Check for updates

*CORRESPONDENCE
Jian Shi,
✉ shijian27@mail.sysu.edu.cn
Fang Wang,
✉ wangf359@mail.sysu.edu.cn

†These authors share first authorship

Single-cell transcriptomics reveals tumor microenvironment remodeling in hepatocellular carcinoma with varying tumor subclonal complexity

Jian Shi[1,2†]*, Yanru Zhang[2†], Lixia Xu[1] and Fang Wang[2]*

[1]Department of Oncology, The First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, China, [2]Institute of Precision Medicine, The First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, China

Introduction: The complexity of tumor cell subclonal structure has been extensively investigated in hepatocellular carcinoma. However, the role of subclonal complexity in reshaping the tumor microenvironment (TME) remains poorly understood.

Methods: We integrated single-cell transcriptome sequencing data from four independent HCC cohorts, involving 30 samples, to decode the associations between tumor subclonal complexity and the TME. We proposed a robust metric to accurately quantify the degree of subclonal complexity for each sample based on discrete copy number variations (CNVs) profiles.

Results: We found that tumor cells in the high-complexity group originated from the cell lineage with FGB overexpression and exhibited high levels of transcription factors associated with poor survival. In contrast, tumor cells in low-complexity patients showed activation of more hallmark signaling pathways, more active cell-cell communications within the TME and a higher immune activation status. Additionally, cytokines signaling activity analysis suggested a link between *HMGB1* expressed by a specific endothelial subtype and T cell proliferation.

Discussion: Our study sheds light on the intricate relationship between the complexity of subclonal structure and the TME, offering novel insights into potential therapeutic targets for HCC.

Introduction

Hepatocellular carcinoma (HCC) is the third most lethal cancer in the world, with patients often experiencing post-surgical recurrence and metastasis, leading to poor survival outcomes (Wang et al., 2021). Tumor heterogeneity is the main cause of drug resistance and treatment failure (Dagogo-Jack and Shaw, 2018; Marusyk et al., 2020; Vasan et al., 2019), which encompasses the cellular diversity and dynamics both among tumors in different patients and within individual tumor. The tumor heterogeneity is manifested through genomic mutations, aberrations in epigenetic modifications, transcriptional alterations, and changes at the protein level (Sun and Yu, 2015). Additionally, extrinsic factors such as

hypoxia, pH levels, and the interactions between tumor cells and other stromal components within the tumor microenvironment (TME) contribute to the diversity and dynamics in tumor genotypes and phenotypes (Sun and Yu, 2015). Previous studies, such as those conducted by TCGA and ICGC, have integrated multi-omics profiles to delineate the heterogeneous molecular features of cancers and explored the subtyping of cancers, guiding personalized therapeutic strategies (Cancer Genome Atlas Research, 2013). Tumor heterogeneity also underpins tumor evolution, as tumor cells continuously accumulate genetic and epigenetic variations, forming heterogeneous subclones which possess different survival fitness and undergo dynamic processes of subclonal selection and elimination under selective pressures such as nutrient availability and metabolism (Davis et al., 2017; Greaves and Maley, 2012). Therefore, understanding the complexity of subclonal structure within tumor is crucial for elucidating tumor biology, disease progression, and response to therapy.

The recent advancements in single-cell sequencing technology (Kashima et al., 2020; Yu et al., 2021) have provided unprecedented insights into the composition, function, and spatial organization of immune and stromal cells within the tumor microenvironment (TME), leading to a deeper understanding of its complexity and diversity. In the process of tumor progression, tumor heterogeneity continuously remodels the tumor microenvironment by altering the transcription levels of target genes in non-malignant cells within the TME, which leads to divergences in developmental trajectories, immune landscapes, and intercellular networks (Wu et al., 2021; Leader et al., 2021; Zhang A. et al., 2022; Voit et al., 2023). Specifically, heterogeneous secretion of cytokines by tumor cells can also modulate immune-tumor interactions, as demonstrated by studies in mouse models (Liu et al., 2013; Knoche et al., 2021). Tumor cells can induce profound phenotypic changes in non-immune stromal components within the TME, impacting the immune components and their activities. For instance, oncogenic BRAFV600E signaling in human melanoma cells has been shown to interfere with T cell-mediated anti-tumor responses by modulating the phenotype of cancer-associated fibroblasts (Khalili et al., 2012; Binnewies et al., 2018). Furthermore, immune cells within the microenvironment can eliminate clones with high mutational burdens (high immunogenicity) through the process of immune editing, maintaining a balance that ultimately favors the selection of tumor variants capable of evading immune surveillance (Tsai et al., 2023; Polyak et al., 2009; Nam et al., 2021; Dunn et al., 2002; O'Donnell et al., 2019). Thus, the genotype and phenotype of tumor cells and the TME are inextricably linked. However, the intricacies of their interactions and how the subclonal complexity reshapes the TME remain largely unexplored.

In this study, we performed an integrated single-cell analysis of HCC patients with untreated primary tumors. We developed a metric to quantify the tumor subclonal complexity based on Shannon entropy theory. Our investigation focused on systematically elucidating the relationship between the subclonal complexity and cancer cell state, the functional activity of the immune cells. We found that cell-cell interaction landscape and TME polarization varied with subclonal complexity. Our findings underscore the significance of subclonal complexity in modulating the tumor microenvironment and its implications for immune responses.

Materials and methods

scRNA-seq datasets

The datasets of HCC were acquired from GEO, including 4 scRNA-seq cohorts (GSE156625 (24), GSE149614 (25), GSE151530 (26), GSE189903 (27)). Specifically, GSE156625 (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA658541&o=acc_s%3Aa) was the raw sequencing data and the other 3 datasets are processed read count matrices. To accurately analyze the impact of tumor subclonal complexity on the remodeling of the tumor microenvironment, we established the following criteria for sample selection: 1) all samples are from the 10X Genomics platform; 2) availability of raw sequencing data or read count expression matrices; 3) samples are untreated; 4) samples are primary tumors without invasion, recurrence and metastasis; 5) for multi-region sequencing data, we selected samples from the tumor core region whenever possible. Finally, only 30 samples were retained for further analysis (Supplementary Table S1).

Pre-processing of scRNA-seq data

Raw reads from GSE156625 were processed to generate gene expression matrices by using the standard internal pipeline based on the Cell Ranger toolkit (3.1.1). Then the expression matrices of each individual sample were converted to Seurat object by the "CreateSeuratObject" function from Seurat package (4.3.0) (Hao et al., 2021). We firstly filtered out cells that had either lower than 200 or higher than 8,000 expressed genes. Furthermore, the cells with the percent of mitochondrial genes over 40% of total expressed genes were discarded. Additionally, we initially removed doublet cells by DoubletFinder (2.0.3) (McGinnis et al., 2019) to avoid the potential influence on analytical results. A total of 67,125 cells were retained for further analysis.

Dimensionality reduction, clustering, DEGs identification, and major cell type annotation

To better reflect the biological features from the data, we used the "NormalizeData" function to normalize the expression matrices for the Seurat object. And the top 2,000 highly variable genes (HVGs) identified by "FindVariableFeatures" function were used to scale data through "ScaleData" function. The principal component analysis (PCA) was carried out by "RunPCA" function to generate 50 PCs, followed by "RunHarmony" function implanted in Harmony package (0.1.1) (Korsunsky et al., 2019) with the top 30 PCs to eliminate batch effect. The clustering analysis was implemented depended on the integrated joint embedding produced by Harmony with the Louvain algorithm after computing a shared nearest-neighbor (SNN) graph. The UMAP technique was used to create a 2D map on which the discovered clusters were displayed. Then we used "FindAllMarkers" function to identify over-expressed genes in specific cluster when compared with the other clusters (adjusted P-value <0.05, only.pos = T and logfc.threshold = 0.25). The well-known cell markers (Ma et al., 2021; Liu et al., 2023) were used to annotate the clusters: Endothelial (*ENG*, *VWF*, *PECAM1*), Fibroblasts (*ACTA2*, *COL1A2*, *PDGFRB*),

T cells (*CD2, CD3D, CD3E, CD3G*), B cells (*CD79A, CD79B, MS4A1, BANK1*), Myeloid (*CD14, CD68*), Plasma (*MZB1, DERL3, SDC1*), Hepatocytes (*APOA2, ALB, APOA1, AMBP*), proliferative T (*TOP2A, MKI67, TUBB*).

Cell subtype annotation

We further clustered T cells, endothelial cells and myeloid cells following a similar pipeline as described above, including normalization, highly variable genes identification, dimensionality reduction, batch effect correction with Harmony and clustering. Within each cell subtype, an iterative procedure that a cell cluster displaying at least two canonical marker expression of 7 major cell type was determined as doublet cells to be removed, if any, and reclustered the remaining cells, was adopted to ensure the reliability of the study. Specifically, we used the following well-known markers (Cheng et al., 2021; Zheng et al., 2021) for subtype identification. NK (*GNLY, KLRF1, NKG7*), CD4+_CCR7 (*CD4, CCR7*), CD4+_IFNG+ (*CD4, IFNG*), CD4+_FOXP3_Treg (*CD4, FOXP3, IL2RA*), CD8+_GZMK (*CD8A, GZMK*), CD8+_TNFSF9 (*CD8A, TNFSF9*), macrophage (*C1QC, APOE, CD68*), monocyte (*FCN1, LYZ*), cDC1_CLEC9A (*CLEC9A, FLT3*), cDC2_CD1C (*CD1C, FCER1A*), cDC3_LAMP3 (*LAMP3, CCR7*), mast (*KIT*). Similarly, endothelial cell were also confirmed by canonical markers (Goveia et al., 2020; Schupp et al., 2021), such as activated_PCV (*ACKR1, SELP, VCAM1, POSTN*), Capillaries_EDNRB (*EDNRB, CA4, HPGD, IL1RL1*), Capillaries_FCN3 (*FCN3, BTNL9, NOSTRIN, EDN*), arteries_GJA5 (*GJA5, EFNB2, SOX17, DKK2*), tip_CXCR4 (*CXCR4, ADM, ANGPT2, APLN*), RBP4+_EC (*RBP4, IGFBP7*), CXCL10+_EC (*CXCL10, GBP1, SOD2*). To accurately annotate proliferative T cells, we calculated the averaged Pearson correlation coefficient and averaged *p*-value for each proliferative T cell against other T cell subtypes and NK subtypes based on single-cell expression profiles. Proliferative T cells were excluded if the mean correlation coefficients were all less than 0.4 or the mean *p*-values all exceeded 0.05. Otherwise, they were assigned to the subtype with the highest average correlation.

CNVs estimation and malignant cells identification

We inferred copy number variations (CNVs) for each cell through the InferCNV (1.10.1) package of R. The non-malignant cells were applied as the reference to estimate the CNVs values of the hepatocytes (observation). The genes were sorted by their genomic locations on each chromosome. Specifically, The InferCNV analysis was implemented with parameters "denoise = T, HMM = T, analysis_mode = "subclusters", cutoff = 0.1, cluster_by_groups = T".

To identify the malignant cells from the hepatocytes, we designed the CNVs score quantifying the degree of CNVs fluctuation for each cell as follows:

$$CNV_{score} = \sqrt{\sum_{i=1} (X_i - X_{mean})^2 \Big/ n}$$

where X_i represented the CNVs value of gene *i* in the cell, X_{mean} was the mean CNVs value of gene *i* in the cell, and n was the number of the genes in the cell. If the CNVs score of a hepatocyte exceeded the

threshold determined by the intersection of the distributions of CNVs score in hepatocytes and non-malignant cells, then the hepatocyte was considered malignant (Supplementary Figure S1C).

Identification of intratumor NMF programs

The non-negative Matrix Factorization (NMF) algorithm was applied to identify the underlying expression programs from the malignant cells based on the NMF R package. We employed NMF (rank = 2:6, nrun = 10) to the relative expression matrices of the top 2000 HVGs in each sample with all negative values converted to 0. We chose the optimal *rank* value at which the cophenetic coefficient started producing the maximum descent (Supplementary Figure S1E). In total, we identified 98 programs from 30 samples (Supplementary Table S2). The 30 genes of each expression program with the top NMF scores were input into "AddModuleScore" function to generate the program score for malignant cells in each sample. We calculated the correlations between the 98 program scores in each sample individually. After excluding clustering groups containing less than half of samples, we finally extracted 3 meta-programs from the 98 programs according to hierarchical clustering of averaged correlations of pairs of programs across all samples (Figure 1F; Supplementary Table S3). Each meta-program, retaining 30 genes with the highest average NMF score, was used to perform pathway enrichment analysis.

CNV complexity score based on single cell CNVs profiles

Shannon entropy was used to quantify the CNV complexity for each sample based on discrete CNVs profiles inferred from InferCNV. Firstly, we calculated the frequency of copy number gains and losses for each gene within the tumor as follows:

$$LVF = n_{loss} \big/ n$$
$$GVF = n_{gain} \big/ n$$

where n represented the total number of malignant cells, n_{loss} was the number of cells in which the copy number of the gene was lost, and n_{gain} was the number of cells in which the copy number of the gene was gained. Specifically, Discrete CNVs value of the gene greater than (less than) 3 was considered to be gained (lost). The frequency of copy number gain and loss were divided into 10 equal-length frequency intervals separately (i.e., Loss:0–0.1, 0.1–0.2, ..., 0.9–1; Gain:0–0.1, 0.1–0.2, ..., 0.9–1). After mapping all the genes to equal-length frequency intervals according to the *LVF* and *GVF* of the genes, we then counted the number of genes belonging to each equal-length frequency interval and calculated the frequency of the number of genes in each interval. Finally, the CNV complexity score could be calculated for each sample as follow:

$$CNV\_complexity\_score = -\sum_i P(X_i)log_2 P(X_i)$$

where i was the *i*th interval, X_i was the number of genes falling into the *i*th interval, and $P(X_i)$ was the frequency of the number of genes in the *i*th interval.

FIGURE 1
Tumor heterogeneity landscape of HCC. **(A)** UMAP plot of all 66609 cells from 30 samples. Cells were annotated based on the well-known markers.
(B) UMAP plot of 30955 malignant cells from 30 samples. Cells were colored by samples. **(C)** UMAP plot of 34777 non-malignant cells from 30 samples.
Cells were colored by sample. **(D)** Distributions of pairwise correlation within intra-tumor and inter-tumor. **(E)** Pairwise correlation of all malignant cells
from 30 samples. Red represents positive correlation and blue represents negative correlation. **(F)** Heatmap of average correlations across
30 samples between pairs of expression programs. **(G)** Dot plot displaying significantly enriched pathways for genes of each meta-program. The size of
the dot represented the number of genes in the indicated hallmark (the larger, the more genes included) and the color shade of the dot indicated the *P*
value after FDR correction (the yellower, the smaller adjusted *P* value).

Simulation data and method evaluation

To evaluate whether the CNV complexity score could accurately distinguish between polyclonal and monoclonal samples, we used

computer simulations to generate single-cell CNV profiles for 100 polyclonal samples and 100 monoclonal samples as the ground truth data, respectively. Briefly, the size of cell population $N \sim U [200, 2000]$ and the total number of clones $C \sim U [5, 20]$ were

generated for each sample. For each clone, the size of clone was generated through binomial distribution B (N, p). For monoclonal samples, we assumed that $p = 0.9$ for one dominant clone and $p = 0.1/(c-1)$ for other rare clones. We considered int (N/2) subclones in the tumor cells, where int takes the nearest integer in the polyclonal samples. We used $p = 2/N$ to generate the size of subclone. For detailed steps of the entire simulation data generation process, please refer to Wang et al. (2023). We then applied the three methods (CNV complexity score, Guo et al.'s and Ma et al.'s) to the simulated single-cell CNV profiles which had been labeled as either monoclonal or polyclonal to calculate scores. A higher score for this sample means that the algorithm is more likely to predict the sample as a polyclonal sample. Finally, we used R package pROC (Robin et al., 2011) to calculate precision, recall, and the area under the receiver operating characteristic curve (AUC).

Gene set enrichment analysis, pathway enrichment analysis and ssGSEA

Gene set enrichment analysis (GSEA) and pathway enrichment analysis were performed using R package clusterProfiler (Yu et al., 2012). The ssGSEA algorithm was used to calculate the activity score of the gene sets for each cell based on GSVA package (Hanzelmann et al., 2013). We obtained 13 cancer functional states from CancerSEA (Yuan et al., 2019), as well as a collection of 50 cancer Hallmark gene sets downloaded from MsigDB (Liberzon et al., 2015). The signature gene lists of macrophages (e.g. Angiogenesis, Phagocytosis, MHC I molecules, MHC II molecules; Supplementary Table S5) and T cells (e.g. Cytotoxic, Exhaustion, Terminally exhaustion, Proliferation, TCR signaling pathway, Type II interferon response; Supplementary Table S6) were collected from previous studies (Rooney et al., 2015; Long et al., 2022; Li et al., 2020; Jiang et al., 2021; Zhang Y. et al., 2022).

CytoTRACE and pseudotime analysis

The R package CytoTRACE v.0.3.3 (47) was applied to predict the differentiation state of cells from the single-cell RNA-seq (scRNA-seq) profiles. The CytoTRACE score range from 0 to 1, with higher scores being associated with greater stemness (less differentiation) and *vice versa*, which help us determine the starting point of differentiation. We employed Monocle2 (Qiu et al., 2017) to discover the differentiation trajectory for T cells and endothelial cells. The top 2,000 DEGs between the cell subgroups were used to create the DDRTree. Utilizing the root state determined by the CytoTRACE, each cell was given a pseudotime value using the "order_cells" function. The cells were ordered and visualized with the "plot_cell_trajectory" function.

TF regulatory network construction and critical TFs identification

The SCENIC (Aibar et al., 2017) was used to construct TF regulatory network for malignant cells in different subclonal complexity. Briefly, GRNBoost2 estimated the co-expression

network and RcisTarget was used to identify the regulons. The regulon activity was then measured by AUCell for every cell. The differently activated TFs regulons between different subclonal complexity were identified by wilcox.test(). The TF-gene regulatory network was visualized by Cytoscape (Shannon et al., 2003).

Evaluation of associations between critical TFs and survival outcomes

To explore the prognostic power of TFs, we downloaded two gene expression datasets from cBioPortal (Gao et al., 2013) and GEO (GSE76427) for HCC patients, which contained clinical information. HCC patients from TCGA were treated as a training dataset, and HCC patients from GSE76427 were used as an independent validation dataset. A univariate Cox proportional hazards regression analysis was performed to evaluate the association between the critical TFs and patients' OS. Only TFs with p-value <0.05 were selected to further conduct variable selection according to stepwise Cox proportional hazards regression analysis. Finally, we created a risk-score formula based on TFs expression weighted by Cox regression coefficients.

$$Risk\ score = \sum_{i}^{N} (coef_i \times expr_i)$$

where N was the number of selected TFs, $expr_i$ was the expression value of the ith gene, and $coef_i$ was the Cox regression coefficient of the ith gene in the univariate Cox regression analysis. Patients were divided into high-risk and low-risk group based on the median risk score. KM curves were performed to compare the OS and DFS between two groups. To evaluate whether risk score could be an independent prognostic factor, Multivariate Cox proportional hazards regression model was conducted with risk score, age, sex, height, weight, race as covariates.

Cell-cell communications analysis

We applied CellPhoneDB (Efremova et al., 2020) to infer cell-cell communication between different cell types. The specific ligand-receptor interactions between cell types were identified based on permutation test. The interaction score represented the total mean of the individual ligand-receptor partner average expression value in the corresponding interacting pairs of cell types. To predict ligands driving the transcriptomic changes of target cells, we further performed NicheNet analysis (Browaeys et al., 2020) between RBP4+_EC and T cells with the "nichenet_seuratobj_aggregate" function. Specifically, the differentially expressed genes (DEGs) among T cells between different subclonal complexity with high complexity group as control were used as gene sets of interest and all expressed genes in T cells were used as background of genes.

Cytokines signaling activity calculation

We used CytoSig algorithm (Jiang et al., 2021) to quantify the signaling activity of cytokines for T cells. The CytoSig was a linear

model predicting cytokines signaling activity score for each cell and covered 51 cytokines provided by the curated training data. We implemented CytoSig for single-cell transcriptional profiles of T cells in Python 3.

Statistical analysis

Continuous variables were compared by the Wilcoxon rank-sum test and categorical variables were compared by the chi-squared test. Pearson correlation was calculated by cor () and cor. test (). The log-rank test was used to compare Kaplan–Meier curves. When $p <$ 0.05, the differences were considered to be significant. All statistical analyses were conducted using R software version 4.1.2 (http://www.rproject.org).

Results

The transcriptional diversity analysis reveals transcriptional programs in primary HCC

We downloaded 30 untreated primary HCC samples from four cohorts of single-cell RNA-seq data (Sharma et al., 2020; Lu et al., 2022; Ma et al., 2021; Ma et al., 2022) (Supplementary Table S1). After quality control, we obtained a total of 66,609 high-quality cells with a median of 2,220 cells per sample. Cell types were annotated by canonical cell markers, identifying 31,832 hepatocytes and 34,777 non-malignant cells (Figure 1A; Supplementary Figure S1A). We inferred copy number variations (CNVs) based on scRNA-seq data using InferCNV((Project. iotTC)) (Supplementary Figure S1B). Malignant and non-malignant hepatocytes were distinguished by their accumulated CNV scores (see Section Methods, Supplementary Figure S1C), ultimately identifying 30,955 malignant cells.

We performed clustering and UMAP visualization separately for malignant and non-malignant cells. Consistent with previous findings, non-malignant cells exhibited distinct separation by cell type but showed a mixture of different samples (Supplementary Figure S1D; Figure 1B), whereas malignant cells exhibited high inter-tumor heterogeneity (Figure 1C) (Puram et al., 2017). The transcriptomic similarity among inter-tumor cells was lower than that among intra-tumor cells (Figure 1D). However, the similarity of intra-tumor malignant cells varied across patients; for instance, HCC04T showed relatively high similarity, while H72 exhibited lower similarity among malignant cells (Figure 1E).

To identify underlying transcriptional programs, we applied non-negative matrix factorization (NMF) (Gaujoux and Seoighe, 2010) to each sample. The identified gene programs tend to be mutually exclusive among malignant cells within sample, and were negatively correlated (Supplementary Figure S1E–G). A total of 98 gene programs were identified across the 30 tumor samples (Supplementary Table S2). We then performed hierarchical clustering on these 98 expression programs, extracting three meta-programs (Figure 1F; Supplementary Table S3) that encompassed more than half of the samples. These three meta-programs were associated with distinct signaling pathways: Meta-program 1 was enriched in cell cycle-related pathways, Meta-

program 2 was enriched in metabolic pathways, and Meta-program 3 was enriched in stress response-related pathways (Figure 1G).

Quantifying intra-tumor subclonal complexity using CNV profiles

To explore intra-tumor subclonal complexity, we proposed a metric to quantify CNV complexity based on the discrete CNVs profiles inferred from InferCNV (Figure 2A, method). We quantified CNV complexity for each HCC sample and classified samples into high (H) and low (L) groups based on the median CNV complexity score. We found that CNV complexity scores were significantly correlated with two other heterogeneity scores proposed by (Ma et al., 2019) and (Guo et al., 2022) respectively (Figures 2B, C). Moreover, the CNV complexity score was more robust than the other two heterogeneity scores, both of which were negatively correlated with the number of the cells (Figures 2D, E). Simulation data further demonstrated the robustness of the CNV complexity score (Supplementary Figure S2A).

Finally, we evaluated the three methods on simulated datasets generated *in silico* (Supplementary Figure S2B, C). The CNV complexity score showed excellent performance in reflecting the degree of subclonal complexity, as demonstrated by a receiver operating characteristic (ROC) curve with the highest AUC value, which was superior to the other methods (Figure 2F). Overall, our constructed CNV complexity score demonstrates excellent stability and sensitivity, and more importantly, it is not influenced by the number of cells, providing a more accurate quantification of subclonal complexity in samples.

Phenotypic regulatory plasticity in HCC samples with varying subclonal complexity

To explore the phenotypic plasticity underlying different subclonal complexity, we used cancer functional gene sets (Supplementary Table S4) along with ssGSEA and identified 37 gene sets with significant differences between the H and L groups, with 34 functional gene sets being significantly upregulated in the L group ($p < 0.05$, Wilcoxon rank-sum test, Figure 3A; Supplementary Figure S3). For instance, the L group was significantly enriched for hypoxia, invasiveness, metastasis, emphasizing varying signaling activity between the H group and L group. Malignant cells are susceptible to epithelial-mesenchymal transition under hypoxic conditions, gaining invasive and metastatic abilities, accompanied by a transition from low to high stemness. Consistent with hypoxia, invasion and metastasis activities mentioned above, it was discovered that the stemness of malignant cells in L group was also significantly higher than that in the H group ($p = 0.0049$, Wilcoxon rank-sum test, Figure 3B) (Gulati et al., 2020), which highlights that stemness of malignant cells is closely related to hypoxia, invasiveness and metastasis.

We further investigated the cell lineage origin of malignant cells using scHCL(Han et al., 2020), and found significantly different distributions of cell lineages between the H and L group ($p < 0.05$, chi-squared test, Figure 3C), with a significantly higher proportion

FIGURE 2
Construction and validation of the robust CNV complexity score. **(A)** Overview of the novel scoring strategy based on Shannon entropy theory. **(B)** Comparative analysis of our method and two other strategies (Ma et al. and Guo et al.) in the real discrete CNV profiles inferred by InferCNV. **(C)** Correlation analysis of CNV complexity score with Ma et al.'s score (left) and Guo et al.'s score (right) respectively. **(D)** Scatterplot showing the correlation between the number of cells and Ma et al.'s score (left), Guo et al.'s score (right). Each dot represents a sample. **(E)** Scatterplot showing the correlation between the number of cells and the CNV complexity score. Each dot represents a sample. **(F)** ROC curve of CNV complexity score, Ma et al.'s score and Guo et al.'s score in the simulated datasets.

of cell lineage overexpressing FGB in the H group ($p < 0.05$, Wilcoxon rank-sum test, Figure 3D). FGB was significantly upregulated in malignant cells from the H group ($p < 2.22e-16$, Wilcoxon rank-sum test, Supplementary Figure S4A). Altered expression of this gene leads to several disorders, including

afibrinogenemia, dysfibrinogenemia, and thrombotic tendency. In summary, malignant cells from samples with varying CNV complexity might undergo diverse transcriptional reprogramming, resulting in different cancer functional states and cell lineage origins.

FIGURE 3
Identification of subclonal complexity-specific TFs. **(A)** Heatmap of the ssGSEA score of 37 cancer functional states significantly differed between L and H groups. Two-sided Wilcoxon rank-sum test. **(B)** Boxplot showing the CytoTRACE score of the L and H groups. The CytoTRACE score for each sample was the averaged CytoTRACE score of malignant cells from the sample. Two-sided Wilcoxon rank-sum test. **(C)** Histogram indicating the proportion of cells lineages in malignant cells from L and H groups. Chi-squared test. **(D)** Boxplot showing the fraction of cell lineages in L and H groups. *p < 0.05; **p < 0.01; ***p < 0.001; NS not significant. **(E)** Heatmap showing the TF activity for malignant cells derived from L and H groups. **(F)** Kaplan-Meier plots of risk score on OS (left) and DFS (right). **(G)** Forest plot showing the HR (95% CI) for OS. The multivariate Cox proportional hazard models. HR, hazard ratios; CI, confidence interval.
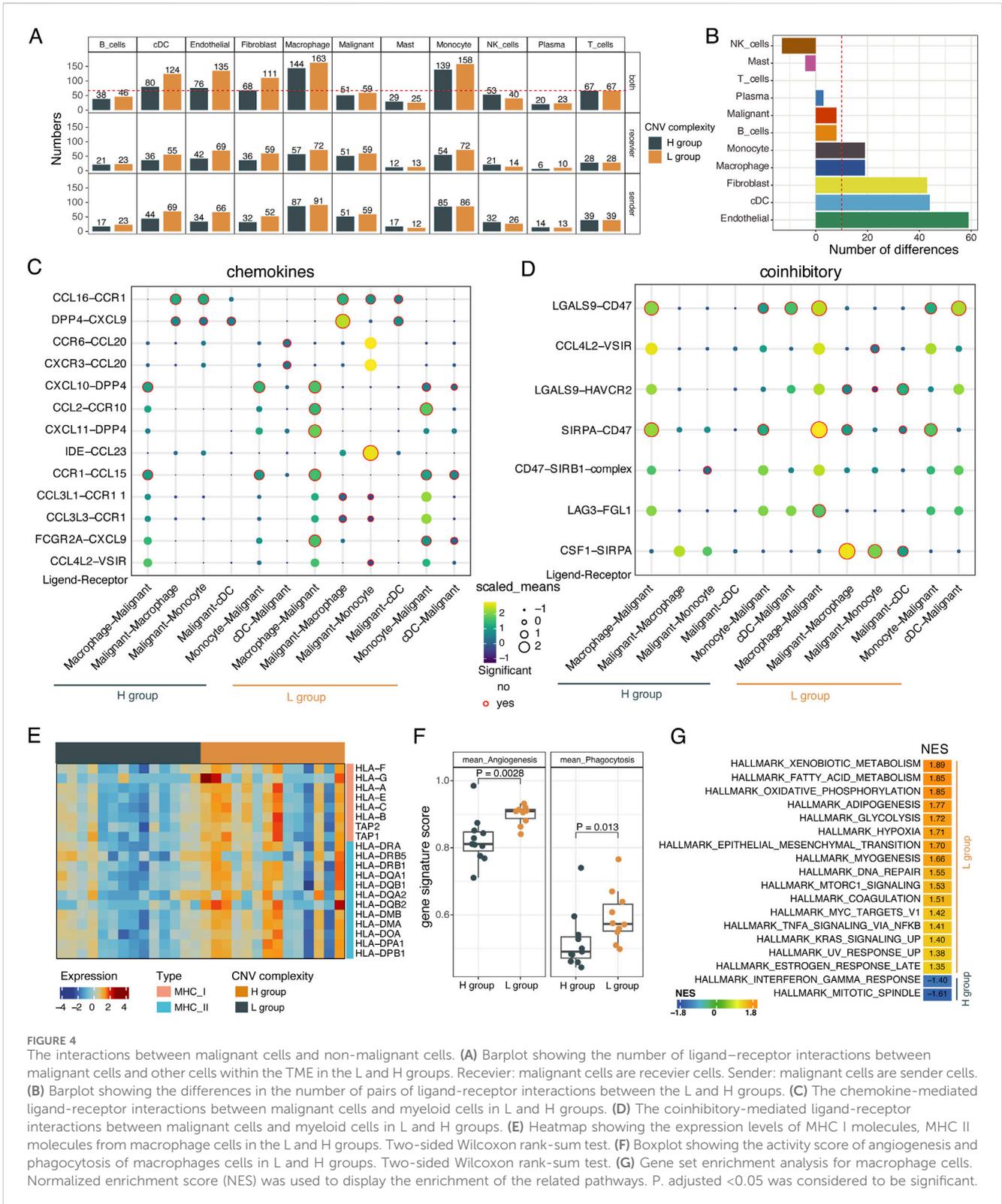
FIGURE 4
The interactions between malignant cells and non-malignant cells. **(A)** Barplot showing the number of ligand–receptor interactions between malignant cells and other cells within the TME in the L and H groups. Receiver: malignant cells are receiver cells. Sender: malignant cells are sender cells. **(B)** Barplot showing the differences in the number of pairs of ligand-receptor interactions between the L and H groups. **(C)** The chemokine-mediated ligand-receptor interactions between malignant cells and myeloid cells in L and H groups. **(D)** The coinhibitory-mediated ligand-receptor interactions between malignant cells and myeloid cells in L and H groups. **(E)** Heatmap showing the expression levels of MHC I molecules, MHC II molecules from macrophage cells in the L and H groups. Two-sided Wilcoxon rank-sum test. **(F)** Boxplot showing the activity score of angiogenesis and phagocytosis of macrophages cells in L and H groups. Two-sided Wilcoxon rank-sum test. **(G)** Gene set enrichment analysis for macrophage cells. Normalized enrichment score (NES) was used to display the enrichment of the related pathways. P. adjusted <0.05 was considered to be significant.

To investigate the regulatory mechanisms contributing to the different cell lineages and functional phenotypes of tumor cells between the H and L group, we employed SCENIC to identify transcription factors (TFs) regulatory modules (Figure 3E). We then constructed TFs regulatory networks corresponding to the H and L groups using differentially expressed and differential CNV genes targeted by identified TFs (Supplementary Figure S4B; Supplementary Figures S5A, B). Functional enrichment analysis of the TF regulatory network revealed that the upregulated genes in the L group's TFs regulatory network were enriched in pathways

such as hypoxia, apoptosis, TNFA signaling via NFkB and regulation of Wnt signaling. In contrast, the upregulated genes in the H group were involved in pathways regulating epithelial cell proliferation and metabolism-associated pathways (Supplementary Figure S4C).

Among these TFs, we identified four transcription factors *MAFG, YY1, EZH2, AR,* associated with overall survival (OS) which were upregulated in the H group. Samples were stratified into high-risk and low-risk groups based on the median risk score quantified by these four TFs. Kaplan-Meier curve analysis demonstrated poorer survival in the high-risk group compared to the low-risk group in terms of OS and disease-free survival (DFS) (OS: $p < 0.0001$, DFS: $p < 0.0001$, log-rank test, Figure 3F). Further multivariate Cox proportional hazards regression analysis confirmed the independent prognostic value of the risk score (Figure 3G; Supplementary Figure S4D). Finally, we validated the prognostic power of the 4-TF signatures in an external validation dataset (GSE76427). The KM curves of the two groups were significantly different ($p = 0.0016$, log-rank test, Supplementary Figure S4E) and the risk score showed the independent prognostic value (Supplementary Figure S4F).

Characterizing cell-cell communications in the tumor microenvironment across different subclonal complexity

To investigate the characteristics of cell-cell communications within the tumor microenvironment under different complexity of subclonal structure, we performed cell-cell interaction analysis using CellPhoneDB (Efremova et al., 2020) between malignant cells and non-malignant cells in the H and L groups, respectively. We found that fibroblast and endothelial cells strongly interacted with tumor cells in the L group rather than the H group (Figure 4A). Additionally, myeloid cells, including macrophage, monocytes and dendritic cells (Supplementary Figure S6A, B), showed strong interactions with tumor cells in both the H and L groups, with increased communications observed in the L group (Figure 4B).

We further explored the ligand-receptor interactions between tumor cells and myeloid cells in the L group and H group. We found that many chemokine-mediated interactions were specifically present in the L group, such as the *CCL4L2-VSIR, FCGR2A-CXCL9, CCL3L3-CCR1, CCL3L1-CCR11, IDE-CCL23,* and *CXCL11-DPP4, CCL2-CCR10* pathways which suggests that tumor cells in the L group might recruit more myeloid cells. In the H group, there were also a few specific interactions, such as the interactions between the dendritic cell-expressed ligands *CCR6* and *CXCR3* with the tumor cells-expressed receptor *CCL20* (Figure 4C). Additionally, we confirmed that coinhibitory interactions mediated by pathways like *CSF1-SIRPA, LAG3-FGL1, CCL4L2-VSIR,* and *LGALS9-HAVCR2* were specifically activated in the L group (Figure 4D). Similar results were also observed in the costimulatory-mediated interactions between malignant cells and myeloid cells (Supplementary Figure S6C). Interestingly, the L group was enriched for both coinhibitor- and costimulator-mediated ligand-receptor interaction pairs. In summary, these results revealed that varying subclonal complexity could alter the cell-cell interactions landscape.

Given the differential interaction strength of myeloid cells and tumor cells between the H and L groups, we investigated the

functional characteristics of myeloid cells. Myeloid cells from the L group showed upregulation of MHC I and MHC II molecules and higher angiogenesis and phagocytosis signature scores ($p < 0.05$, Wilcoxon rank-sum test, Supplementary Table S5; Figures 4E, F). Based on gene set enrichment analysis (GSEA) of the differentially expressed genes identified in myeloid cells between the L and H groups (Supplementary Figure S6D), more hallmark signaling pathways were enriched in myeloid cells from the L group, including hypoxia, EMT, TNFA signaling via NFkB, DNA repair, apoptosis, and angiogenesis (Figure 4G; Supplementary Figure S6E, F). These results were consistent with what we observed in tumor cells from samples with different complexity of subclonal structure.

Differential activation and functional states of T Cells with varying subclonal complexity

Although there was no significant difference in tumor-T cells interaction strength between the H and L groups, T cells exhibited significantly different activation statuses based on ssGSEA analysis of functional gene sets (Supplementary Table S6). Both CD4[+] and CD8[+] T cells from the L group had higher scores for TCR signaling pathway, exhaustion and proliferation but lower type II interferon response score ($p < 0.05$, Wilcoxon rank-sum test, Figure 5A). Moreover, cytotoxicity-related genes (e.g., *GZMA*, *GZMB*, *GZMH*, *NKG7*, *PRF1*) and immune checkpoint genes (e.g., *LAG3*, *PDCD1*, *LCK*) were significantly upregulated in T cells from the L group ($p < 0.05$, Wilcoxon rank-sum test, Supplementary Figure S7A, B). The scores of these functional gene sets were significantly correlated with CNV complexity score, implying the potential of subclonal complexity in predicting the activation status of T cells (Figure 5B).

We performed subclustering on T/NK cells, resulting in seven cell subpopulations: CD4_CCR7, CD4_IFNG+, CD4_FOXP3+_Treg, CD8_GZMK+, CD8+_TNSF9+, proliferative T cells, and NK cells (Figure 5C; Supplementary Figure S7C). The proliferative T cells were further annotated according to their similarity with other T subtypes and NK subtype (see methods, Figure 5D). Based on trajectory analysis of CD4[+] and CD8[+] T cells, we found that both CD4[+] and CD8[+] T cells from the L group tended to locate in the intermediate of the trajectory whereas those from the H group were primarily located at the terminal of trajectory (Figure 5E, F). Additionally, T cells from L group showed increased CytoTRACE score, indicating a heightened activation status of T cells in the L group (CD8+: $p = 0.043$, CD4+: $p = 0.075$, Wilcoxon rank-sum test, Figures 5E, F; Supplementary Figure S7D, E).

Immune activation mediated by endothelial cell subtypes with varying subclonal complexity

To investigate the regulatory mechanisms underlying the different activation states of T cells, we estimated the signaling activities of 51 cytokines on T cells and their correlation with T cell proliferation using the CytoSig model (Jiang et al., 2021). The activity of the cytokine *HMGB1* showed a significant positive
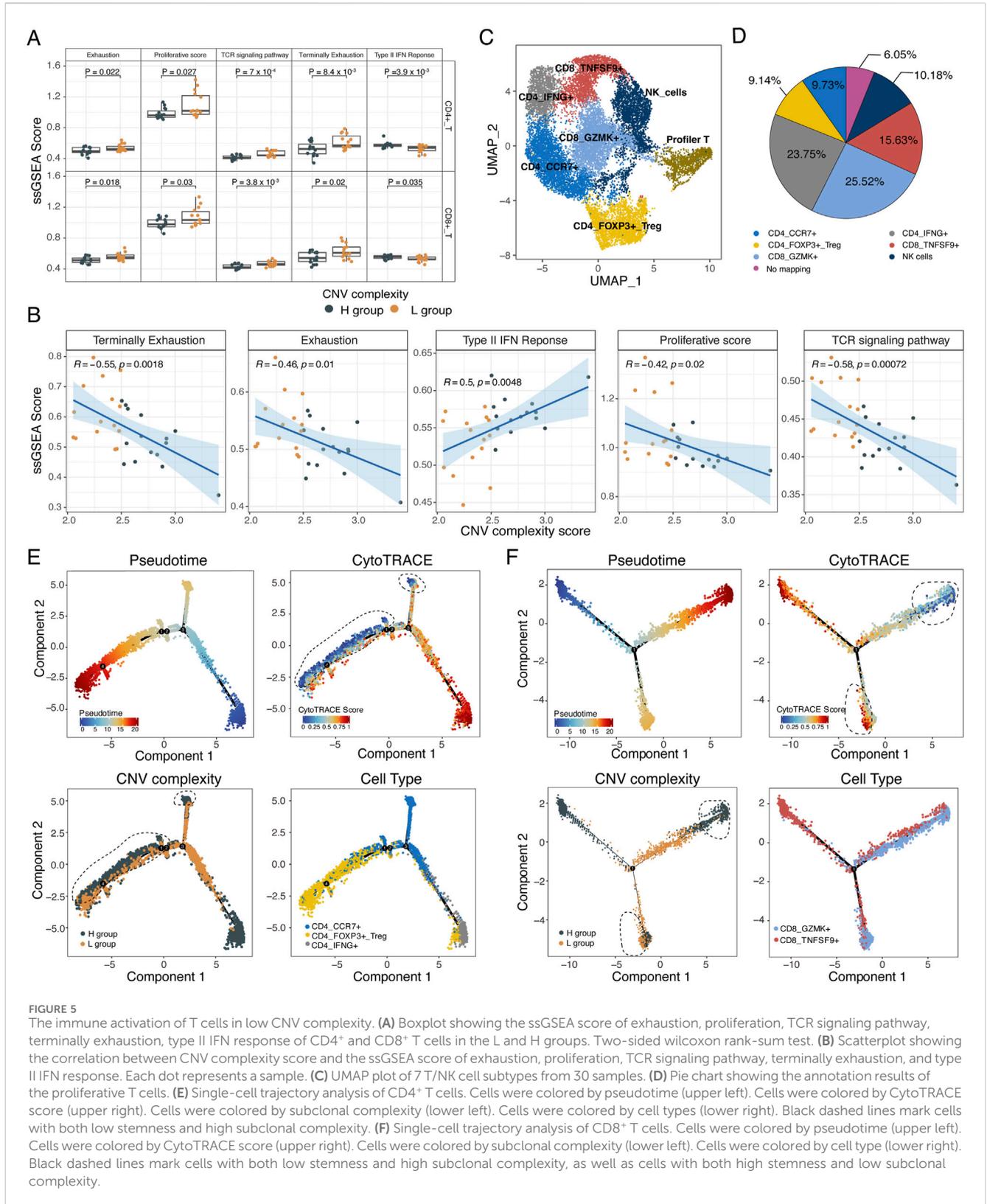
FIGURE 5

The immune activation of T cells in low CNV complexity. **(A)** Boxplot showing the ssGSEA score of exhaustion, proliferation, TCR signaling pathway, terminally exhaustion, type II IFN response of CD4+ and CD8+ T cells in the L and H groups. Two-sided wilcoxon rank-sum test. **(B)** Scatterplot showing the correlation between CNV complexity score and the ssGSEA score of exhaustion, proliferation, TCR signaling pathway, terminally exhaustion, and type II IFN response. Each dot represents a sample. **(C)** UMAP plot of 7 T/NK cell subtypes from 30 samples. **(D)** Pie chart showing the annotation results of the proliferative T cells. **(E)** Single-cell trajectory analysis of CD4+ T cells. Cells were colored by pseudotime (upper left). Cells were colored by CytoTRACE score (upper right). Cells were colored by subclonal complexity (lower left). Cells were colored by cell types (lower right). Black dashed lines mark cells with both low stemness and high subclonal complexity. **(F)** Single-cell trajectory analysis of CD8+ T cells. Cells were colored by pseudotime (upper left). Cells were colored by CytoTRACE score (upper right). Cells were colored by subclonal complexity (lower left). Cells were colored by cell type (lower right). Black dashed lines mark cells with both low stemness and high subclonal complexity, as well as cells with both high stemness and low subclonal complexity.

correlation with the proliferative score of T cells ($p = 0.014$, R = 0.45, Figure 6A). Moreover, *HMGB1* signaling activity was significantly higher in the L group compared to the H group ($p = 0.011$, Wilcoxon rank-sum test, Figure 6B). *HMGB1* was primarily expressed in endothelial and fibroblast cells (Figure 6C). However, the

expression levels of *HMGB1* in endothelial cells, rather than fibroblast cells, were upregulated in the L group ($p = 0.0023$, Wilcoxon rank-sum test) and significantly correlated with T cell proliferation (Figures 6D, E; Supplementary Figure S8A). Additionally, the expression levels of *HMGB1* in endothelial cells

FIGURE 6
HMGB1 drived the reprogramming of the TME. **(A)** Scatterplot showing the correlation between *HMGB1* activity and proliferative score of T cells. **(B)** Boxplot showing the *HMGB1* activity in L and H groups. Two-sided Wilcoxon rank-sum test. **(C)** Violin plot of expression levels of *HMGB1* in various cell subtypes, colored by the mean expression levels of *HMGB1* in cell subtypes. **(D)** Boxplot showing the expression levels of *HMGB1* in L and H groups. Two-sided Wilcoxon rank-sum test. **(E)** Scatterplot showing the correlation between the expression level of *HMGB1* and proliferative score of T cells. **(F)** UMAP plot of 7 endothelial cell subtypes from 30 samples. **(G)** Boxplots illustrating the fraction of endothelial subtypes in L and H groups. Two-sided Wilcoxon rank-sum test. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; NS not significant. **(H)** Boxplot showing the expression of *HMGB1* from RBP4_EC in the H and L groups. Two-sided Wilcoxon rank-sum test. **(I–K)** Single-cell trajectory analysis of endothelial cells. **(I)** Cells were colored by pseudotime. **(J)** Cells were colored by cell subtypes. **(K)** The changes in expression levels of *HMGB1* along with the pseudotime.

was positively correlated with the activity scores of T cell exhaustion and TCR signaling pathway, and negatively correlated with the activity score of type II interferon response score (Supplementary Figure S8B).

We further dissected the subtypes of endothelial cells, identifying seven distinct endothelial cell subtypes (Figure 6F; Supplementary Figure S8C). The proportion of *RBP4* positive endothelial cells (RBP4+EC) was significantly higher in the L group compared to the H group ($p < 0.05$, Wilcoxon rank-sum test, Figure 6G). Moreover, *HMGB1* expression which was present only in RBP4+ ECs was significantly increased in the L group compared to the H group ($p = 0.009$, Wilcoxon rank-sum test, Figure 6H; Supplementary Figure S8D). Trajectory analysis demonstrated that RBP4+ECs persisted throughout the differentiation trajectory and showed highly expressed *HMGB1* in the L group along with the pseudotime (Figures 6I–K). Moreover, CytoTRACE analysis suggested that RBP4+ECs in the L group

showed a less differentiated status ($p = 0.0079$, Wilcoxon rank-sum test, Supplementary Figure S8E–G).

To further investigate the regulatory mechanisms of RBP4+ECs on T cell activation, we identified regulators from RBP4+ECs for T cell activation through Nichenet analysis (Browaeys et al., 2020) (Figure 7A). We found that RBP4+ECs regulated the expression of 29 target genes in T cells by upregulating *HMGB1* in the L group, including *IFNG*, which enhances tumor immune function and promotes T cell exhaustion (Benci et al., 2019), and *CCL3*, which facilitates dendritic cells (DC) recruitment for antigen presentation and T cell activation (Castellino et al., 2006). Interaction analysis between RBP4+ECs and T cells showed more and stronger chemokine-mediated interactions between RBP4+ECs and CD8+T in the L group including *CXCR6-CXCL16*, *NR3C1-CCL2*, *CXCL12-CXCR3*, *CXCL12-CXCR4*, and *PGRMC2-CCL4L2* (Figure 7B). Similar results were observed in the interactions between RPB4+ECs and CD4+ T cells (Figure 7C).

FIGURE 7
The interactions between RBP4+_EC and T cells. **(A)** Heatmaps showing potential ligands driving transcriptomic changes of target genes in T cells. **(B,C)** The chemokine-mediated ligand-receptor interactions between RBP4+_EC and T cells in L and H groups. **(B)** CD8+ T cells. **(C)** CD4+ T cells.
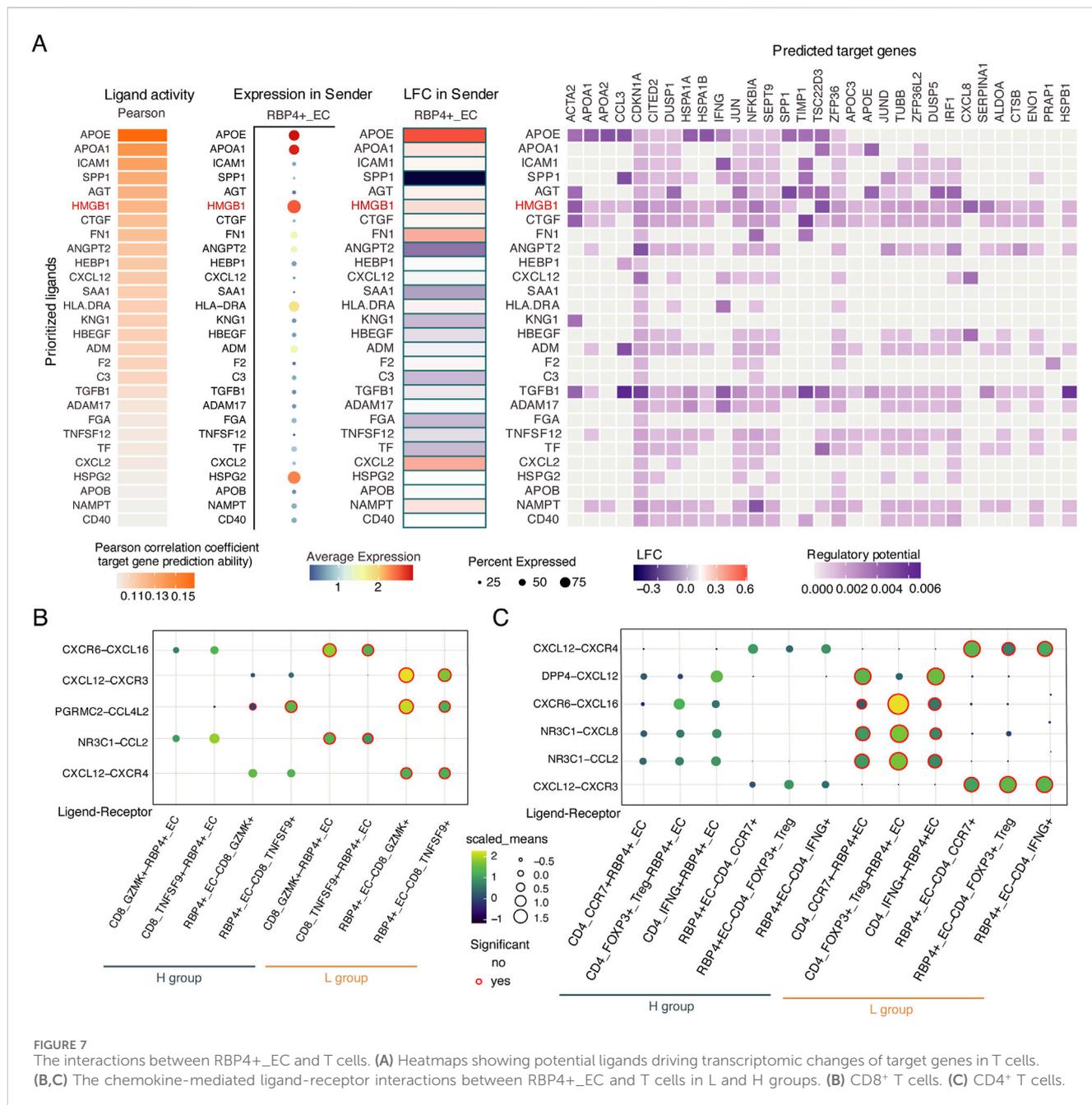
Discussion

Tumor heterogeneity pervades across various cancers, profoundly influencing the survival of patients (Dagogo-Jack and Shaw, 2018). In this study, we observed sample-specific clustering patterns among malignant cells, whereas non-malignant cells exhibited clustering patterns predominantly defined by cell types. The correlation analysis of malignant cells revealed that intra-tumor heterogeneity was variable among samples and intra-tumor cells similarity was generally greater than that of inter-tumor cells. Previous research has identified complex transcriptional programs across diverse malignancies, including cell cycle, interferon response, partial epithelial-mesenchymal transition (pEMT), and complete EMT (cEMT) (Barkley et al., 2022). Here,

we identified three meta-programs that were associated with distinct hallmark functions using NMF. Several strategies have been proposed to quantify the degree of tumor heterogeneity. For instance, Ma et al. (2019) employed PCA to project all malignant cells to the eigenvector space to derive PCs and calculated the averaged distance of malignant cells from the centroid (arithmetic mean of PCs of all malignant cells within the corresponding tumor) in principal component space. This averaged distance was used as a measure of heterogeneity for that particular sample. Guo et al. first calculated inter cell distances based on the Pearson correlation between any two malignant cells within a sample. The distance between cells was defined as D = 1- R, where R is the Pearson correlation coefficient between the CNV profiles of any two cells. Then, the metric of

heterogeneity was defined as the median distance of distribution of pairwise cell distances (Guo et al., 2022). However, these methods implicitly involved the cell number in the calculation of heterogeneity, thus making them potentially influenced by the number of cells. Therefore, we proposed a new metric that avoided the involvement and interference of cell number and accurately quantified the complexity of subclonal structure without bias from cell numbers.

Data originating from glioblastoma (Patel et al., 2014; Neftel et al., 2019), oligodendroglioma (Tirosh et al., 2016), head and neck cancer (Puram et al., 2017), and melanoma (Rambow et al., 2018) consistently demonstrated a spectrum of differentiation within tumors, ranging from stem-like or progenitor states to fully differentiated cells, alongside a plethora of heterogeneous functional states in malignant cells, such as stress responses, interferon reactions, and hypoxia (Puram et al., 2017; Patel et al., 2014; Neftel et al., 2019; Moncada et al., 2020; Yu et al., 2022). We identified a total of 37 significantly different cancer functional states between low and high-complexity groups. The low-complexity group was significantly enriched in functional states associated with hypoxia, invasiveness, metastasis, stemness, apoptosis, and DNA repair, suggesting their potential roles in tumor progression and therapeutic resistance, driven by highly competitive dominant clones (Davis et al., 2017; Greaves and Maley, 2012). Furthermore, an investigation into lineage origins revealed that distinct compositions of cellular lineages in samples with varying subclonal complexity. Additionally, the TFs regulatory networks associated with the low and high-complexity groups reveal a correlation between subclonal complexity and survival outcomes in HCC.

Previous research has established that tumor heterogeneity is a critical cornerstone for reconstructing tumor evolution and driving the dynamics and diversity of the tumor microenvironment (Dentro et al., 2021). Genetic and epigenetic alterations collectively define the transcriptomic and phenotypic heterogeneity of malignant cells, leading directly or indirectly to the reprogramming of the TME (Flavahan et al., 2017; Vitale et al., 2021; Alizadeh et al., 2015). In this study, we observed markedly distinct interaction strength between malignant cells and other cell types within the microenvironment in high- versus low-complexity group. Specifically, malignant cells exhibited a substantially higher number of ligand-receptor interactions with myeloid cells, fibroblasts, and endothelial cells in low-complexity group compared to the high-complexity group. Notably, endothelial cells displayed the largest differences between the two groups, suggesting a potential role for endothelial cells in shaping the TME according to the complexity of subclonal structure. Our subsequent analysis revealed a remodeled immunological landscape in the low-complexity group, characterized by enhanced antigen-presenting capabilities of macrophages and heightened T cells immunoreactivity.

Cytokines play an important role in remodeling the tumor microenvironment (Binnewies et al., 2018). In this study, we identified HMGB1, a critical cytokine expressed by RBP4+ECs, which exerts immunostimulatory effects on T cells. Our results demonstrated that both the expression and signaling activity of HMGB1 were significantly higher in the low complexity group compared to the high complexity group. A notable correlation was observed between the expression levels of HMGB1 in endothelial cells and immune activity score of T cells. We further confirmed that RBP4+ECs had more frequent chemokine-mediated ligand-receptor interaction pairs with T-cells in the low complexity group. These findings collectively highlighted the intricate interplay between the subclonal complexity and the TME, particularly emphasizing the role of subclonal complexity in orchestrating immunological landscape. The increased interaction strength between RBP4+ECs and T cells and activation of T cells in the low complexity group, mediated by HMGB1, suggests a potential mechanism through which the cytokines can either promote or suppress antitumor immunity. Future studies should focus on elucidating the molecular mechanisms underlying the varying expression of HMGB1 and its impact on T cells, potentially leading to the identification of novel biomarkers and therapeutic targets for cancer treatment.

Data availability statement

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2024.1467682/full#supplementary-material

References

Aibar, S., Gonzalez-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* 14 (11), 1083–1086. doi:10.1038/nmeth.4463

Alizadeh, A. A., Aranda, V., Bardelli, A., Blanpain, C., Bock, C., Borowski, C., et al. (2015). Toward understanding and exploiting tumor heterogeneity. *Nat. Med.* 21 (8), 846–853. doi:10.1038/nm.3915

Barkley, D., Moncada, R., Pour, M., Liberman, D. A., Dryg, I., Werba, G., et al. (2022). Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nat. Genet.* 54 (8), 1192–1201. doi:10.1038/s41588-022-01141-9

Benci, J. L., Johnson, L. R., Choa, R., Xu, Y., Qiu, J., Zhou, Z., et al. (2019). Opposing functions of interferon coordinate adaptive and innate immune responses to cancer immune checkpoint blockade. *Cell* 178 (4), 933–948. doi:10.1016/j.cell.2019.07.019

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., et al. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* 24 (5), 541–550. doi:10.1038/s41591-018-0014-x

Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* 17 (2), 159–162. doi:10.1038/s41592-019-0667-5

Cancer Genome Atlas Research, N. (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499 (7456), 43–49. doi:10.1038/nature12222

Castellino, F., Huang, A. Y., Altan-Bonnet, G., Stoll, S., Scheinecker, C., and Germain, R. N. (2006). Chemokines enhance immunity by guiding naive CD8+ T cells to sites of CD4+ T cell-dendritic cell interaction. *Nature* 440 (7086), 890–895. doi:10.1038/nature04651

Cheng, S., Li, Z., Gao, R., Xing, B., Gao, Y., Yang, Y., et al. (2021). A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* 184 (3), 792–809 e23. doi:10.1016/j.cell.2021.01.010

Dagogo-Jack, I., and Shaw, A. T. (2018). Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* 15 (2), 81–94. doi:10.1038/nrclinonc.2017.166

Davis, A., Gao, R., and Navin, N. (2017). Tumor evolution: linear, branching, neutral or punctuated? *Biochim. Biophys. Acta Rev. Cancer* 1867 (2), 151–161. doi:10.1016/j.bbcan.2017.01.003

Dentro, S. C., Leshchiner, I., Haase, K., Tarabichi, M., Wintersinger, J., Deshwar, A. G., et al. (2021). Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cell* 184 (8), 2239–2254.e39. doi:10.1016/j.cell.2021.03.009

Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J., and Schreiber, R. D. (2002). Cancer immunoediting: from immunosurveillance to tumor escape. *Nat. Immunol.* 3 (11), 991–998. doi:10.1038/ni1102-991

Efremova, M., Vento-Tormo, M., Teichmann, S. A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* 15 (4), 1484–1506. doi:10.1038/s41596-020-0292-x

Flavahan, W. A., Gaskell, E., and Bernstein, B. E. (2017). Epigenetic plasticity and the hallmarks of cancer. *Science* 357 (6348), eaal2380. doi:10.1126/science.aal2380

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal* 6 (269), pl1. doi:10.1126/scisignal.2004088

Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinforma.* 11, 367. doi:10.1186/1471-2105-11-367

Goveia, J., Rohlenova, K., Taverna, F., Treps, L., Conradi, L. C., Pircher, A., et al. (2020). An integrated gene expression landscape profiling approach to identify lung tumor endothelial cell heterogeneity and angiogenic candidates. *Cancer Cell* 37 (1), 421–436 e13. doi:10.1016/j.ccell.2020.03.002

Greaves, M., and Maley, C. C. (2012). Clonal evolution in cancer. *Nature* 481 (7381), 306–313. doi:10.1038/nature10762

Gulati, G. S., Sikandar, S. S., Wesche, D. J., Manjunath, A., Bharadwaj, A., Berger, M. J., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* 367 (6476), 405–411. doi:10.1126/science.aax0249

Guo, L., Yi, X., Chen, L., Zhang, T., Guo, H., Chen, Z., et al. (2022). Single-cell DNA sequencing reveals punctuated and gradual clonal evolution in hepatocellular carcinoma. *Gastroenterology* 162 (1), 238–252. doi:10.1053/j.gastro.2021.08.052

Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., et al. (2020). Construction of a human cell landscape at single-cell level. *Nature* 581 (7808), 303–309. doi:10.1038/s41586-020-2157-4

Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., 3rd, Zheng, S., Butler, A., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* 184 (13), 3573–3587.e29. doi:10.1016/j.cell.2021.04.048

Jiang, P., Zhang, Y., Ru, B., Yang, Y., Vu, T., Paul, R., et al. (2021). Systematic investigation of cytokine signaling activity at the tissue and single-cell levels. *Nat. Methods* 18 (10), 1181–1191. doi:10.1038/s41592-021-01274-5

Kashima, Y., Sakamoto, Y., Kaneko, K., Seki, M., Suzuki, Y., and Suzuki, A. (2020). Single-cell sequencing techniques from individual to multiomics analyses. *Exp. Mol. Med.* 52 (9), 1419–1427. doi:10.1038/s12276-020-00499-2

Khalili, J. S., Liu, S., Rodriguez-Cruz, T. G., Whittington, M., Wardell, S., Liu, C., et al. (2012). Oncogenic BRAF(V600E) promotes stromal cell-mediated immunosuppression via induction of interleukin-1 in melanoma. *Clin. Cancer Res.* 18 (19), 5329–5340. doi:10.1158/1078-0432.CCR-12-1632

Knoche, S. M., Larson, A. C., Sliker, B. H., Poelaert, B. J., and Solheim, J. C. (2021). The role of tumor heterogeneity in immune-tumor interactions. *Cancer Metastasis Rev.* 40 (2), 377–389. doi:10.1007/s10555-021-09957-3

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., et al. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* 16 (12), 1289–1296. doi:10.1038/s41592-019-0619-0

Leader, A. M., Grout, J. A., Maier, B. B., Nabet, B. Y., Park, M. D., Tabachnikova, A., et al. (2021). Single-cell analysis of human non-small cell lung cancer lesions refines tumor classification and patient stratification. *Cancer Cell* 39 (12), 1594–1609.e12. doi:10.1016/j.ccell.2021.10.009

Li, Y., Jiang, T., Zhou, W., Li, J., Li, X., Wang, Q., et al. (2020). Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. *Nat. Commun.* 11 (1), 1000. doi:10.1038/s41467-020-14802-2

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Liu, R. B., Engels, B., Schreiber, K., Ciszewski, C., Schietinger, A., Schreiber, H., et al. (2013). IL-15 in tumor microenvironment causes rejection of large established tumors by T cells in a noncognate T cell receptor-dependent manner. *Proc. Natl. Acad. Sci. U. S. A.* 110 (20), 8158–8163. doi:10.1073/pnas.1301022110

Liu, Y., Xun, Z., Ma, K., Liang, S., Li, X., Zhou, S., et al. (2023). Identification of a tumour immune barrier in the HCC microenvironment that determines the efficacy of immunotherapy. *J. Hepatol.* 78 (4), 770–782. doi:10.1016/j.jhep.2023.01.011

Long, Z., Sun, C., Tang, M., Wang, Y., Ma, J., Yu, J., et al. (2022). Single-cell multiomics analysis reveals regulatory programs in clear cell renal cell carcinoma. *Cell Discov.* 8 (1), 68. doi:10.1038/s41421-022-00415-0

Lu, Y., Yang, A., Quan, C., Pan, Y., Zhang, H., Li, Y., et al. (2022). A single-cell atlas of the multicellular ecosystem of primary and metastatic hepatocellular carcinoma. *Nat. Commun.* 13 (1), 4594. doi:10.1038/s41467-022-32283-3

Ma, L., Heinrich, S., Wang, L., Keggenhoff, F. L., Khatib, S., Forgues, M., et al. (2022). Multiregional single-cell dissection of tumor and immune cells reveals stable lock-and-

key features in liver cancer. *Nat. Commun.* 13 (1), 7533. doi:10.1038/s41467-022-35291-5

Ma, L., Hernandez, M. O., Zhao, Y., Mehta, M., Tran, B., Kelly, M., et al. (2019). Tumor cell biodiversity drives microenvironmental reprogramming in liver cancer. *Cancer Cell* 36 (4), 418–430. doi:10.1016/j.ccell.2019.08.007

Ma, L., Wang, L., Khatib, S. A., Chang, C. W., Heinrich, S., Dominguez, D. A., et al. (2021). Single-cell atlas of tumor cell evolution in response to therapy in hepatocellular carcinoma and intrahepatic cholangiocarcinoma. *J. Hepatol.* 75 (6), 1397–1408. doi:10.1016/j.jhep.2021.06.028

Marusyk, A., Janiszewska, M., and Polyak, K. (2020). Intratumor heterogeneity: the rosetta stone of therapy resistance. *Cancer Cell* 37 (4), 471–484. doi:10.1016/j.ccell.2020.03.007

McGinnis, C. S., Murrow, L. M., and Gartner, Z. J. (2019). DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* 8 (4), 329–337. doi:10.1016/j.cels.2019.03.003

Moncada, R., Barkley, D., Wagner, F., Chiodin, M., Devlin, J. C., Baron, M., et al. (2020). Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* 38 (3), 333–342. doi:10.1038/s41587-019-0392-8

Nam, A. S., Chaligne, R., and Landau, D. A. (2021). Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nat. Rev. Genet.* 22 (1), 3–18. doi:10.1038/s41576-020-0265-5

Neftel, C., Laffy, J., Filbin, M. G., Hara, T., Shore, M. E., Rahme, G. J., et al. (2019). An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* 178 (4), 835–849. doi:10.1016/j.cell.2019.06.024

O'Donnell, J. S., Teng, M. W. L., and Smyth, M. J. (2019). Cancer immunoediting and resistance to T cell-based immunotherapy. *Nat. Rev. Clin. Oncol.* 16 (3), 151–167. doi:10.1038/s41571-018-0142-8

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344 (6190), 1396–1401. doi:10.1126/science.1254257

Polyak, K., Haviv, I., and Campbell, I. G. (2009). Co-evolution of tumor cells and their microenvironment. *Trends Genet.* 25 (1), 30–38. doi:10.1016/j.tig.2008.10.012

Project. iotTC (2019). Available at: https://github.com/broadinstitute/inferCNV.

Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., et al. (2017). Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 171 (7), 1611–1624. doi:10.1016/j.cell.2017.10.044

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., et al. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* 14 (10), 979–982. doi:10.1038/nmeth.4402

Rambow, F., Rogiers, A., Marin-Bejar, O., Aibar, S., Femel, J., Dewaele, M., et al. (2018). Toward minimal residual disease-directed therapy in melanoma. *Cell.* 174 (4), 843–855. doi:10.1016/j.cell.2018.06.025

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinforma.* 12, 77. doi:10.1186/1471-2105-12-77

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160 (1-2), 48–61. doi:10.1016/j.cell.2014.12.033

Schupp, J. C., Adams, T. S., Cosme, C., Jr., Raredon, M. S. B., Yuan, Y., Omote, N., et al. (2021). Integrated single-cell atlas of endothelial cells of the human lung. *Circulation* 144 (4), 286–302. doi:10.1161/CIRCULATIONAHA.120.052318

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303

Sharma, A., Seow, J. J. W., Dutertre, C. A., Pai, R., Bleriot, C., Mishra, A., et al. (2020). Onco-fetal reprogramming of endothelial cells drives immunosuppressive macrophages in hepatocellular carcinoma. *Cell* 183 (2), 377–394. doi:10.1016/j.cell.2020.08.040

Sun, X. X., and Yu, Q. (2015). Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment. *Acta Pharmacol. Sin.* 36 (10), 1219–1227. doi:10.1038/aps.2015.92

Tirosh, I., Venteicher, A. S., Hebert, C., Escalante, L. E., Patel, A. P., Yizhak, K., et al. (2016). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* 539 (7628), 309–313. doi:10.1038/nature20123

Tsai, C. H., Chuang, Y. M., Li, X., Yu, Y. R., Tzeng, S. F., Teoh, S. T., et al. (2023). Immunoediting instructs tumor metabolic reprogramming to support immune evasion. *Cell Metab.* 35 (1), 118–133.e7. doi:10.1016/j.cmet.2022.12.003

Vasan, N., Baselga, J., and Hyman, D. M. (2019). A view on drug resistance in cancer. *Nature* 575 (7782), 299–309. doi:10.1038/s41586-019-1730-1

Vitale, I., Shema, E., Loi, S., and Galluzzi, L. (2021). Intratumoral heterogeneity in cancer progression and response to immunotherapy. *Nat. Med.* 27 (2), 212–224. doi:10.1038/s41591-021-01233-9

Voit, R. A., Tao, L., Yu, F., Cato, L. D., Cohen, B., Fleming, T. J., et al. (2023). A genetic disorder reveals a hematopoietic stem cell regulatory network co-opted in leukemia. *Nat. Immunol.* 24 (1), 69–83. doi:10.1038/s41590-022-01370-4

Wang, W., Guo, Y., Zhong, J., Wang, Q., Wang, X., Wei, H., et al. (2021). The clinical significance of microvascular invasion in the surgical planning and postoperative sequential treatment in hepatocellular carcinoma. *Sci. Rep.* 11 (1), 2415. doi:10.1038/s41598-021-82058-x

Wang, Y., Zhang, M., Shi, J., Zhu, Y., Wang, X., Zhang, S., et al. (2023). Cracking the pattern of tumor evolution based on single-cell copy number alterations. *Brief. Bioinform* 24 (6), bbad341. doi:10.1093/bib/bbad341

Wu, F., Fan, J., He, Y., Xiong, A., Yu, J., Li, Y., et al. (2021). Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat. Commun.* 12 (1), 2540. doi:10.1038/s41467-021-22801-0

Yu, F., Cato, L. D., Weng, C., Liggett, L. A., Jeon, S., Xu, K., et al. (2022). Variant to function mapping at single-cell resolution through network propagation. *Nat. Biotechnol.* 40 (11), 1644–1653. doi:10.1038/s41587-022-01341-y

Yu, F., Sankaran, V. G., and Yuan, G. C. (2021). CUT&RUNTools 2.0: a pipeline for single-cell and bulk-level CUT&RUN and CUT&Tag data analysis. *Bioinformatics* 38 (1), 252–254. doi:10.1093/bioinformatics/btab507

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Yuan, H., Yan, M., Zhang, G., Liu, W., Deng, C., Liao, G., et al. (2019). CancerSEA: a cancer single-cell state atlas. *Nucleic Acids Res.* 47 (D1), D900-D908–D8. doi:10.1093/nar/gky939

Zhang, A., Miao, K., Sun, H., and Deng, C. X. (2022a). Tumor heterogeneity reshapes the tumor microenvironment to influence drug resistance. *Int. J. Biol. Sci.* 18 (7), 3019–3033. doi:10.7150/ijbs.72534

Zhang, Y., Vu, T., Palmer, D. C., Kishton, R. J., Gong, L., Huang, J., et al. (2022b). A T cell resilience model associated with response to immunotherapy in multiple tumor types. *Nat. Med.* 28 (7), 1421–1431. doi:10.1038/s41591-022-01799-y

Zheng, L., Qin, S., Si, W., Wang, A., Xing, B., Gao, R., et al. (2021). Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* 374 (6574), abe6474. doi:10.1126/science.abe6474