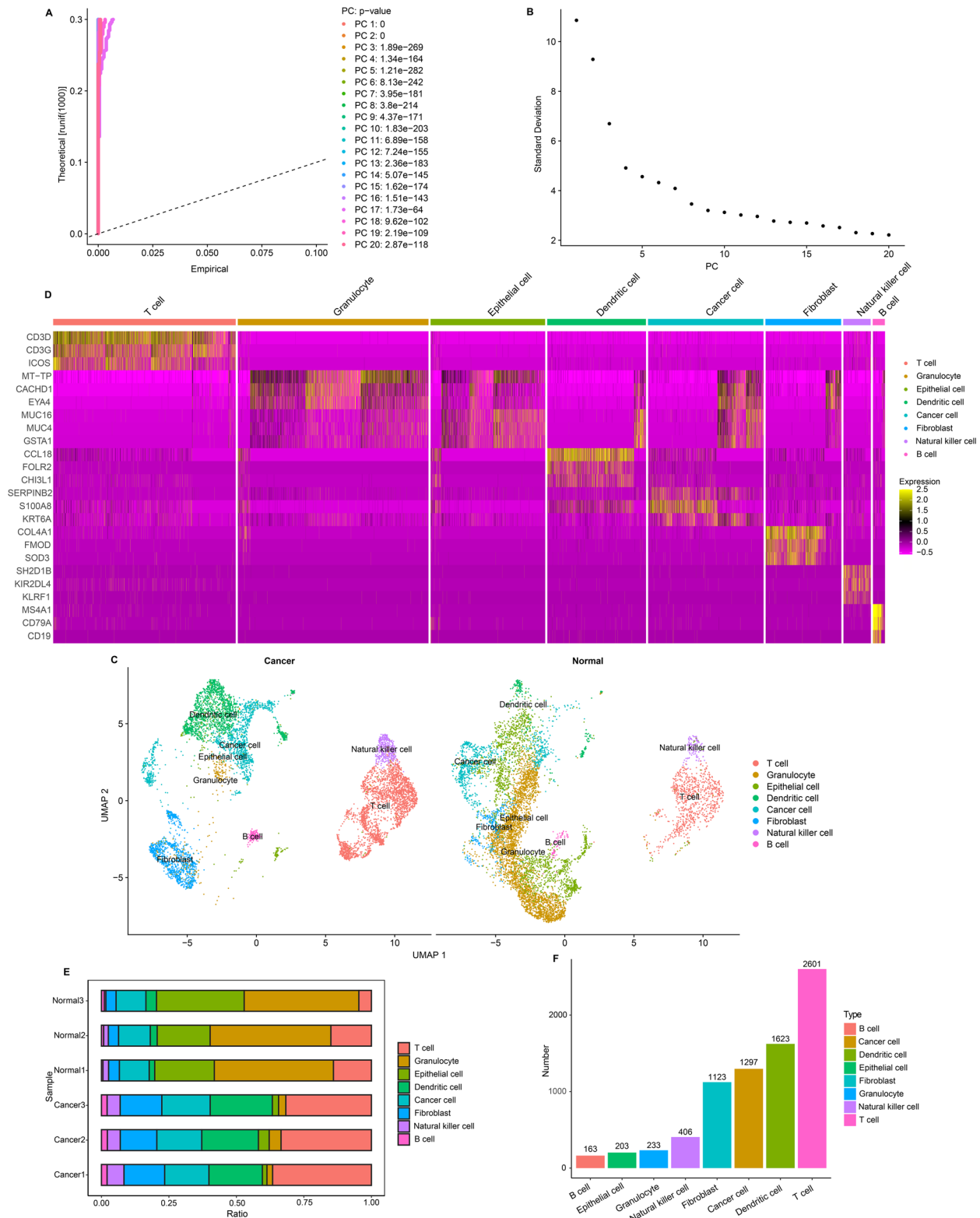
Figure 2. A single-cell atlas and cellular/transcriptional heterogeneity of tumour microenvironment (TME) in laryngeal cancer (LC). (a): Identification of 20 putative clusters based on the top highly variable genes (HVGs), using principal component analysis (PCA) with $p < .05$. (B): elbow plot shows how well each principal component explains the variance of the data and is ranked. (C): Heatmap plot showing representative marker genes across cell types. Colour intensity corresponds to the relative expression of specific genes. (D): uniform manifold approximation and projection (UMAP) plot of eight cell types annotated as T cells, granulocytes, epithelial cells, dendritic cells, cancer cells, fibroblast, natural killer cells, B cells in LC and normal tissues. (E): bar plot displaying the distribution of eight cell types in each LC and normal samples. (F): bar plot displaying and ranking the numbers of eight cell types in LC samples.
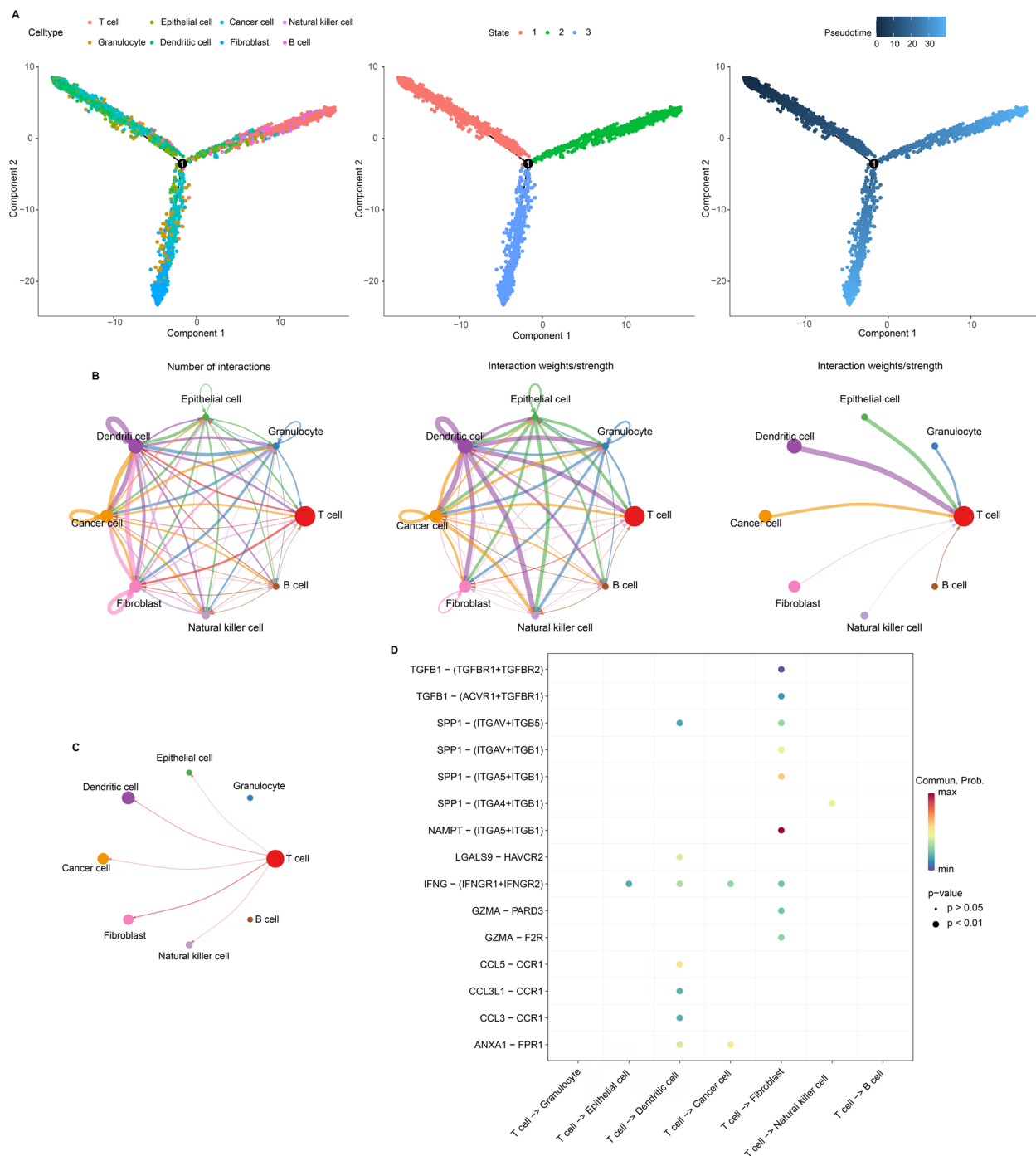
Figure 3. Analysis of the cellular developmental course and complex intercellular communication networks in LC, especially T cells. (A) The pseudo temporal curve and its development trajectory of the cell types are depicted in the pseudo temporal trajectory diagram of monocle 2. The development pathway of each cell type and the differentiation mechanism of cell types as inferred by monocle 2. (B): circle plots depicting the interaction numbers and interaction strength between T cells and other cells. The arrows indicate the direction of intercellular communication. (C) Schematic illustration of the cell-cell interaction networks (T cells act as outputs and other cells as receivers) in the LC TME. (D): plots showing selected interactions between the T cells and other cell types in the LC TME. The interactions are mediated through ligand-receptor pairs known to have immune functions. Dot colour reflects communication probabilities and dot size represents computed p values computed from the one-sided permutation test. Empty space means the communication probability is zero.

extracellular structure organization, ECM structural constituent and so on. As for the KEGG, they were primarily involved in proteoglycans in cancer, ECM−receptor interaction, AGE−RAGE signalling pathway, PI3K−Akt signalling pathway, focal adhesion, relaxin signalling pathway, etc. (Figure 4(D)).

Identification and construction of TCRG classifier through multiple machine learning approaches

To create a consistent prognostic model based on T cell-related molecular, we incorporated 55 identified genes into our integration program (LOOCV framework). We designed prediction models employing 54 algorithm combinations and 10-fold cross-validation in the TCGA OS training cohort. For the other five testing cohorts, we calculated the average C-index for each algorithm. The final model was determined to be the combination of StepCox[both] + Ridge, which exhibited the highest average C-index (0.660), as illustrated in Figure 5(A) (Supplementary Material 4). Ultimately, we constructed a prognostic model, named the TCRG classifier, based on eight hub TCRGs (CDK4, COL3A1,
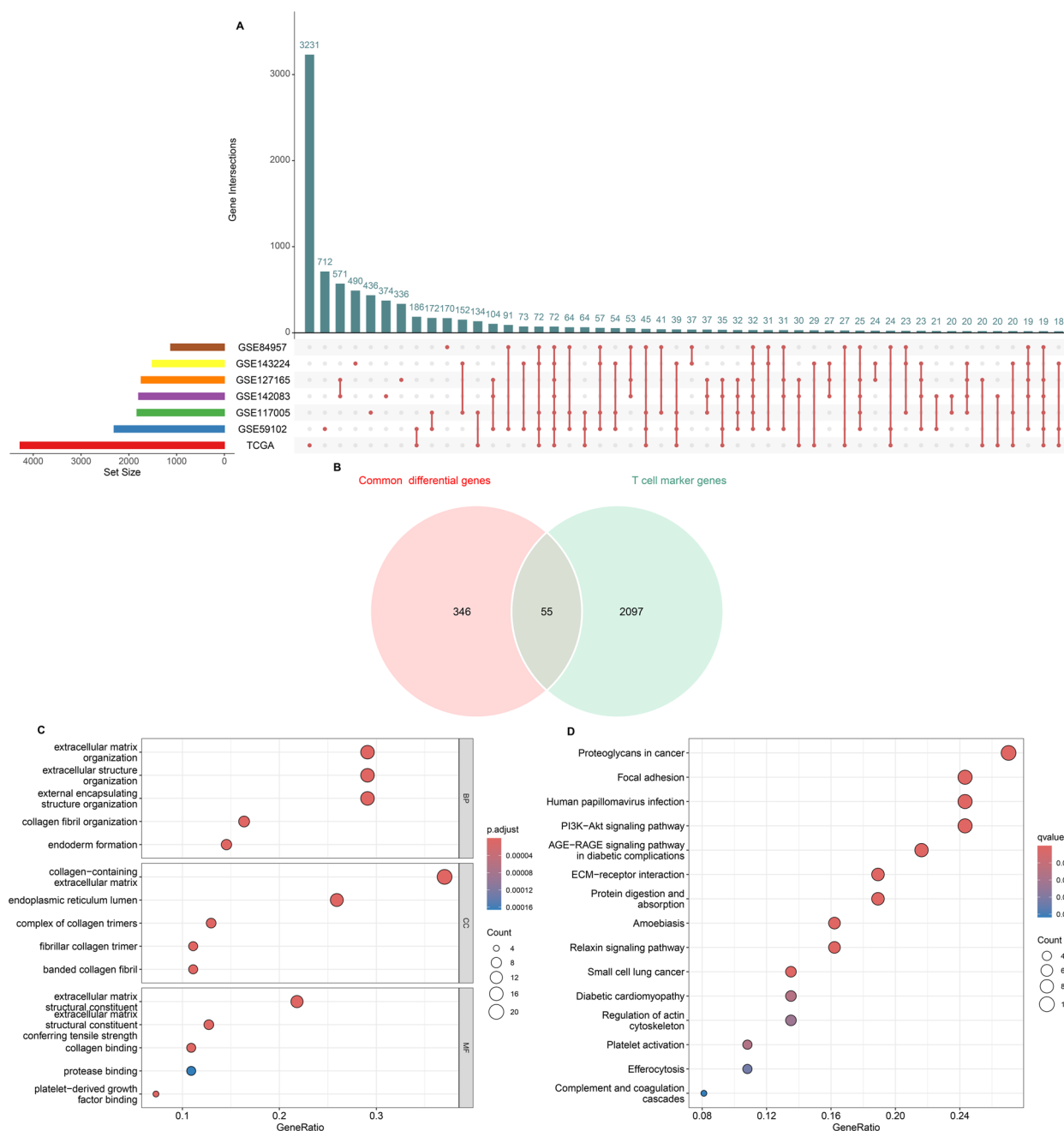

Figure 4. Identification of T cell-specific genes and functional enrichment analysis in LC. (A): UpSet plot presents the intersection of seven datasets to identify differentially expressed genes (DEGs) in LC. (B) Venn diagrams present the shared genes termed T cell-specific genes in common DEGs and T cell marker genes in LC. (C, D): functional enrichment analysis of T cell-specific genes in LC. (C): gene ontology (GO) analysis on the biological process (BP), cellular component (CC) and molecular function (MF). (D): Kyoto Encyclopedia of genes and Genomes (KEGG) pathways.

ITM2A, TREM2, SERPINH1, MMP12, MKNK2 and PIK3R1). In addition, TCRG scores were calculated for every sample across all six cohorts by using the expression profiles of eight genes that were incorporated in the TCRG classifier.

Risk stratification, prognostic value and clinical usefulness of TCRG classifier

The patients were categorized based on their TCRG scores, and the TCGA and GEO cohorts depict the landscape of associated clinical characteristics (Figure 5(B–G)). Notably, higher TCRG scores were shown to be significantly correlated to unfavourable prognostic information (OS, RFS, DFS) ($p < .05$), III-IV T stages ($p < .01$), more positive lymph node (LN) numbers ($p < .01$), higher LN ratio (LNR) ($p < .01$), positive margin status ($p < .01$).

Additionally, a superior predictive capability was demonstrated by time-dependent ROC curves for TCRG. The corresponding AUC values for the three- and five-year OS were 0.906 and 0.956, while they were 0.898 and 0.84 for RFS in the TCGA training cohort (Figure S2E and S2F). Additionally, the KM curves for OS and RFS revealed that among the TCGA training cohort, those with low TCRG had significantly extended survival times ($p < .001$ in OS and RFS, Figure S2K and S2L). In external GEO cohorts, comparable outcomes were observed (Figure S2).

Then, univariate and multivariate analyses demonstrated the independent role of TCRG in predicting unfavourable OS (HR: 1.16, 95% CI 1.10–1.22, $p = .001$) and RFS (HR: 1.03, 95% CI 1.02–1.04, $p < .001$) in TCGA dataset, DFS (HR: 1.30, 95% CI 1.07–1.57, $p < .001$) in GSE25727 dataset, DFS (HR: 1.59, 95% CI 1.34–1.89, $p = .001$) in GSE27020 dataset, OS (HR: 1.16, 95% CI 1.09–1.24, $p < .001$) and RFS (HR: 1.13, 95% CI 1.07–1.19, $p < .001$) in GSE65858 dataset, following control for additional clinical variables (Table S1 and Figure 5(H)).

Furthermore, the TCRG showed superior performance in the prediction of OS, RFS and DFS in comparison with TNM stage or other clinical characteristics (age, margin status and lymphovascular invasion) in TCGA and GEO cohorts (Figure S3A–S3F) (all $p < .05$). In particular, the DCA diagram shows that TCRG is superior to the TNM stage and other clinical variables (age, lymphovascular invasion and margin status) in TCGA and GEO cohorts as measured by the potential death threshold's continuity (x-axis) and the net benefit of employing the model for risk categorization (y-axis) (Figure S3G–S3L).

Analyses hub TCRGs in single-cell RNA-seq data

We used violin plots to visualize the expression of eight hub TCRGs in different cell types (Figure 6(A)). Consequently, we found that ITM2A is predominantly expressed in T cells, while four hub TCRGs are almost not expressed in T cells, such as SERPINH1 and CDK4 are primarily expressed in cancer-associated fibroblasts (CAF), and TREM2 is primarily expressed in DCs. Additionally, we investigated the effects of eight hub TCRGs in cell differentiation and found that they are all involved in the developmental trajectory of T cells (Figure 6(B)). COL3A1, ITM2A, SERPINH1, MMP12 and PIK3R1 exhibited a progressive increase in the later stage of the pseudo-time trajectory, while MMP12 and TREM2 gradually decreased.

Correlation of TCRG classifier/TCRGs with immunological infiltration

The variation in immune profile composition among various TCRG patterns in the LC TME was examined. ESTIMATE algorithm uncovered that the ESTIMATE score and immune score were substantially higher in the low TCRG subgroup than in the high TCRG subgroup, but tumour purity was significantly lower (Figure 7(A–D) and Figure S4A–S4D) in GEO and TCGA datasets. Then, the immune cell proportion in LC TME was assessed utilizing the CIBERSORTx and the ssGSEA algorithms. Furthermore, the CIBERSORTx analysis of GEO datasets showed that the low TCRG subgroup had lower levels of M0 macrophages ($p = .002$) but higher levels of activated NK cells ($p = .049$) and CD8 T cells ($p = .009$) in comparison with the high TCRG subgroup (Figure 7(E)). The ssGSEA results demonstrated that activated CD4 T cell ($p = .025$), activated CD8 T cell ($p = .013$), effector memory CD4 T cell ($p = .003$), effector memory CD8 T cell ($p = .012$) was considerably increased and macrophages ($p = .012$), Treg ($p = .007$), myeloid-derived suppressor cells (MDSC) ($p = .003$) were lower in the low TCRG subgroup as opposed to the high TCRG subgroup (Figure 7(F)). The TCGA datasets also showed similar results (Figure S4E and S4F).

Additionally, according to the results of the correlation analysis, TCRG was also favourably linked to macrophages ($p < .05$), Treg ($p < .01$) and MDSC ($p < .01$), but negatively correlated with activated CD4 T cell ($p < .001$) and activated CD8 T cell ($p < .01$) (Figure 7(G)) in GEO cohorts. Noticeably, hub TCRGs SERPINH1 was significantly positively associated with macrophages ($p < .001$), Treg ($p < .01$) and MDSC ($p < .01$), and ITM2A was considerably positively linked to CD4 T cell
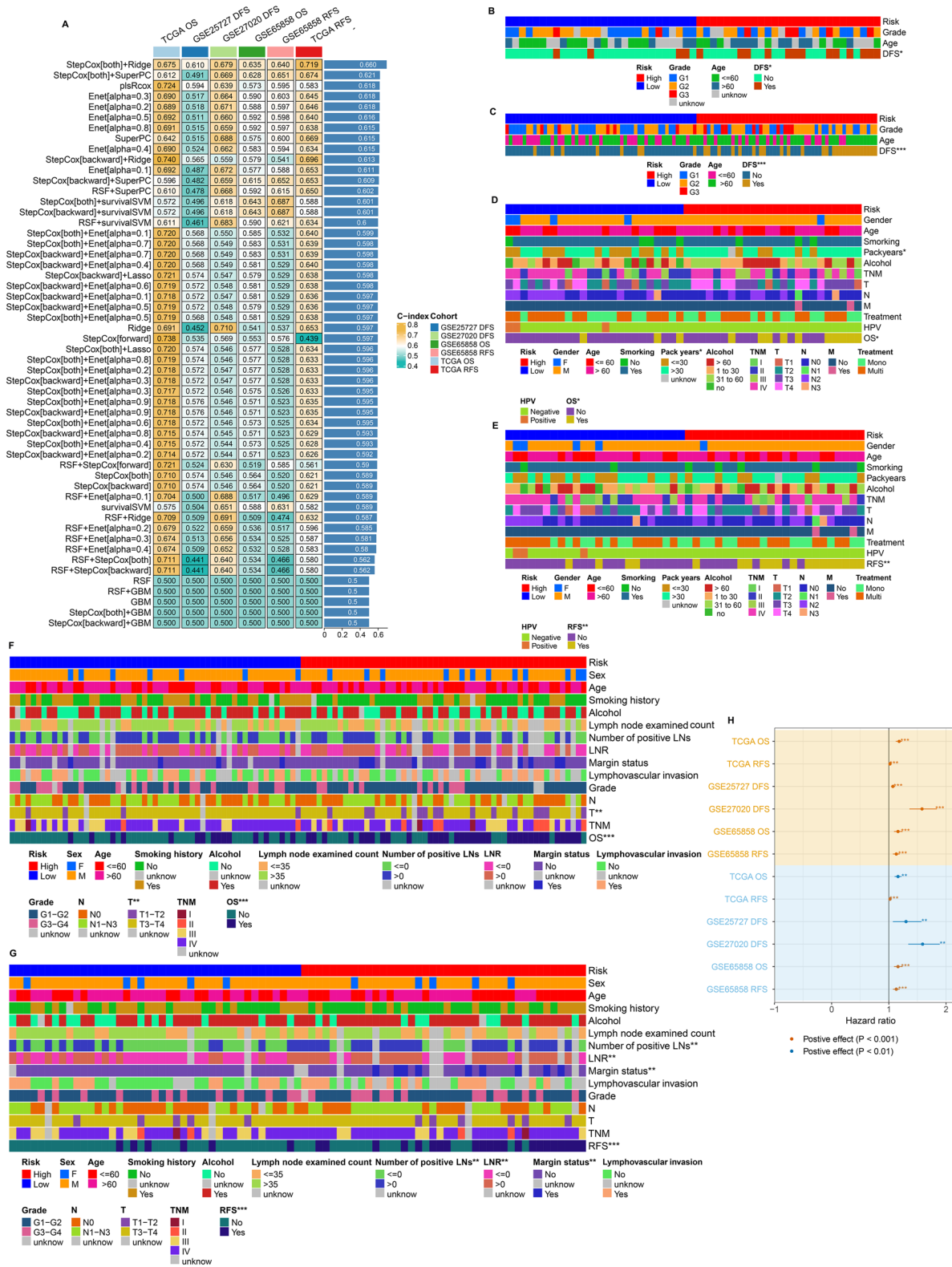
Figure 5. Construction and testing of the artificial intelligence-derived prognostic signature (TCRG classifier). (A) the C-indexes of 54 machine-learning algorithm combinations in the training (TCGA OS) and five testing cohorts (GSE25727 DFS, GSE27020 DFS, GSE65858 OS, GSE65858 RFS, TCGA RFS). (B–G) correlation of TCRG classifier with clinicopathologic characteristics. The heatmap depicted the distribution of clinical characteristics arranged by the increasing TCRG score in the GSE25727 DFS (B), GSE27020 DFS (C), GSE65858 OS (D), GSE65858 RFS (E), TCGA OS (F) and TCGA RFS (G). The asterisks indicate a statistically significant *p* value calculated using the Chi-square test (**p* < .05; ***p* < .01; ****p* < .001). (H) Forest plots showing univariable and multivariable cox regression analysis (hazards ratio with 95% confidence intervals) of TCRG classifier for survival prediction in TCGA and GEO database. The yellow represent univariable cox regression analysis, the light blue represent multivariable cox regression analysis. **p* < .05; ***p* < .01; ****p* < .001.
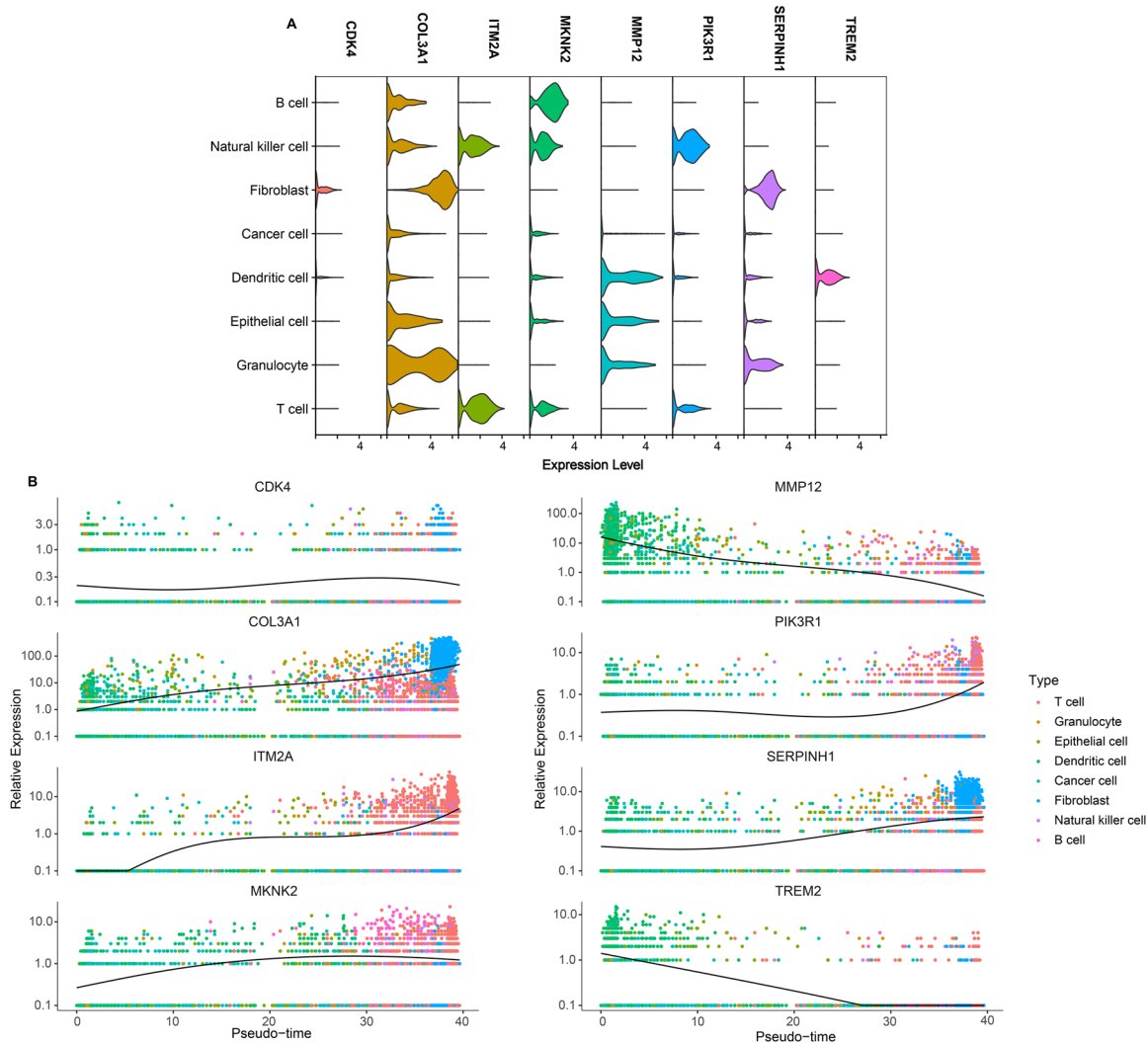
Figure 6. Analyses hub T cell-related genes (TCRGs) in single-cell RNA-seq data. (A) Violin plot depicting the expression of hub TCRGs (CDK4, COL3A1, ITM2A, TREM2, SERPINH1, MMP12, MKNK2 and PIK3R1) in different cell types. (B) Pseudo-temporal change curve of expression level of hub TCRGs (CDK4, COL3A1, ITM2A, TREM2, SERPINH1, MMP12, MKNK2 and PIK3R1) in different cell types.

($p < .001$) and activated CD8 T cell ($p < .001$) (Figure 7(G)). The findings from the TCGA cohorts were similar (Figure S4G).

Comprehensive analysis of molecular characteristics in different TCRG subgroups

To investigate potential biological pathways between the TCRG subgroups in LC, ssGSEA and GSVA were conducted to determine the enrichment score of pre-defined biological processes and KEGG gene sets, respectively. As for pre-defined biological processes, we revealed that T cell co-stimulation, T cell co-inhibition and APC co-stimulation, were enriched in the TCRG low subgroup relative to the TCRG high subgroup in GEO (Figure 8(A)) and TCGA (Figure 8(B)) cohorts. In terms of KEGG gene sets, we discovered that in comparison with the TCRG low subgroup,

pathways in cancer, TGF beta signalling pathway, ECM receptor interaction and focal adhesion were enriched in the TCRG high subgroup, whereas butanoate metabolism and T cell receptor signalling pathway were mainly involved in TCRG low subgroup in GEO (Figure 8(C)) and TCGA (Figure 8(D)) cohorts.

Additionally, correlation analysis uncovered that TCRG was strongly positively linked to focal adhesion, TGF beta signalling pathway and ECM receptor interaction, yet negatively associated with T cell co-stimulation, butanoate metabolism and T cell receptor signalling pathway in GEO (Figure 8(E)) and TCGA (Figure 8(F)) cohorts. Noticeably, hub TCRGs SERPINH1 showed a significant positive link to focal adhesion, TGF beta signalling pathway and ECM receptor interaction, and ITM2A was significantly positively linked to T cell receptor signalling pathway and T cell co-stimulation.
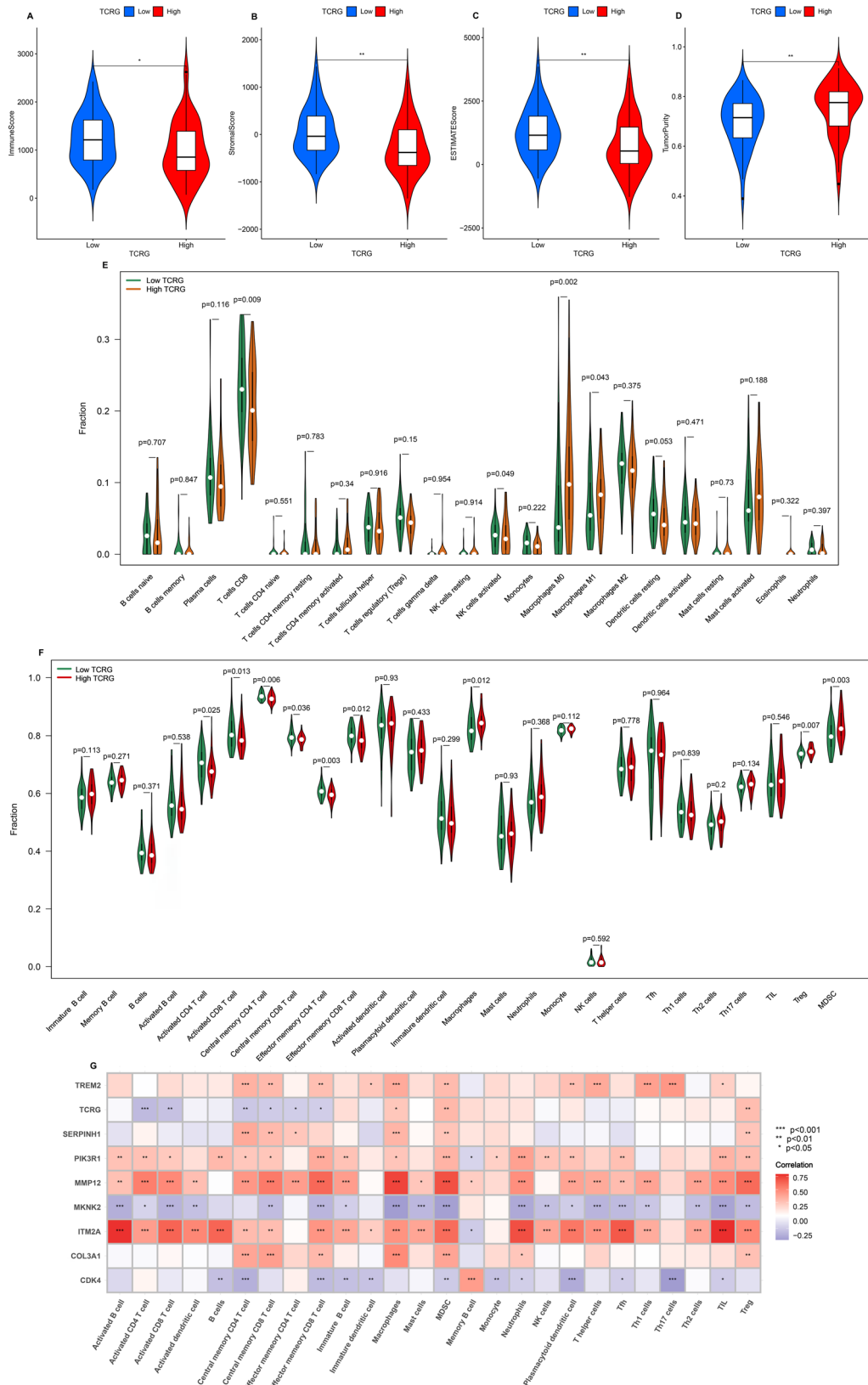
Figure 7. Correlation of TCRG classifier/TCRGs with immune infiltration in GEO datasets. (A–D) Differential analysis of immune score (a), stromal score (B), ESTIMATE score (C) and tumour purity (D) in different TCRG subgroups. The asterisks indicate a significant statistical *p* value calculated using the Wilcoxon test (*p < .05; **p < .01; ***p < .001). (E, F) Comparison of infiltrating immune cells between high and low TCRG subgroups by CIBERSORTx tool (E) and ssGSEA algorithm(F). *p* Value was calculated using the Wilcoxon test. (G): correlation between TCRG classifier/hub TCRGs and infiltrating immune cells. Correlation coefficient and *p* value were calculated by Spearman correlation analysis. The asterisks indicate a statistically significant *p* value calculated using the Spearman correlation analysis (*p < .05; **p < .01; ***p < .001).
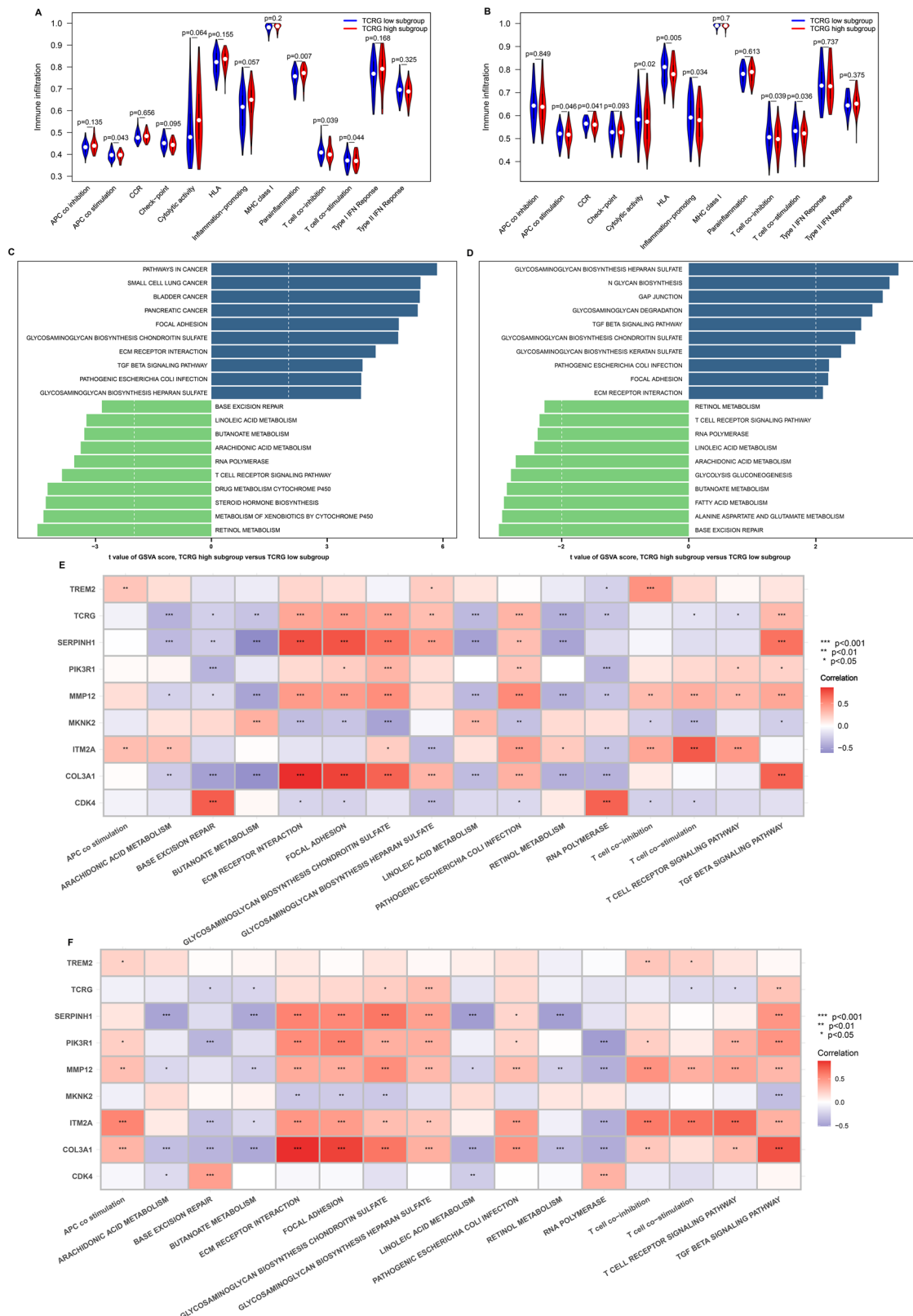
Figure 8. Comprehensive analysis of molecular characteristics in different TCRG subgroups in GEO and TCGA datasets. (A, B) Comparison of pre-defined biological processes between TCRG different subgroups in GEO (A) and TCGA (B) datasets based on ssGSEA algorithm. *p* Value was calculated using the Wilcoxon test (C, D). Comparison of KEGG gene sets from Molecular Signatures Database between TCRG different subgroups in GEO (C) and TCGA (D) datasets based on gene set variation analysis (GSVA). t value and *p* value were calculated using the limma R package. (E, F) The heatmap plot depicted the correlation between TCRG and shared biological pathways in in GEO (E) and TCGA (F) datasets. Correlation coefficient and *p* value were calculated by Spearman correlation analysis. The asterisks indicate a statistically significant *p* value calculated using the Spearman correlation analysis (*p* < .05; **p* < .01; ***p* < .001).

Correlation of TCRG with immunological properties of the TME in LC

We compared the mRNA levels of key immunological markers across TCRG subgroups to gain a deeper understanding of the relationship between TCRG and the immunological properties of the TME in LC. A significant differential expression of various ICIs-related genes (CD160, IDO1, IL10RB and PDCD1) and HLA-related genes (particularly the MHC II molecule) was noted in the GEO (Figure 9(A,B)) and TCGA (Figure 9(C,D)) cohorts. There was also strong evidence from correlation analysis that TCRG was considerably linked to genes associated with ICIs and HLA in the GEO (Figure 9(E,F)) and TCGA (Figure 9(G,H)) cohorts.

We evaluated the connection between TCRG and well-established biological markers in TCGA cohorts to assess the potential effectiveness of the TCRG classifier in predicting immunotherapy responses. Figure 10 displays the findings, which show that low TCRG subgroups were characterized by low levels of ploidy and tumour purity, along with a high level of KS of neoantigens, AUC of neoantigens, clonal neoantigens, subclonal neoantigens, TMB and MSI. Moreover, a significant positive link was observed between TCRG and TMB ($R = 0.13$, $p = .02$) (Figure 10(M)) and MSI ($R = 0.17$, $p = .043$) (Figure 10(N)).

Role of TCRG classifier in immunotherapy

Importantly, we further clarified the effects of TCRG classifiers in immunotherapy. The IPS analysis uncovered that compared to high TCRG subgroups. Notably, higher IPS CTLA4 (+) PD1(+) and IPS CTLA4(+) PD1(−) scores were recorded in the low TCRG subgroups (Figure 11(C,D)), while IPS CTLA4 (+) PD1(+) and IPS CTLA4(+) PD1 (−) scores did not achieve a statistically significant difference (Figure 11(A,B)).

Additionally, the TCIC analysis found CD8 T cell recruiting and T cell recruiting were substantially increased in low TCRG subgroups than in high TCRG subgroups (Figure 11(E,F)). Then, the TIDE web platform revealed that the low TCRG subgroups exhibited higher CD8 and IFNG scores, lower CAF, exclusion and TIDE scores (Figure 11(G–K)), as well as higher immunotherapy response rates (84% *vs* 54%, two-sided Chi-square test, $p < .05$) (Figure 11(L)).

Based on eight TCRGs, we constructed a TCRG classifier in the IMvigor210 cohort. In the IMvigor210 cohort, the KM curve for OS showed that the high TCRG group exhibited considerably shorter survival ($p < .001$, Figure 11(M)). The AUC for TCRG was 0.684 after one year of OS and 0.673 after two years of OS,

indicating that it had a greater predictive ability according to the time-dependent ROC curves (Figure 11(N)). Additionally, compared with stable disease (SD)/progressive disease (PD) after immunotherapy, complete response (CR)/partial response (PR) after immunotherapy has a lower TCRG score ($p < .001$) (Figure 11(O)). Immunotherapy treatment was more effective in 54/174 of low TCRG patients compared with those with high TCRG (24/174) (two-sided Chi-square test, $p < .05$) (Figure 11(P)).

Discussion

In this study, we collected three LC samples and paired adjoining normal samples to perform 10X Genomics single-cell RNA-seq followed by comprehensive bioinformatic analyses, uncovered that T cells are the key components of the LC TME, significantly involved in cell differentiation trajectory and have a critical function in cell-cell interactions. These findings prompt us to comprehend the potential clinical value of T cell-related molecular in LC through a combination of single-cell and bulk RNA-seq data. Accordingly, we started by developing and validating an eight-gene TCRG classifier utilizing 54 distinct combinations of ML algorithms in four separate multicentre cohorts that had diverse prognostic outcomes. Our TCRG classifier showed strong and better prediction power in comparison with many common clinicopathological characteristics. Additionally, the identified eight TCRGs (CDK4, COL3A1, ITM2A, TREM2, SERPINH1, MMP12, MKNK2 and PIK3R1) are expected to be new potential targets for LC therapy. Importantly, we substantiated that the low TCRG subgroup showed considerably prolonged DFS, OS and RFS, had a significant enrichment in immune-associated pathways (especially T cell activation), higher neoantigens and frequency mutations, increased ICIs and HLA expression, relatively activated TME and better immunotherapeutic efficacy in GEO and TCGA cohorts, vice versa. As a result, our TCRG classifier has the potential to lay a solid foundation for LC personalized decision-making in clinical contexts.

The TME's regulatory function in tumour formation has been the subject of intense investigation in several fields, including LC. The efficacy of immuno-oncology treatments is likely to be a contributing factor. Furthermore, immunotherapy effectiveness is dictated by processes that are regulated by intricate communication among TME-resident immunological cells, stromal cells and epithelial cells [22]. Tumour immunotherapy using ICIs has demonstrated the need to target the TME rather than the tumour itself. Novel technologies like single-cell RNA-seq allow for a
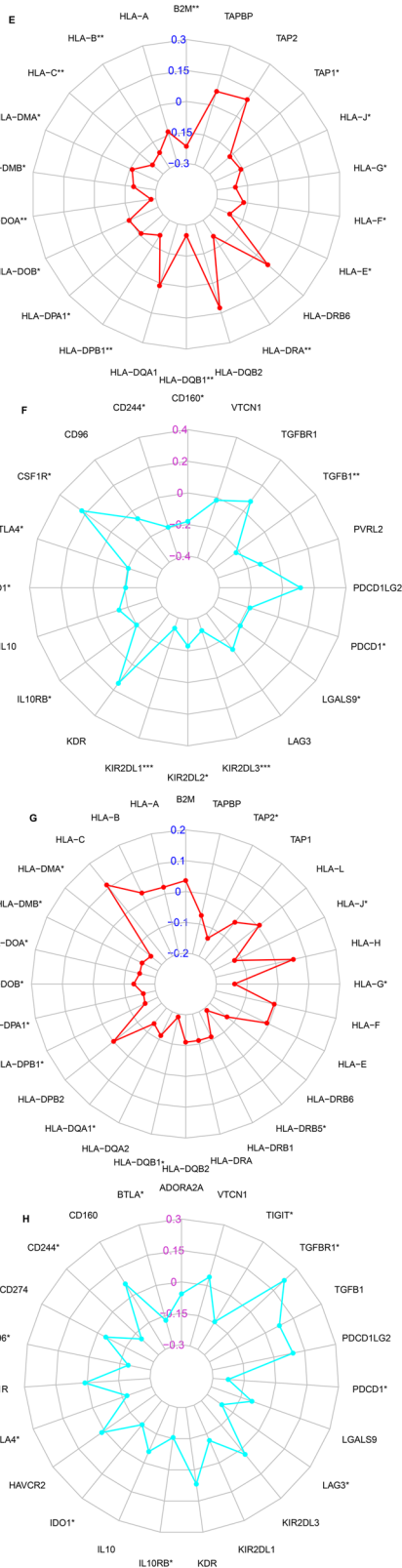
Figure 9. Analysing the correlation between TCRG and immunomodulators (ICIs-related genes and HLA-related genes) in GEO and TCGA datasets. (A, B): Differential expression of ICIs-related genes (A) and HLA-related genes (B) between different TCRG expression subgroups in GEO datasets. (C, D): Differential expression of ICIs-related genes (C) and HLA-related genes (D) between different TCRG expression subgroups in TCGA datasets. p Value was calculated using the Wilcoxon test. (E, F) The radar plot depicted correlation between TCRG and ICIs-related genes (E) and HLA-related genes (F) in GEO datasets. (G, H) The radar plot depicted correlation between TCRG and ICIs-related genes (E) and HLA-related genes (F) in TCGA datasets. Correlation coefficient and p value were calculated by Spearman correlation analysis. The asterisks indicate a statistically significant p value calculated using the Spearman correlation analysis (*p < .05; **p < .01; ***p < .001).
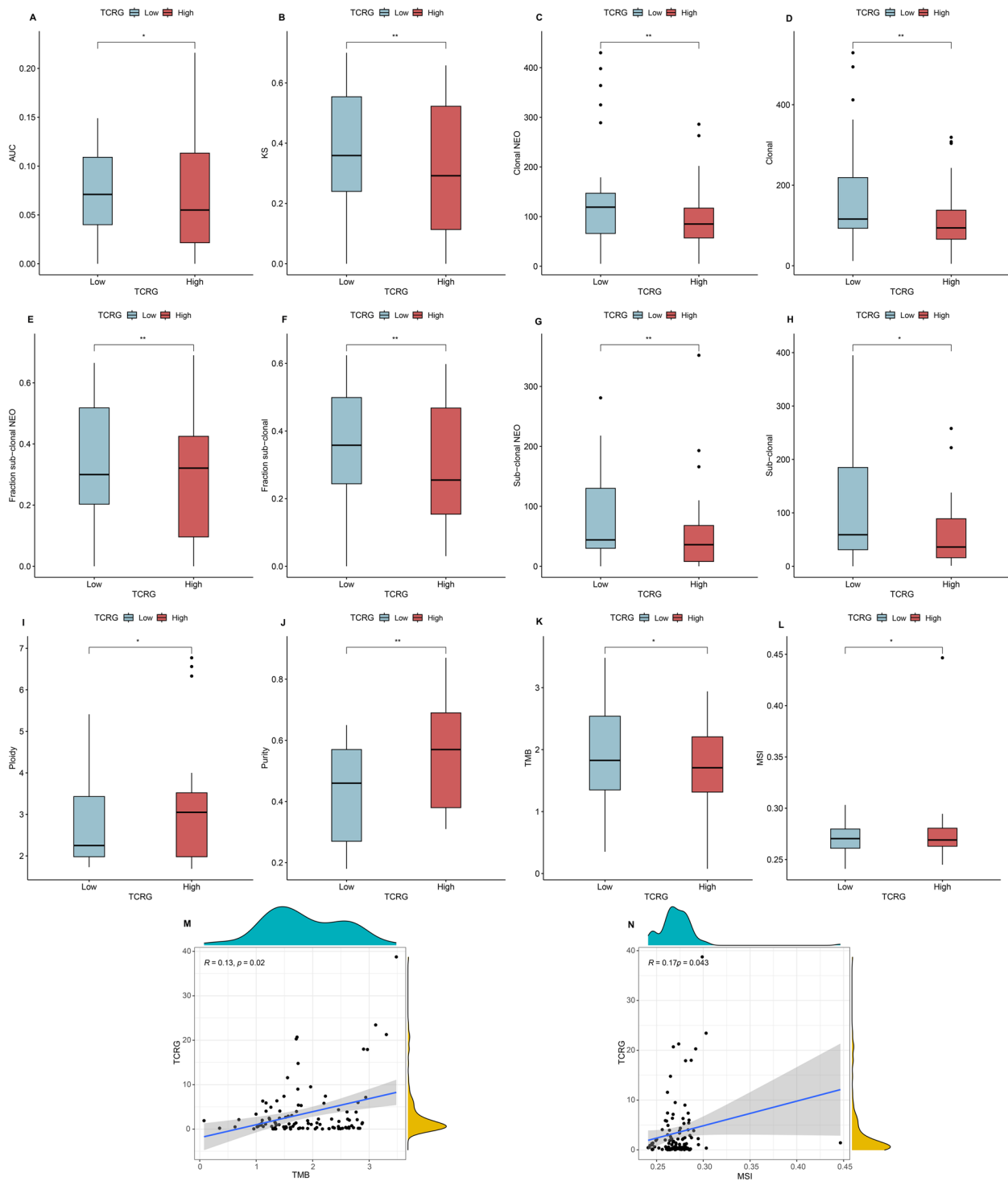
Figure 10. Analysing the correlation between TCRG and neoantigen, tumour mutation burden (TMB) and microsatellite instability (MSI) in TCGA datasets. (A–L): Differential analysis of AUC of neoantigens (a), KS of neoantigens (B), the number of clonal neoantigens (C), the number of clonal (D), fraction subclonal (E), fraction subclonal neoantigens (F), the number of subclonal neoantigens (G), the number of subclonal (H), ploidy(I), purity (J), TMB (K) and MSI (L). The asterisks indicate a significant statistical p value calculated using the Wilcoxon test (*$p<.05$; **$p<.01$; ***$p<.001$). (M, N): correlation between TCRG expression and TMB (M) and MSI (N). Correlation coefficient and p value were calculated by spearman correlation analysis.

multidimensional analysis of tumour heterogeneity and immune microenvironment. The emphasis of numerous single-cell research has been on T cells since they are among the key effector cell types that drive anti-cancer immunity and are the target of several presently developing immunotherapy medicines. A better knowledge of the LC TME at the single-cell level is vital for the development of more effective
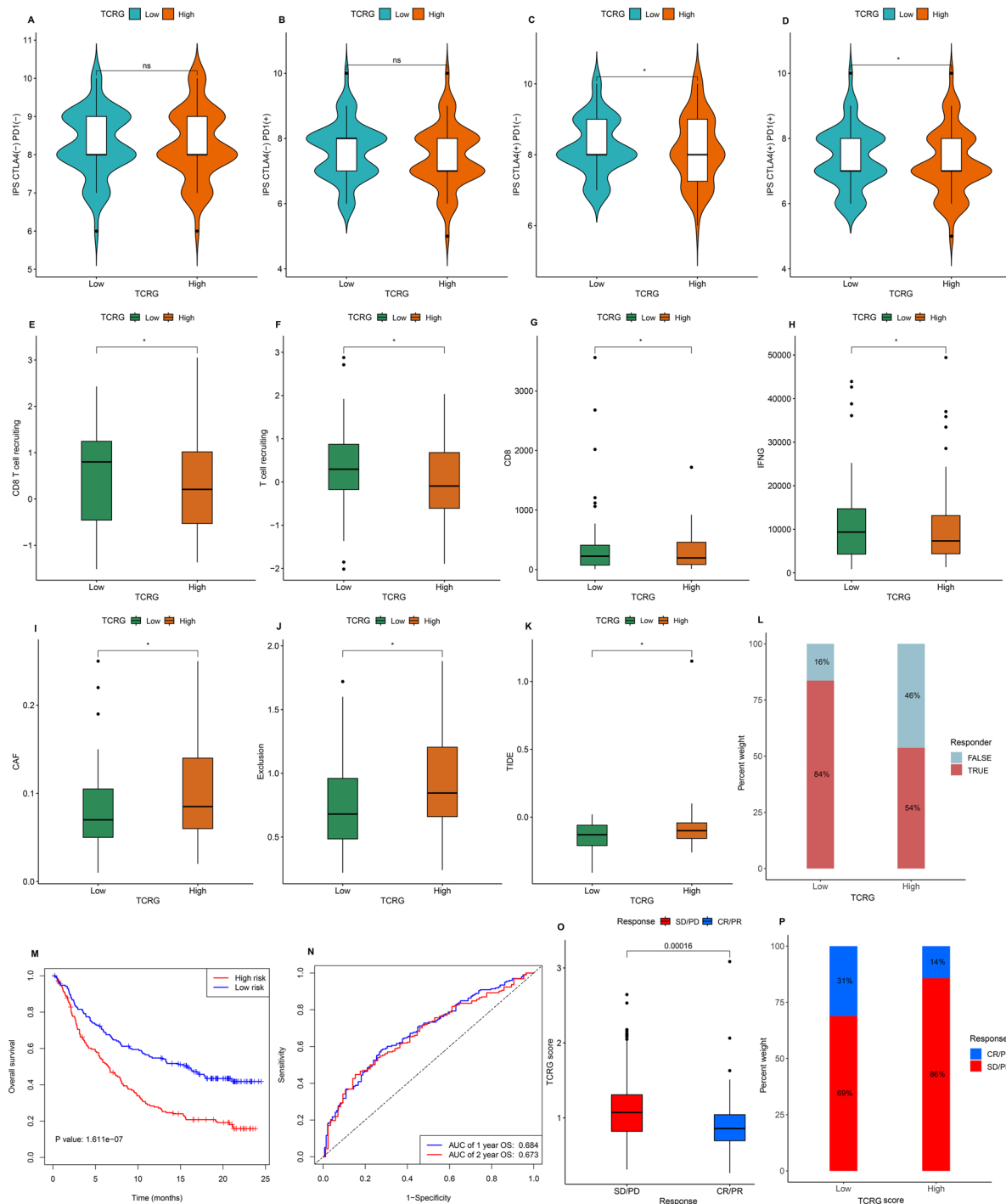
Figure 11. Evaluation of TCRG potential in predicting immunotherapeutic response based on IPS (immunophenoscore), TCIC (the cancer immunity cycle), TIDE (tumour immune dysfunction and exclusion) and IMvigor210 cohort analysis. (A–K): Differential analysis of IPS CTLA4(−) PD1(−) score (a), IPS CTLA4(−) PD1(+) score (B), IPS CTLA4(+) PD1(−) score (C), IPS CTLA4(+) PD1(+) score (D), CD8 T cell recruiting (E), T cell recruiting (F), CD8 (G), IFNG (H), cancer associated fibroblasts (CAF) (I), tumour immune exclusion (J) and TIDE score (K). The asterisks indicate a significant statistical p value calculated using the Wilcoxon test (*$p<.05$; **$p<.01$; ***$p<.001$). (L): the distribution of immunotherapeutic response in two groups stratified by TCRG in TCGA LC cohort based on the TIDE algorithm. A two-sided chi-square test was used to analyse contingency tables for ICI responders. (M–P): construction a TCRG classifier in IMvigor210 cohort, further evaluation of the prognostic value and immunotherapy response of TCRG classifier in IMvigor210 cohort. (M): Kaplan-Meier survival curves with log-rank test for different TCRG subgroups to compare the OS differences in IMvigor210 cohort. (N): time-independent ROC curves with AUC values to evaluate predictive efficacy of TCRG based on OS in IMvigor210 cohort. (O): comparison of TCRG score between stable disease (SD)/progressive disease (PD) and complete response (CR)/partial response (PR) groups. The asterisks indicate a significant statistical p value calculated using the Wilcoxon test (*$p<.05$; **$p<.01$; ***$p<.001$). (P): the distribution of immunotherapeutic response in two groups stratified by TCRG in IMvigor210 cohort. A two-sided chi-square test was used to analyse contingency tables for ICI responders.

treatments. Recently, one single-cell RNA-seq study on laryngeal squamous cell carcinoma (LSCC) *in situ* [23] and another single-cell RNA-seq study with lymphatic metastasis [24] have been reported. Song et al. [23] collected two LSCC tissue samples and Sun et al. [24] collected three metastatic LSCC tissue samples to conduct a single-cell RNA-seq analysis, delineate a landscape of the LSCC intratumor heterogeneity and detect a complex but unique LSCC TME which was consistent with our study. However, Song et al. [23] discovered that the abundance of tumour cells was highest, followed by immune cells. Sun et al. [24] confirmed that there were a significant number of epithelial cells, with T cells following closely after, and our study showed that the majority of the cells were T cells. This result is due to the heterogeneity of the tumour and remains to be determined. Additionally, they did not identify and analyse T cell-related molecules in depth. Notably, Song et al.'s study [23] did not decipher these complex intercellular crosstalk and Sun et al.'s study [24] did not elucidate the unique cellular communication of T cells with other immune cells. In our study, we highlighted the function of T cells/T cells-related molecules in LC TME, identified potential therapeutic targets and deciphered T cells and other cells' complex intercellular crosstalk. As a result, we identified 2096T cell type-specific DEGs using the FindAllMarkers and Wilcoxon test, in preparation for subsequent bulk RNA sequencing data analysis. Cell-cell communications analysis uncovered that T cells can influence epithelial cells, dendritic cells, cancer cells, fibroblast and NK cell functions through the IFNG, CCL5, ANXA1, NAMPT and SPP1 signalling pathways. These findings suggest that T cells have a critical function in intercellular communication of LC TME and future basic experiments could further explore the molecular mechanisms by which T cells interact with other cells in LC.

Accurately evaluating prognosis and predicting the effectiveness of LC patients is a challenge in the age of precision medicine and anatomy-based TNM staging falls far short of clinicians' expectations [25]. Although several LC prognostic markers have been developed recently, they may not be able to shed light upon key pathophysiological changes or uncover the mechanisms at work. Pathophysiologically, thousands of genes are differentially expressed in response to oncogenesis and progression of LC [26], and some genes can potentially provide valuable insights into prognosis and prediction. Furthermore, we observed only moderate predictive performance and poor methodological quality overall [27]. Most importantly, most of them are based on TCGA cohorts, yet there is a lack of sufficient external validation [28,29], lower statistical

power and model generalizability. In this study, we systematically reviewed available LC bulk RNA-seq datasets from inception to date and emphasized the effects of T cell-related molecules in LC. As a result, considering T cell type-specific DEGs at single-cell resolution, we identified 55T cell-specific genes based on six GEO and TCGA cohorts, which were chosen to determine the key hub genes. Also, previous research shows that individuals often choose modelling algorithms according to their preferences and the extent of their knowledge. We compiled a list of 10 machine-learning methods that may be employed to create biological prognostic markers to address this limitation. We further merged them into 54 algorithm combinations, with the functions of variable screening and data dimensionality reduction performed by CoxBoost, RSF, LASSO and Stepwise Cox. The final model was determined to be the combination of StepCox[both] and Ridge that yielded the largest average C-index (0.660) among the five remaining testing cohorts.

When developing biomedical models using AI and ML, one of the most troublesome issues is over-fitting, which occurs when a model fits the training data well but fails to do so when tested with other external validation datasets [30]. A final eight-gene signature, the TCRG classifier, was formed using Ridge after StepCox[both] minimized redundant data. The TCRG classifier demonstrated outstanding predictive and prognostic performance in the training cohort as well as five test cohorts with varying survival outcomes (OS, DFS and RFS), according to the results of risk stratification analysis, KM analysis, time-dependent ROC curve and univariate and multivariate Cox regression. Moreover, when pitted against several common clinicopathologic variables such as classical TNM stage, age, grade, margin status, or lymphovascular invasion, our TCRG classifier showed significantly improved accuracy.

Therapy decision-making for LC patients generally makes use of the TNM stage system [31], developed by the American Joint Committee on Cancer based on anatomical information. As a result, a significant proportion of patients with erroneous staging may get either too much or too little therapy. Using the TCRG to make decisions on LC survival and recurrence was shown to have a higher net benefit according to DCA findings than using TNM stage or other clinicopathologic variables, or treating all patients or none at all in the TCGA and GEO datasets. Additionally, the most recent National Comprehensive Cancer Network Guidelines place an extensive emphasis on tumour TNM staging when making decisions on larynx-preserving therapy, which is consistent with the worldwide tendency towards

organ-preserving approaches for LC patients. The larynx-preserving procedure can be challenging for surgeons to perform, but it may be beneficial for patients with advanced tumours at the TNM stage [32]. Investigating combination techniques and personalized treatment plans to significantly enhance the prognostic classification of LC patients is a highly urgent matter. Currently, a multifaceted evaluation found that our TCRG classifier outperforms TNM staging to guide personalized management in LC, but it remains to be tested in clinical practice. To further increase prediction accuracy and enable personalized patient therapy, future studies should integrate the TCRG classifier with other clinical data, like the classical TNM stage, because each mirrors various pathophysiological aspects.

The current state of prognostic biological markers and models is inadequate for LC since they have mostly been used to estimate the overall prognosis. Stratifying patients to direct treatments like immunotherapy ought to be the focus of the additional data (We have already mentioned that the poor response rate of patients is a fundamental restriction of immunotherapy efficacy). The next step was to compare the immunological profiles of the high- and low-TCRG groups. Notably, the results of ESTIMATE, CIBERSORTx and ssGSEA demonstrated larger immune scores and immune cell type levels (e.g. activated CD4+/CD8+ T cells and NK cells) and a lower tumour purity in the low TCRG subgroup in TCGA and meta-GEO cohorts. It is well known that the basis for the beginning of the antitumour immune response is the activation of immune cells. The low TCRG subgroup was characterized by the activation of adaptive immunity, corresponding to the immune-inflamed phenotype; the high TCRG subgroup was found to have suppressed immunity, which fits the description of an immune-desert phenotype. Conversely, the immune-inflamed phenotype, often referred to as a hot tumour, exhibits a large number of effector immune cell infiltration in TME [33], which enhances antitumour immunity. Conversely, the immune-desert phenotypes are often typified by immune tolerance and ignorance and the absence of activated and priming T-lymphocytes [34]. In agreement with ssGSEA and GSVA results for potential biological pathways in different TCRG subgroups, High TCRG subgroups exhibited strongly active carcinogenic pathways, including ECM receptor interaction and the TGF beta signalling pathway and tended to have the immune-desert phenotype (cold tumour phenotype). Furthermore, the low TCRG subgroup is rich in immune activation-related pathways (T cell co-stimulation and T cell receptor signalling pathway) and may have stronger antitumour immunity.

Tumour immunogenicity is another factor that triggers the antitumour immune response. In general, cancerous cells with more frequent somatic mutations may produce more new antigens, which might enhance the immune system's ability to destroy these cells [35,36]. Our study found that in comparison to patients in the high TCRG subgroup, patients in the low TCRG subgroup had considerably higher neoantigen (clonal, clonal neoantigens, subclonal, subclonal neoantigens), TMB and MSI, which are among the most reliable indicators of response to anti-PD-L1/PD-1 ICIs that have been tested so far. Besides, TCRG is significantly positively related to neoantigen, TMB and MSI. Notably, distinctive features within the TME, which include immune infiltration and tumour purity, might impact the tumour's immunogenicity [37]. These findings align with our previous results of characterizing the infiltration of cells in the TME.

Based on these findings, we tested whether TCRG can reliably predict the efficacy of immunotherapy by exhaustively characterizing TME features in LC. In line with expectations, we noted that the low TCRG patients had substantially higher expression of ICIs-related genes including IDO1 and PDCD1, indicating that they could potentially gain more benefit from ICI therapy. Three well-known methods of estimating tumour patients' ICI sensitivity using expression profiles are IPS, TCIC and TIDE. Our previous finding that immunotherapy was more effective in patients with low TCRG was consistently supported by the data. Importantly, we applied TCRG to a real immunotherapy cohort (IMvigor210 cohort treated with a PD-L1-targeting antibody) and found that immunotherapy was more effective in the low TCRG population than that in the high TCRG subgroup. Overall, our results suggested that our TCRG might act as a benchmark for the early detection of LC patients who are responsive to immunotherapy and are undergoing first-line immunotherapy.

One prominent finding of this study was that we identified eight TCRGs (CDK4, COL3A1, ITM2A, TREM2, SERPINH1, MMP12, MKNK2 and PIK3R1) that are anticipated to be new potential targets for LC therapy. SERPINH1 (Serpin peptidase inhibitor clade H, member 1), is located in the endoplasmic reticulum and functions as a molecular chaperone that is specific to collagen during collagen biosynthesis. A recent pan-cancer study [37] found that SERPINH1 overexpression substantially reduced the OS and disease-specific survival in multiple cancers. Additionally, SERPINH1 expression exhibited a close link to the infiltration of immunological cells, ICI markers, immunoregulators in brain lower-grade glioma (LGG) and testicular germ cell tumours. We found that high SERPINH1 levels correlated

to a significant reduction in OS, DFS and RFS, was specifically expressed in CAF, gradually increased at the later stage of the pseudo-time trajectory for T cell development, significantly positively correlated with classical immunosuppressive cell (MDSC and Tregs) and had a significant enrichment of ECM receptor interaction and TGF beta signalling pathway. These findings hint that CAF-derived SERPINH1 may mediate T-cell function leading to an immunosuppressive TME that reduces the anti-tumour effects in patients with LC. Conversely, ITM2A (integral membrane protein 2A), belongs to the Type II ITM2 family [38]and is preferentially expressed in T lineage cells among haematopoietic cells. Our study implies that ITM2A may be able to activate T cell immune function (T cell co-stimulation, activated CD4 and CD8 cells), leading to an immune-inflamed TME that enhances anti-tumour effects in LC. Therefore, further characterization of molecules including SERPINH1, ITM2A, TREM2, CDK4, COL3A1, MMP12, MKNK2, and PIK3R1 will offer new insights into the oncogenesis and progression of LC and aid in finding possible treatment targets for LC patients.

Compared with earlier published research, our study features several notable distinctions. (1) Through single-cell RNA-seq techniques, we first highlighted the role of T cells/T cells-related molecules in LC TME, identified potential therapeutic targets and deciphered T cells and other cells' complex intercellular crosstalk. (2) We systematically reviewed available LC bulk RNA-seq datasets from inception to date and integrated single-cell and bulk RNA-seq data (six GEO and TCGA cohorts) to identify convincing hub T cell-specific genes. (3) We systematically collected four multicentre cohorts with different survival outcomes (OS, DFS and RFS) and developed our TCRG classifier using the algorithms that had the highest average C-index across all five testing cohorts, which improved the model's stability and prognostic power. (4) We merged 10 well-known ML algorithms into 54 different combinations and then chose the most accurate one to avoid the inappropriate modelling methods owing to subjective preferences. (5) We identified two TCRGs (SERPINH1 and ITM2A) that play important roles in LC TME and are anticipated to be new potential targets for LC therapy. Our study was as thorough and meticulous as possible, but there are certain caveats to consider. First, the sizes and sequencing platforms of the included cohorts varied even though these variations were mitigated by utilizing correction algorithms. Second, since we only compile four separate multi-centre cohorts, the therapeutic benefit of TCRG should be further confirmed in bigger, prospective multicentre cohorts. Third, we found that TCRG influences LC tumour prognosis, immunological properties and TME remodelling. Nonetheless, *in vivo* validation should be performed in future investigations to investigate the specific mechanisms underlying the eight TCRGs. Lastly, the comprehensive molecular mechanisms *in vivo* and *in vitro* ought to be further investigated in future research to optimize therapies, such as combining ICBs with anti-SERPINH1 drugs or an ITM2A agonist to improve the outcomes of LC patients.

Conclusion

We used 54 different combinations of ML algorithms to develop and verify a consensus prognostic signature (TCRG classifier) based on eight TCRGs from the single-cell RNA-seq cohort, training cohort and five testing cohorts. We found that our TCRG had strong and much better prediction power, which has significant implications for the clinical care and personalized treatment of LC (such as immunotherapy). Overall, through integrating single-cell and bulk RNA-seq data and cutting-edge computational algorithms, this work offers a promising method for clinical practice that may be used to assess prognosis, stratify patients according to risk and provide personalized care for those with LC.

Ethics approval and consent to participate

The studies involving human samples were reviewed and approved by the Research Ethics Committee of Tianjin Medical University Cancer Institute and Hospital and complied with the Declaration of Helsinki (No. Ek2022226). Prior to sample and data collection, written informed consent was obtained from the eligible patients.

Authors' contributions

JC, XDW and YPO conceived and designed the study. JC, XDW and YPO drafted the manuscript. JC, XDW, MHW, YPO, KY, YSW, GLL and YSD analysed and interpreted all the data. JC, XDW, MHW, KY, YSW, ZC and YSD prepared the figures and tables. JC, XDW, MHW, YPO, KY, YSW, GLL, YSD and ZC reviewed and revised the manuscript. All authors have read and approved the manuscript for publication.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The datasets analysed during the current study are available from the corresponding author on reasonable request *via* email.

References

[1] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin. 2016;66(2):115–132. doi: 10.3322/caac.21338.

[2] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7–30. doi: 10.3322/caac.21590.

[3] Steuer CE, El-Deiry M, Parks JR, et al. An update on larynx cancer. CA Cancer J Clin. 2017;67(1):31–50. doi: 10.3322/caac.21386.

[4] Chow LQM. Head and neck cancer. N Engl J Med. 2020;382(1):60–72. doi: 10.1056/NEJMra1715715.

[5] Hammerman PS, Hayes DN, Grandis JR. Therapeutic insights from genomic studies of head and neck squamous cell carcinomas. Cancer Discov. 2015;5(3):239–244. doi: 10.1158/2159-8290.CD-14-1205.

[6] Ferris RL, Blumenschein GJr, Fayette J, et al. Nivolumab for recurrent squamous-cell carcinoma of the head and neck. N Engl J Med. 2016;375(19):1856–1867. doi: 10.1056/NEJMoa1602252.

[7] Qi H, Chen W, Zhang C, et al. Epidemiological analysis of 1234 cases of laryngeal cancer in Shanxi Province, China. Cancer Control. 2021;28:10732748211041236. doi: 10.1177/10732748211041236.

[8] Zhang J, Liu Y, Xia L, et al. Constructing heterogeneous single-cell landscape and identifying microenvironment molecular characteristics of primary and lymphatic metastatic head and neck squamous cell carcinoma. Comput Biol Med. 2023;165:107459. doi: 10.1016/j.compbiomed.2023.107459.

[9] Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. Nature. 2013;501(7467):328–337. doi: 10.1038/nature12624.

[10] Bolon B, Aeffner F. A primer for oncoimmunology (immunooncology). Toxicol Pathol. 2017;45(5):584–588. doi: 10.1177/0192623317713318.

[11] Chen L, Weng Y, Cui X, et al. Comprehensive analyses of a CD8+ T cell infiltration related gene signature with regard to the prediction of prognosis and immunotherapy response in lung squamous cell carcinoma. BMC Bioinformatics. 2023;24(1):238. doi: 10.1186/s12859-023-05302-3.

[12] Cai DQ, Cai D, Zou Y, et al. Construction and validation of chemoresistance-associated tumor-infiltrating exhausted-like CD8+ T cell signature in breast cancer: cr-TILCD8TSig. Front Immunol. 2023;14:1120886. doi: 10.3389/fimmu.2023.1120886.

[13] Chen J, Wang Z, Zhu Q, et al. Comprehensive analysis and experimental verification of the mechanism of action of T cell-mediated tumor-killing related genes in colon adenocarcinoma. Transl Oncol. 2024;43:101918. doi: 10.1016/j.tranon.2024.101918.

[14] Johnson E, Dickerson KL, Connolly ID, et al. Single-cell RNA-sequencing in glioma. Curr Oncol Rep. 2018;20(5):42. doi: 10.1007/s11912-018-0673-2.

[15] Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. Nature. 2017;541(7637):331–338. doi: 10.1038/nature21350.

[16] Cui J, Chen Y, Ou Y, et al. Cancer germline antigen gene MAGEB2 promotes cell invasion and correlates with immune microenvironment and immunotherapeutic efficiency in laryngeal cancer. Clin Immunol. 2022;240:109045. doi: 10.1016/j.clim.2022.109045.

[17] Satija R, Farrell JA, Gennert D, et al. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015;33(5):495–502. doi: 10.1038/nbt.3192.

[18] Hou Y, Chen Z, Wang L, et al. Characterization of immune-related genes and immune infiltration features in epilepsy by multi-transcriptome data. J Inflamm Res. 2022;15:2855–2876. doi: 10.2147/JIR.S360743.

[19] Lu J, Chen R, Ou Y, et al. Characterization of immune-related genes and immune infiltration features for early diagnosis, prognosis and recognition of immunosuppression in sepsis. Int Immunopharmacol. 2022;107:108650. doi: 10.1016/j.intimp.2022.108650.

[20] Chen Z, Lu J, Liu G, et al. Comprehensive characterization of cytokines in patients under extracorporeal membrane oxygenation: evidence from integrated bulk and single-cell RNA sequencing data using multiple machine learning approaches. Shock. 2025;63(2):267–281. doi: 10.1097/SHK.0000000000002425.

[21] Chen Z, Zeng L, Liu G, et al. Construction of autophagy-related gene classifier for early diagnosis, prognosis and predicting immune microenvironment features in sepsis by machine learning algorithms. J Inflamm Res. 2022;15:6165–6186. doi: 10.2147/JIR.S386714.

[22] Sharma P, Hu-Lieskovan S, Wargo JA, et al. Primary, adaptive, and acquired resistance to cancer immunotherapy. Cell. 2017;168(4):707–723. doi: 10.1016/j.cell.2017.01.017.

[23] Song L, Zhang S, Yu S, et al. Cellular heterogeneity landscape in laryngeal squamous cell carcinoma. Int J Cancer. 2020;147(10):2879–2890. doi: 10.1002/ijc.33192.

[24] Sun Y, Chen S, Lu Y, et al. Single-cell transcriptomic analyses of tumor microenvironment and molecular reprograming landscape of metastatic laryngeal squamous cell carcinoma. Commun Biol. 2024;7(1):63. doi: 10.1038/s42003-024-05765-x.

[25] Cui J, Wang L, Tan G, et al. Development and validation of nomograms to accurately predict risk of recurrence for patients with laryngeal squamous cell carcinoma: cohort study. Int J Surg. 2020;76:163–170. doi: 10.1016/j.ijsu.2020.03.010.

[26] Cui J, Wang L, Zhong W, et al. Identification and validation of methylation-driven genes prognostic signature for recurrence of laryngeal squamous cell carcinoma by integrated bioinformatics analysis. Cancer Cell Int. 2020;20(1):472. doi: 10.1186/s12935-020-01567-3.

[27] Cui J, Wang L, Zhong W, et al. Development and validation of epigenetic signature predict survival for patients with laryngeal squamous cell carcinoma. DNA Cell Biol. 2021;40(2):247–264. doi: 10.1089/dna.2020.5789.

[28] Sun Q, Qin X, Zhao J, et al. Cuproptosis-related LncRNA signatures as a prognostic model for head and neck squamous cell carcinoma. Apoptosis. 2023;28(1-2):247–262. doi: 10.1007/s10495-022-01790-5.

[29] Dai Y, Wang Z, Yan E, et al. Development of a novel signature derived from single cell RNA-sequencing for preoperative prediction of lymph node metastasis in head and neck squamous cell carcinoma. Head Neck. 2022;44(10):2171–2180. doi: 10.1002/hed.27126.

[30] Zhang S, Zhang W, Zhang J. 8-Gene signature related to CD8+ T cell infiltration by integrating single-cell and bulk RNA-sequencing in head and neck squamous cell carcinoma. Front Genet. 2022;13:938611. doi: 10.3389/fgene.2022.938611.

[31] Berger DMS, Wassenberg RM, Jóźwiak K, et al. Inter-observer variation in the histopathology reports of head and neck melanoma; a comparison between the seventh and eighth edition of the AJCC staging system. Eur J Surg Oncol. 2019;45(2):235–241. doi: 10.1016/j.ejso.2018.10.529.

[32] Colevas AD, Yom SS, Pfister DG, et al. NCCN guidelines insights: head and neck cancers, version 1.2018. J Natl Compr Canc Netw. 2018;16(5):479–490. doi: 10.6004/jnccn.2018.0026.

[33] Zhang B, Wu Q, Li B, et al. m6A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. Mol Cancer. 2020;19(1):53. doi: 10.1186/s12943-020-01170-0.

[34] Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. Nature. 2017;541(7637):321–330. doi: 10.1038/nature21349.

[35] Gubin MM, Artyomov MN, Mardis ER, et al. Tumor neoantigens: building a framework for personalized cancer immunotherapy. J Clin Invest. 2015;125(9):3413–3421. doi: 10.1172/JCI80008.

[36] Zhang J, Zhou N, Lin A, et al. ZFHX3 mutation as a protective biomarker for immune checkpoint blockade in non-small cell lung cancer. Cancer Immunol Immunother. 2021;70(1):137–151. doi: 10.1007/s00262-020-02668-8.

[37] Wang Y, Gu W, Wen W, et al. SERPINH1 is a potential prognostic biomarker and correlated with immune infiltration: a pan-cancer analysis. Front Genet. 2021;12:756094. doi: 10.3389/fgene.2021.756094.

[38] Zhang R, Xu T, Xia Y, et al. ITM2A as a tumor suppressor and its correlation with PD-L1 in breast cancer. Front Oncol. 2020;10:581733. doi: 10.3389/fonc.2020.581733.