

RESEARCH

Open Access



# Integrative single-cell analysis: dissecting CD8 + memory cell roles in LUAD and COVID-19 via eQTLs and Mendelian Randomization

Jintao Wu<sup>1†</sup>, Xiaocheng Mao<sup>2,3†</sup>, Xiaohua Liu<sup>2,3</sup>, Junying Mao<sup>4</sup>, Xianxin Yang<sup>5</sup>, Xiangwu Zhou<sup>6</sup>, Lu Tianzhu<sup>8,9</sup>, Yulong Ji<sup>10</sup>, Zhao Li<sup>7</sup> and Huijuan Xu<sup>11\*</sup> 

## Abstract

Lung adenocarcinoma exhibits high incidence and mortality rates, presenting a significant health concern. Concurrently, the COVID-19 pandemic has emerged as a grave global public health challenge. Existing literature suggests that T cells, pivotal components of cellular immunity, are integral to both antiviral and antitumor responses. Yet, the nuanced alterations and consequent functions of T cells across diverse disease states have not been comprehensively elucidated. We gathered transcriptomic data of peripheral blood mononuclear cells from lung adenocarcinoma patients, COVID-19 patients, and healthy controls. We followed a standardized analytical approach for quality assurance, batch effect adjustments, and preliminary data processing. We discerned distinct T cell subsets and conducted differential gene expression analysis. Potential key genes and pathways were inferred from GO and Pathway enrichment analyses. Additionally, we implemented Mendelian randomization to probe the potential links between pivotal genes and lung adenocarcinoma susceptibility. Our findings underscored a notable reduction in mature CD8 + central memory T cells in both lung adenocarcinoma and COVID-19 cohorts relative to the control group. Notably, the down-regulation of specific genes, such as TRGV9, could impede the immunological efficacy of CD8 + T cells. Comprehensive multi-omics assessment highlighted genetic aberrations in genes, including TRGV9, correlating with heightened lung adenocarcinoma risk. Through rigorous single-cell transcriptomic analyses, this investigation meticulously delineated variations in T cell subsets across different pathological states and extrapolated key regulatory genes via an integrated multi-omics approach, establishing a robust groundwork for future functional inquiries. This study furnishes valuable perspectives into the etiology of multifaceted diseases and augments the progression of precision medicine.

**Keywords** Single-cell transcriptomics, Immune function, Lung adenocarcinoma, COVID-19, Mendelian randomization

<sup>†</sup>Jintao Wu and Xiaocheng Mao contributed equally to this work.

\*Correspondence:

Huijuan Xu

xuhuijuan2014@163.com

Full list of author information is available at the end of the article



## Introduction

Lung Adenocarcinoma (LUAD) represents a prevalent subtype of lung cancer, and its incidence ranks among the highest globally, especially among smokers [1]. The prevalence varies across different regions, yet it constitutes a significant proportion of all lung cancer cases. The 5-year survival rate in late stages is relatively low primarily due to the typical late-stage diagnosis when the disease has already advanced; however, early detection and improved treatment methods can significantly enhance survival chances [2]. Treatment for LUAD is generally comprehensive, encompassing surgery, chemotherapy, radiation therapy, targeted therapy, and immunotherapy [3–5]. Currently, researchers are exploring biomarkers for diagnosing, prognosticating, and monitoring LUAD, with the aim of identifying superior treatment plans for patients [6, 7].

The COVID-19 pandemic, triggered by the SARS-CoV-2 infection, presents a spectrum of outcomes ranging from asymptomatic infections to life-threatening viral pneumonia and Acute Respiratory Distress Syndrome (ARDS) [8]. While host factors like age, gender, and Body Mass Index (BMI) are known to correlate with disease severity, they don't fully explain the observed inter-individual variations [9]. Despite the availability of COVID-19 vaccines, treating the disease remains imperative [10]. Nonetheless, numerous uncertainties persist regarding the genetic underpinnings of susceptibility to SARS-CoV-2 infection and the genetic determinants of COVID-19 severity. Early studies have suggested that the genes dictating an individual's blood type might impact susceptibility to SARS-CoV-2, and other immunity-related genes could also affect infection risk [11]. Polymorphisms in the ACE2 gene and the presence of TMPRSS2 (Transmembrane Serine Protease 2) are believed to increase the risk of SARS-CoV-2 infection. Moreover, research has shown that males are more susceptible to SARS-CoV-2 infection than females [12]. Several variants affecting the expression of ACE2 and TMPRSS2 receptors related to COVID-19 have been associated with susceptibility and risk factors for the disease [13]. Some genetic studies have identified potential gene variants correlated with susceptibility to SARS-CoV-2 infection and severity of COVID-19. For instance, certain gene variants might affect the binding and entry of the virus to cell surface receptors, thereby influencing infection risk and disease severity [14]. Beyond genetic factors, disease severity is also associated with age, gender, BMI, and medical history. Ethnicity too correlates with susceptibility and severity of COVID-19 [15]. By elucidating the genetic determinants of COVID-19 severity and susceptibility to SARS-CoV-2 infection, risk stratification for prioritizing immunization among high-risk individuals is feasible. Furthermore, a

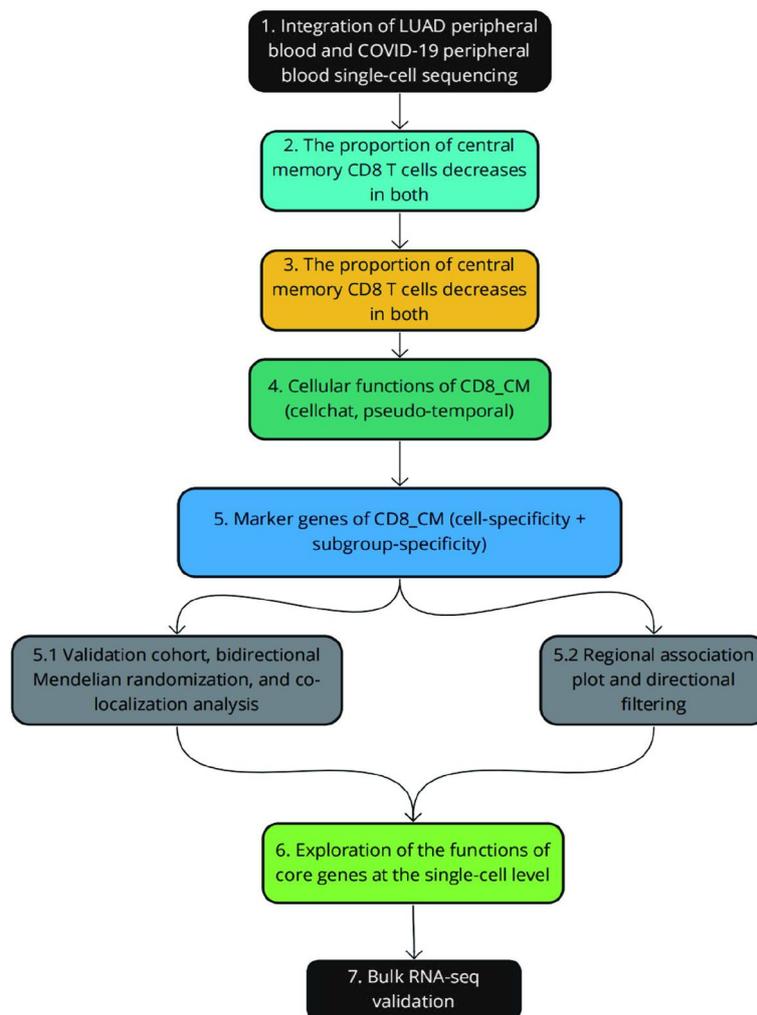
deeper understanding of these genes could guide personalized treatment approaches [14]. These studies and findings illustrate an evolving understanding of COVID-19, yet numerous unknown aspects require further exploration. With the accumulation of more research and data, a better grasp of the genetic basis of SARS-CoV-2 infection, and how genetic information can be utilized to assess and mitigate COVID-19 risk is anticipated.

Single-cell technology, a revolutionary research method, elucidates cellular heterogeneity and functional diversity by analyzing various biological attributes of individual cells. Through single-cell technology, researchers can delve into cell states, activities, and interactions among cells, enriching multifaceted studies in the life sciences domain. The continuous advancement of single-cell technology, especially in multiplex analysis, high throughput, high resolution, and accuracy, has enabled a comprehensive depiction of a cell's genetic landscape [16]. Recent progress includes single-cell epigenomics, single-cell genomics for lineage tracing, spatially resolved single-cell transcriptomics, and single-cell omics sequencing technologies based on third-generation sequencing platforms [17]. Due to rapid technological developments, encompassing improvements in throughput, accuracy, automation, and commercialization, single-cell RNA sequencing (scRNA-seq) has been extensively utilized to address pivotal biological and medical questions [17]. Single-cell multi-omics technology has impacted cell lineage tracing, the creation of tissue and cell-specific atlases, tumor immunology, and cancer genetics, alongside fundamental and translational research in cellular spatial information mapping [16]. These technologies have been deployed in research concerning tumors, microbiology, neurology, reproduction, immunology, digestive, and urinary systems, unveiling the crucial role of single-cell sequencing in both fundamental and clinical research [18]. Single-cell technology facilitates the generation of cell and tissue atlases, exploration of complex disease biology, thus offering a potent tool to solve various problems in the life sciences domain [16]. Recently, an increasing number of researchers are employing single-cell sequencing technologies to investigate cell-level specific mechanisms, cell phenotypes, and cell-specific gene expression, aiming for a better understanding of cellular biological characteristics and functions. The ongoing progress and widespread application of single-cell technology are propelling multifaceted research in life sciences, providing robust support to address key questions in biology and medicine. As the technology continues to evolve, broader applications of single-cell technology are envisioned, and its potential to significantly contribute to future scientific research is anticipated.

Mendelian Randomization (MR) epitomizes a robust statistical methodology for exploring causal associations, particularly playing a pivotal role in drug target research. This method leverages common genetic variants as an unconfounded and unbiased "natural randomized trial" to address causality issues [19]. MR has been extensively applied to various disease research endeavors. For instance, in COVID-19 research, investigators can employ MR to probe the association between specific gene variants and viral infection, as well as how these variants influence the course and severity of the disease post-infection [20]. In Parkinson's disease research, MR analysis could elucidate whether a causal link exists between specific gene variants and the risk of developing Parkinson's disease [21]. Mendelian Randomization, a potent tool for causal inference, offers opportunities to delve into the

relationship between diseases and drug treatment efficacy. By thoroughly comprehending the principles and applications of this method, a better understanding of disease pathogenesis and potential treatment strategies is achievable.

In this study, we aim to identify potential targets between COVID-19 and lung cancer, conducting MR analysis by integrating eQTL discovered in blood with two independent LUAD genome-wide association study (GWAS) datasets. The association between TRGV9 and lung cancer was examined. Our study represents a secondary analysis of publicly available data. According to the original GWAS protocols, all participants provided informed consent, and all ethical approvals for the GWAS were obtained by the original GWAS authors. The workflow of this study is illustrated in Fig. 1.



**Fig. 1** The workflow of the present study

## Methods

### Data collection

We downloaded STAR-counts data and clinical information from both 54 normal and 503 LUAD samples from the TCGA dataset repository (<https://portal.gdc.cancer.gov>). From this dataset, we extracted TPM-formatted data, subsequently undergoing normalization processing via  $\log_2(\text{TPM} + 1)$ . We retained samples possessing RNAseq data and clinical information for further analyses. The samples were utilized for subsequent analyses. The log-rank test was employed to assess the survival differences between the aforementioned two or more groups using Kaplan–Meier survival analysis. Datasets GSE171555 and GSE162498 were retrieved from the GEO database (<https://www.ncbi.nlm.nih.gov/>). We selected samples from three COVID-19 infected patients (GSM5227108, GSM5227109, GSM5227110), three healthy individuals (GSM5227117, GSM5227130, GSM5227134), and three peripheral blood samples from patients with lung adenocarcinoma (GSM4952957, GSM4952958, GSM4952959) for 10X single-cell sequencing. The FastQC tool assessed sequencing data quality, and Cell Ranger was employed for read mapping and count matrix generation. We also accessed the GSE43458 dataset as an external cohort for bulk-RNA analysis and risk model validation. To ascertain the potential causal link between differentially expressed genes and lung adenocarcinoma, we executed bidirectional Mendelian Randomization (MR) on two distinct datasets. Initially, we treated differentially expressed genes as risk factors with lung adenocarcinoma as the outcome and subsequently reversed this approach. A two-step MR method then evaluated the regulatory roles of pivotal genes in lung adenocarcinoma. The SNPs utilized as genetic instruments were derived from a comprehensive European GWAS (<https://gwas.mrcieu.ac.uk/>) [22, 23].

### Single-cell data processing and analysis

In our study, we extracted data from various samples including normal, COVID-19 infected, and tumor tissues, and generated a Seurat object for each. These objects were then amalgamated into a single comprehensive Seurat object, facilitating unified analysis. To ensure data integrity, we assessed the proportion of mitochondrial and erythrocyte genes in each cell, applying quality control criteria to exclude cells of suboptimal quality. We established a gene count range of 200 to 4000 and a mitochondrial gene proportion threshold of less than 10% for quality control. Subsequently, data normalization was performed to mitigate inter-sample variability. During dimensionality reduction and clustering, we initially

selected 2000 genes with the highest variability for principal component analysis (PCA). To further minimize batch effects, we employed the Harmony [24] algorithm and visualized the data in two dimensions using the UMAP technique, which revealed various cell subpopulations in an unsupervised manner. Cell subpopulation annotation leveraged the Cell Marker database, with representative marker genes displayed via VlnPlot and FeaturePlot. Manual cell type labels were assigned to each subpopulation and annotations were preserved. In addition to manual annotation, we explored automated annotation with the SingleR package, presenting the results through DimPlot. Following satisfactory annotation, we proceeded with differential gene expression analysis, adopting a 1.5-fold change and an adjusted  $p$ -value of less than 0.05 as significance criteria, showcased via VolcanoPlot. Differential genes were functionally annotated and their expression across cell subpopulations was depicted in a heatmap. For data encompassing temporal sequences or developmental processes, we utilized the Monocle method for cell trajectory reconstruction and the cellchat package to examine intercellular communication and regulatory dynamics.

### CD8\_CM Key marker gene eQTL and LUAD's Mendelian randomization analysis

Gene expression data preprocessing involved normalization, batch effect mitigation, and missing value handling. Key marker genes of CD8\_CM, in comparison with other T cells and cells, were identified. An eQTL analysis for the CD8\_CM marker gene followed. Gene symbols converted to ENSEMBL IDs ensured data uniformity. Low-quality SNPs underwent removal, and genotype fitting occurred. A stringent eQTL  $P$ -value threshold, such as  $5 \times 10^{-8}$ , was set. From the 'finn-b-C3\_NSCLC\_ADENO\_EXALLC' GWAS dataset, SNPs related to the key marker gene were extracted as MR analysis instrumental variables. Following the  $R^2$  and F-statistic calculations for each SNP, only high-quality SNPs remained. Outcome data corresponding to these instrumental variables facilitated Mendelian Randomization analysis using the TwoSampleMR package.

### Validation set MR analysis, bidirectional Mendelian MR and colocalization analysis

The validation set underwent initial MR analysis. By combining outcome and instrumental variable data, a harmonized dataset ('harmonised\_dat') emerged. The MR analysis utilized a 'mr\_modified' function, estimating a gene's causal impact on the target disease and computing the Phenotypic Variance Explained (PVE). Visualization tools like the 'forest' function presented the MR results. Subsequently, a Bidirectional Mendelian Randomization

analysis was conducted. Instrumental variable data linked to the reverse Mendelian disease ("ieu-a-985") were extracted and merged with the corresponding outcome data to produce a 'harmonized LUAD gene.' The *vcfR* package processed the gene expression's eQTL data. Post data organization and merging, a colocalization analysis was conducted, employing the 'coloc.abf' function to evaluate Bayes factors under various colocalization hypotheses. The results were then interpreted in light of the Bayes factors' magnitude and distribution.

#### **Regional association plot, phenoscanner analysis and directional filtering**

Initially, we constructed the regional association plot. By reading the genotype data *eQTL-a-ENSG00000211695.vcf* (in VCF format) and related association data, we extracted eQTL information relevant to the target gene. Based on this, eQTLs located within a specified region were selected and organized into a format suitable for drawing the regional association plot. The drawing process utilized the *locuscomparer* package to visualize the association information of eQTL and GWAS, providing an intuitive graphical presentation for subsequent analysis. Next, *PhenoScanner* analysis was conducted. By loading existing GWAS summary statistics, the custom *mr\_phenosscanner* function was used to query the association information of each SNP in directories such as GWAS, eQTL, and pQTL. This process employed the *PhenoScanner* tool to query different directories based on SNP associations and to organize the results. Finally, through Steiger filtering, for each trait and PMID combination, the most significant SNP was screened and summarized into a comprehensive result table. Lastly, a directional filtering analysis was conducted. The *steiger\_filtering* function was applied to the SNPs in *harmonised\_dat* to perform the Steiger test, assessing the R2 difference between SNP with exposure and SNP with outcome. This aids in verifying the position of SNP in the causal chain. The *directionality\_test* function was also used to test the directional relationship between the SNP, exposure, and outcome.

#### **Exploring the function of exposure factors at the single-cell level**

First, through single-cell RNA expression analysis, we explored the expression of the target gene at the single-cell level and used visualization tools such as *DotPlot* and *FeaturePlot* to display the expression patterns of key genes. Next, trajectory analysis was conducted. Using the *slingshot* tool, we constructed cell developmental trajectories and visualized them with *UMAP*. In addition, key genes driving the trajectory were analyzed. Functions like *find\_switch\_logistic\_fastglm* were employed to identify

switch genes that might play an essential role in the trajectory development. The *plot\_timeline\_ggplot* was used to depict gene expression patterns over pseudotime. Subsequently, cell communication analysis was executed. Using the *CellChat* tool, communication between different cell clusters was analyzed, including identifying ligand-receptor pairs, projecting onto protein interaction networks, and calculating communication probabilities. Functions like *netVisual\_circle* and *netVisual\_bubble* visually displayed the communication network structure between different cell clusters. In metabolic analysis, the *scMetabolism* tool was used to assess the metabolic activity of macrophages. We presented the activity differences of cells from different gene groups in specific metabolic pathways using the metabolism pathway's *DotPlot*. Finally, differential gene analysis was conducted. Using *Seurat's FindAllMarkers* function, differentially expressed genes in the *CD8\_CM* cell subgroup were identified. The heatmap was utilized to draw the heatmap of the expression matrix, displaying the gene expression patterns across samples. Differential gene analysis was also conducted at the bulk data level by reading external datasets.

#### **Trajectory inference**

We employed the *monocle3* R package [25, 26] to infer the differentiation trajectories of T cells. Initially, the data underwent preprocessing steps including quality control, normalization, and dimensionality reduction. Subsequently, following the guidelines provided by the official documentation, we configured the parameters before conducting cell state inference and transcriptomic trajectory analysis.

#### **Cell culture**

A549, PC-9, and H1299 cell lines obtained from the Cell Bank of the Chinese Academy of Sciences. The cells were cultured at 37 °C with a high-glucose medium containing 10% fetal bovine serum and 5% CO<sub>2</sub>.

#### **Western blotting**

Using RIPA buffer supplemented with phenylmethylsulfonyl fluoride (PMSF), protease inhibitors (PI), and phosphatase inhibitors (PPI). The total protein concentration of the supernatants was quantified using the Pierce BCA Protein Assay Kit according to the manufacturer's instructions (Thermo Fisher Scientific, USA). The proteins were then separated by SDS-PAGE and transferred onto PVDF membranes (Millipore, Billerica, MA, USA) for subsequent analysis. Membranes were blocked with 5% non-fat milk in Tris-buffered saline containing 0.1% Tween-20 (TBST) for two hours at room temperature, followed by overnight incubation at 4 °C with the primary

antibody TRGV9(TRGC1) (abcam, USA, ab192031). After washing thrice with TBST, the membranes were incubated with horseradish peroxidase-conjugated secondary antibody. Chemiluminescent detection was employed to visualize and capture the immunoreactive bands.

#### Data analysis

All data analyses were performed based on R 4.1.3, with  $p < 0.05$  considered statistically significant.

## Results

### Single-cell transcriptomic analysis of lung adenocarcinoma, COVID-19, and normal groups

In this study, we selected three COVID-19 patient samples, three healthy controls, and three sets of peripheral blood from lung adenocarcinoma patients for 10X single-cell RNA-seq analysis from datasets GSE171555 and GSE162498. Initial screening was performed to eliminate low-quality data (Supplementary Fig. 1a), retaining 56,851 cells for subsequent analysis. Data normalization was displayed (Supplementary Fig. 1b), and to mitigate batch effects among different samples, we integrated and standardized the samples using the Harmony method, followed by normalization, principal component analysis (PCA) dimensionality reduction, and clustering (Supplementary Fig. 1c). Visualization of each cluster was accomplished using the UMAP technique based on the first 15 principal components (Supplementary Fig. 1d). Specific marker genes were employed to identify and annotate different cell subpopulations within the single-cell RNA sequencing data. Using the Seurat package's VlnPlot function, we generated violin plots of marker gene expression for various cell types, including B cells, natural killer (NK) cells, T cells, monocytes, dendritic cells (DC), megakaryocytes/platelets, and erythrocytes, demonstrating their distribution across cell subpopulations (Supplementary Fig. 2a). Furthermore, the FeaturePlot function was used to generate feature distribution maps of these marker genes, visualizing their spatial distribution within cells (Supplementary Fig. 2b). We confirmed the number of cell subpopulations defined by unsupervised clustering algorithms and annotated them using the RenameIdents function, aligning each cluster with its corresponding cell type. Cell distributions were visualized using the UMAP algorithm and the DimPlot function (Fig. 2a) and segmented according to tissue type (Fig. 2b). Subsequently, we renamed 26 clusters using the Single R package and visualized cell distributions with UMAP reduction and DimPlot (Fig. 2c), segmenting by tissue type (Fig. 2d). To ensure the reliability of annotations, manual annotation was also performed, visualizing

cell distributions with UMAP and DimPlot (Fig. 2e) and segmenting by tissue type (Fig. 2f) to enhance the understanding of cell type distributions across different tissues.

### GO, KEGG, and WikiPathway enrichment analysis

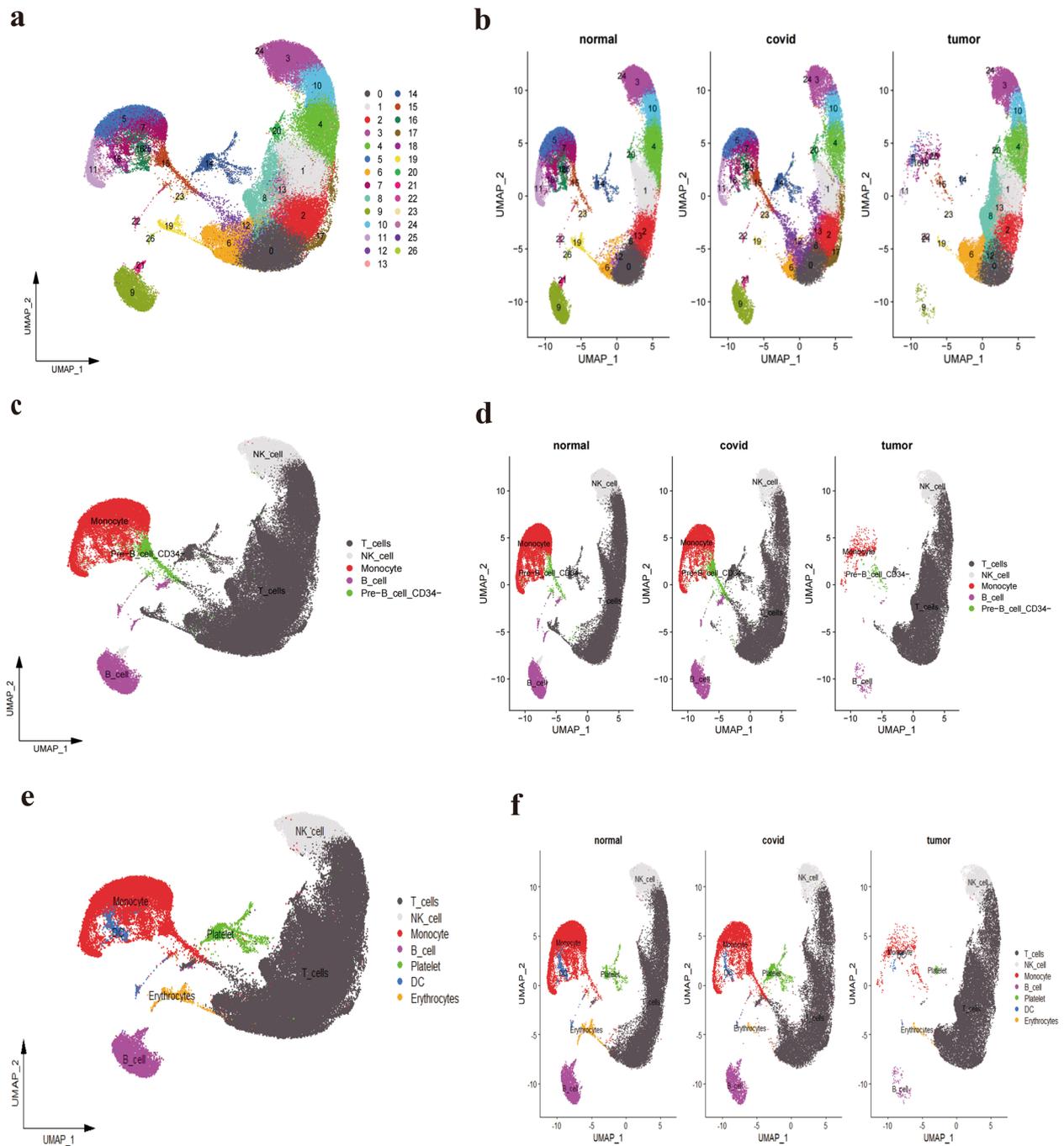
Based on the 7 cell types identified, genes with fold changes  $\geq 1.5$  or  $\leq -1.5$  were considered differentially expressed. We visualized the top 5 overexpressed and underexpressed genes in T\_cells, NK\_cell, Monocyte, B\_cell, platelet, DC, and Erythrocytes (Fig. 3a). To decipher the function of differentially expressed genes in various cell clusters, we referenced databases like TF (Transcription Factors), CSPA (Cell Surface Protein Atlas), GO\_BP (Gene Ontology Biological Processes), KEGG (Kyoto Encyclopedia of Genes and Genomes), and WikiPathway. Our results offer a preliminary understanding of the function of differentially expressed genes in each cell type (Fig. 3b).

### Single-cell transcriptomic analysis of T cells

T cells play multifaceted roles in both cancer and COVID-19, encompassing direct cytotoxicity against infected or aberrant cells, assisting the activity of other immune cells, and maintaining immunological balance. A thorough understanding of T cell functions and interactions is crucial for the development of effective therapeutic and vaccine strategies in both contexts. Initially, T cell subpopulations were extracted from comprehensive single-cell RNA sequencing data, followed by preprocessing, dimension reduction, clustering, and visualization analyses. The distribution of 17 T cell clusters was visualized using UMAP and DimPlot (Fig. 4a), segmented by tissue type (Fig. 4b). Cell proportion charts (Fig. 4c) and Dot Plots (Fig. 4d) illustrated the distribution and expression of characteristic genes within different T cell subpopulations, providing a visual analytical foundation for a deeper understanding of their functions and characteristics. We visualized the distribution of four T cell subpopulations with UMAP and DimPlot (Fig. 4e), also segmented by tissue type (Fig. 4f). Figure 4g shows the cell proportions across different samples, where we noted a higher proportion of CD8\_CM in the normal group compared to the COVID and tumor groups, and a higher proportion of CD4\_Naïve cells in the COVID and tumor groups relative to the normal group.

### Enrichment analysis of four T\_cell subgroups based on GO, KEGG and WikiPathway

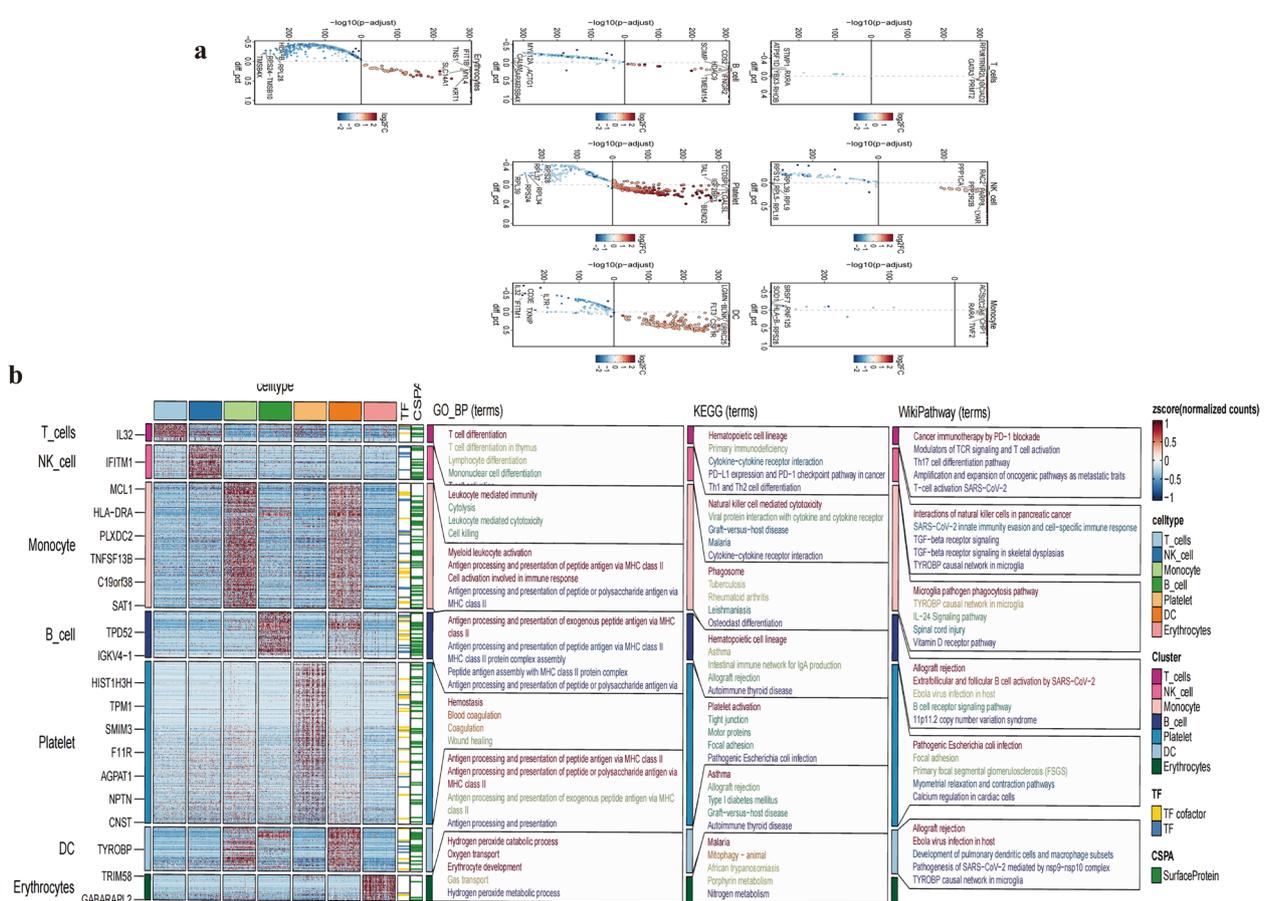
Based on the identified four cell types, differential gene expression was determined using the findallmarker function, with genes exhibiting a fold change of  $\geq 1.5$  or  $\leq -1.5$  considered significant. We visualized the top five upregulated and downregulated genes in the CD4\_Naïve,



**Fig. 2** Single-cell transcriptomics landscape of various samples. **a, b** Depict the single-cell transcriptomic profiles of different samples, showcasing the diversity and distribution of cell clusters. **c, d** Using the Single R package, we annotated the 26 clusters. The results of this automated classification are depicted, showing various cell populations. **e, f** Post manual annotation, seven distinct cell populations were identified: T\_cells, NK\_cell, Monocyte, B\_cell, platelet, DC, and Erythrocytes

CD4\_EM, CD8\_CM, and CD8\_EM subgroups (Fig. 5a). GO\_BP analysis indicated the CD4\_EM group's differential genes primarily enriched in pathways related to cellular response to antibiotics and UV-B radiation. For

the CD8\_CM group, these genes were chiefly associated with leukocyte-mediated immunity, cell killing, and lymphocyte-mediated cytotoxicity pathways. KEGG analysis demonstrated that the CD8\_CM subgroup's genes

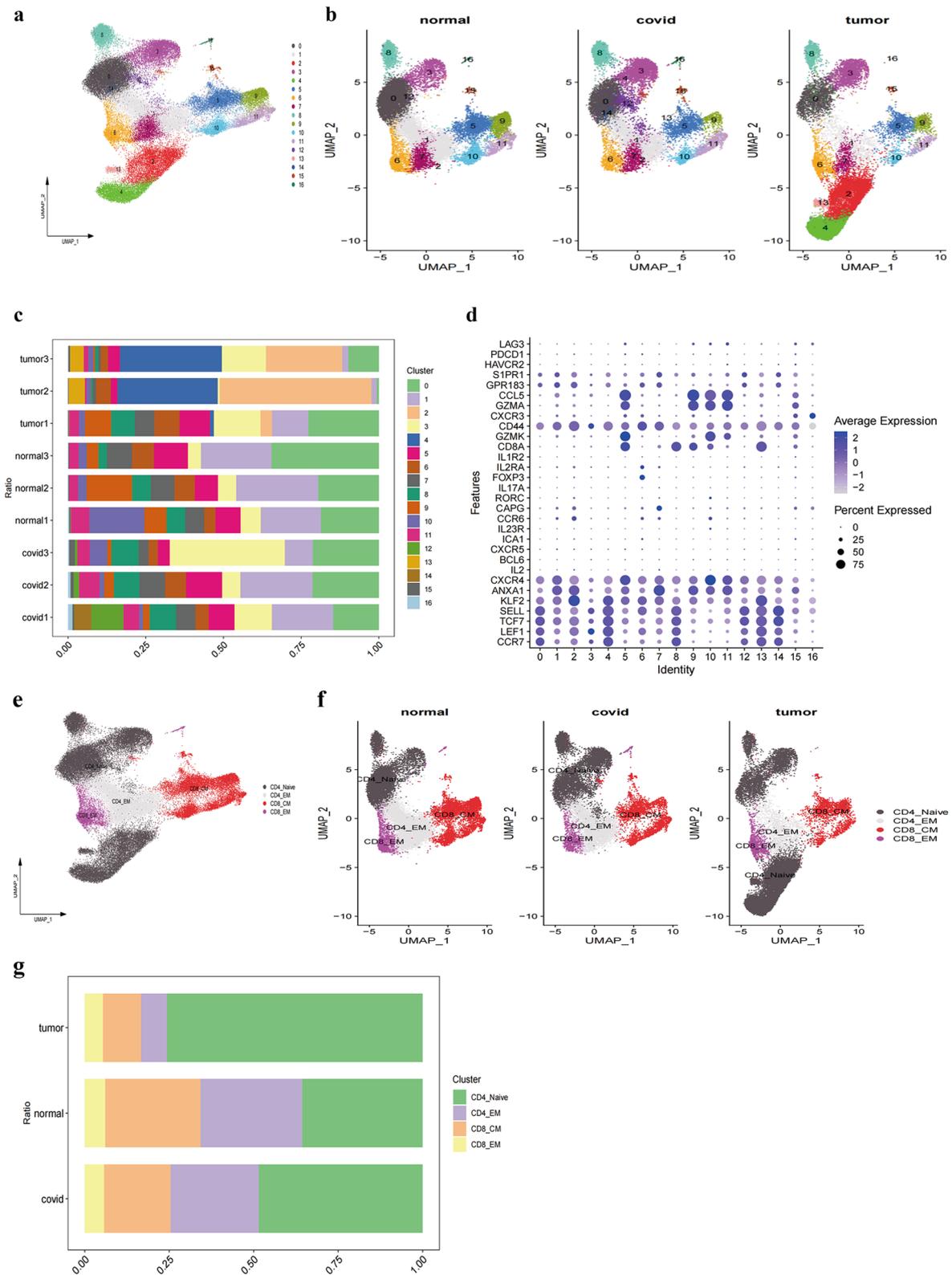


were enriched in natural killer cell-mediated cytotoxicity, whereas the CD8\_EM subgroup's genes were linked to inflammatory bowel disease, antigen processing and presentation, and Th17 cell differentiation pathways. WikiPathway analysis showed enrichment in interactions of natural killer cells in pancreatic cancer for the CD8\_CM group, and allograft rejection for the CD8\_EM group (Fig. 5b). Collectively, these findings offer insights into the distinct functional roles and biological features of various T cell subgroups, enhancing our understanding of their involvement in immune responses and disease processes.

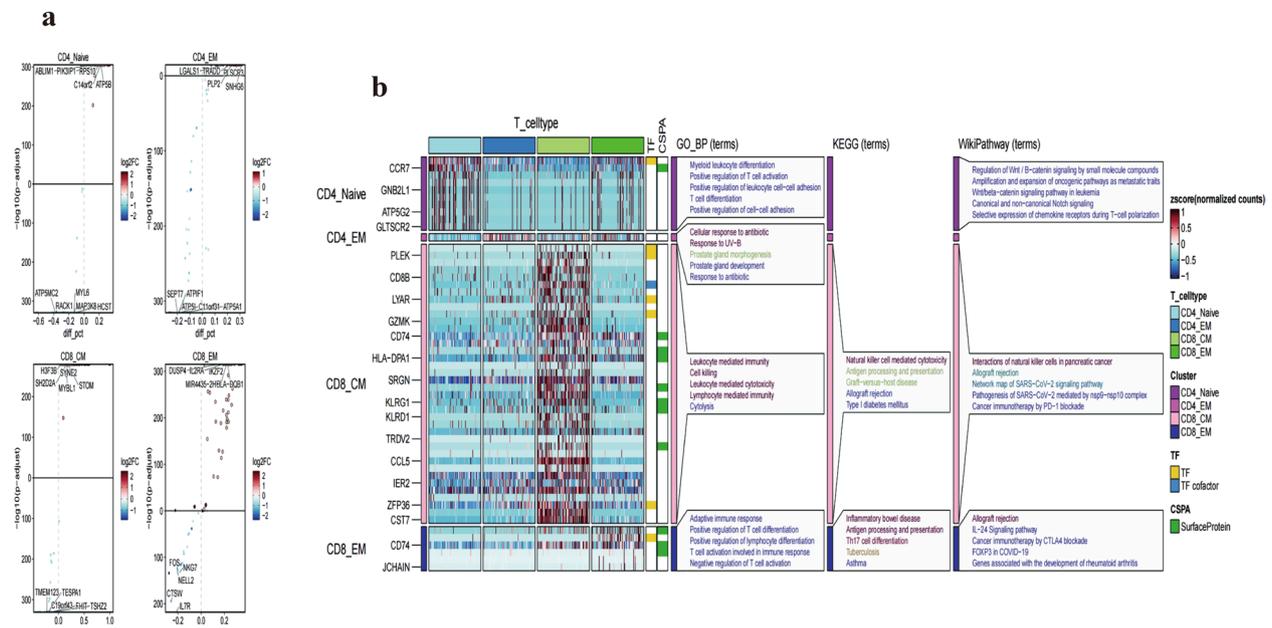
### Trajectory analysis and cellular communication of primary T cell types

T cell development originates in the bone marrow, where lymphocyte precursor cells undergo a series of differentiation steps, inclusive of T cell precursors. This differentiation is progressive, characterized by continuous rather than discrete changes. Cells at different developmental stages express distinct gene sets.

With scRNA-seq, we can capture cells at various stages, enabling exploration of continuous differentiation trajectories. The UMAP dimensionality reduction algorithm outperforms other algorithms in capturing the global and topological structures of datasets, including the spatial positioning of individual cells. We employed the slingshot R package to assess cellular differentiation processes. The trajectory analysis identified CD4\_Naïve as the starting point, culminating in CD8\_CM, suggesting CD8\_CM as terminally differentiated CD8\_T cells (Fig. 6a). To discern the role of CD8\_CM in both COVID-19 and LUAD and its intercellular relationships, we conducted cellular communication analysis. By analyzing samples from COVID-19 and LUAD, we constructed a communication network of CD8\_CM with other cells (Fig. 6b, d). Comparative analysis of pathways enriched in CD8\_CM and other cells revealed frequent communication through MIF -(CD74 + CD44) and MIF -(CD74 + CXCR4) signals with B cells, NK cells, and CD8\_EM cells in both diseases (Fig. 6c, e).



**Fig. 4** In-depth Analysis and Visualization of T-cell Subsets. **a, b** Featureplot showcasing the distribution of T-cells across 17 clusters. **c** Displays the proportion of each cell subset in different samples. **d** Dot Plot illustrating the expression of characteristic genes across various cell subsets. **e, f** Manually annotated T-cell subgroups including CD4\_Naive, CD4\_EM, CD8\_CM, and CD8\_EM. **g** Depicts the distribution of T-cell subgroups in different samples



**Fig. 5** GO, KEGG and WikiPathway enrichment analysis based on 4 subgroups of T-cells (a). Display of the top five highly and lowly expressed genes in the 4 cell groups in the volcano plot (b). GO, KEGG and WikiPathway enrichment analysis

Thus, CD8\_CM is posited as a pivotal player in both COVID-19 and LUAD.

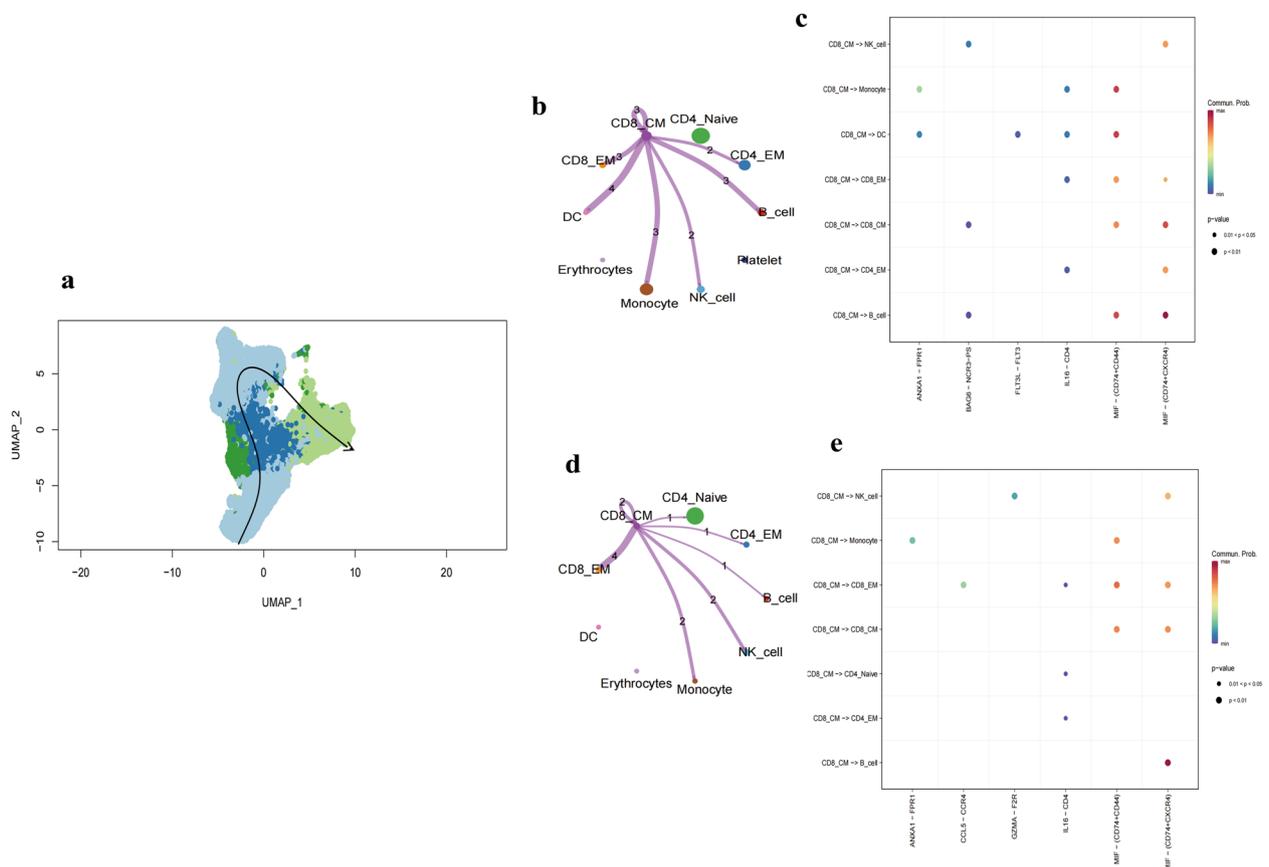
**MR analysis using key marker genes identifies three novel causal genes for LUAD**

We conducted Mendelian randomization (MR) analysis on single-cell RNA sequencing data for lung cancer to uncover key genes potentially influencing lung adenocarcinoma risk. Initially, using the Seurat tool, we identified 70 key genes (Supplementary file 1) distinguishing CD8\_CM T cells from other cell types and subtypes and determined significant markers for subsequent analysis. For deeper insight, we successfully converted gene symbols to corresponding ENSEMBL IDs using org.Hs.eg.db. This was followed by a two-sample MR analysis, involving extraction of SNP data related to our genes of interest as exposure data, retrieving lung adenocarcinoma outcome data from the EBI database, and harmonizing between exposure and outcome datasets. The MR analysis highlighted several genes significantly associated with lung cancer risk. Specifically, the RNF125 gene exhibited an odds ratio (OR) of 0.5826 (95% CI: 0.3474–0.9771,  $p=0.0406$ ). For the CD8B gene, the OR was 3.2331 (95% CI: 1.0898–9.5920,  $p=0.0344$ ). The TRGV9 gene displayed an OR of 0.3927 (95% CI: 0.2283–0.6757,  $p=0.0007$ ). These associations hint at potential genetic markers influencing the aforementioned cancer risk. To visually represent the  $-\log_{10}$ -transformed  $p$ -values against  $\ln(\text{OR})$  for each gene, we generated a volcano plot, with

genes showing significant  $p$ -values highlighted (Fig. 7a). The plot distinctly reveals genes with significant positive and negative associations. Subsequently, we developed a forest plot, visualizing the OR and 95% CI for each significant gene, emphasizing the robustness and direction of each gene’s association (Fig. 7b, c, d and e). In the two-sample Mendelian randomization (MR) analysis, it was observed that RNF125, CD8B, and TRGV9 exhibit no heterogeneity concerning lung adenocarcinoma (LUAD). The results are detailed in (Supplement Table 2). Employing the Inverse Variance Weighted method, the analysis revealed that RNF125, CD8B, and TRGV9 do not exhibit horizontal pleiotropy in the MR analysis of LUAD (RNF125:  $P=0.755703216$ ; CD8B:  $P=0.711380024$ ; TRGV9:  $P=0.571487335$ ) (Supplement Table 3, Supplement Table 1). These tests evaluated the consistency of genetic associations across different SNPs and identified potential horizontal pleiotropy. The analysis validated our identified associations, indicating minimal biases and confounding. The results of this study offer fresh perspectives and directions for subsequent biological validation and mechanistic research.

**Mendelian randomization and colocalization analysis of genes RNF125, CD8B and TRGV9 with lung cancer risk**

To verify the reliability of our findings, we conducted a validation set analysis using Mendelian randomization. We explored the association between three genes, namely RNF125, CD8B, and TRGV9, and lung cancer risk.



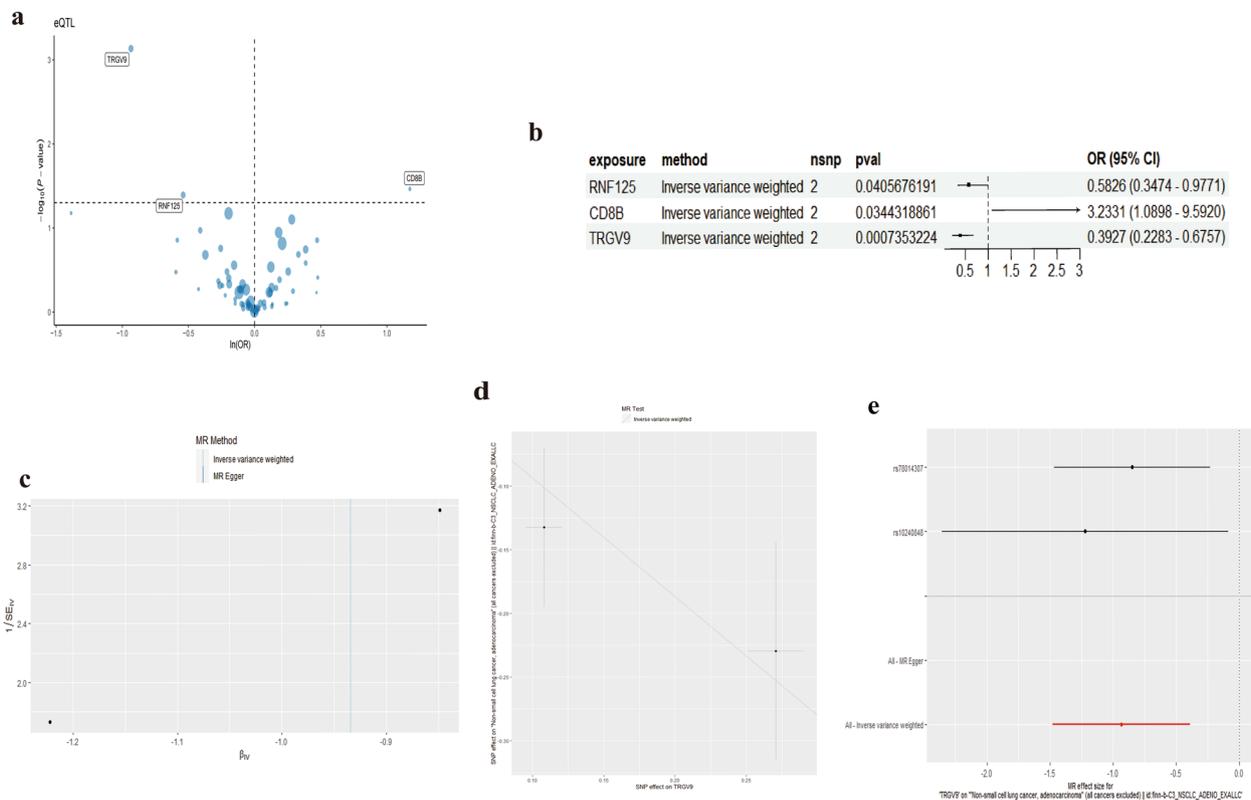
**Fig. 6** Cellular Differentiation and Communication Analysis in COVID-19 and LUAD. **a** Cellular differentiation trajectory plot using UMAP, showcasing the progression from CD4\_Naive to CD8\_CM, implying that CD8\_CM represents terminally differentiated CD8\_T cells. **b, d** Cellular communication networks constructed for CD8\_CM and other cells in both COVID-19 and LUAD samples. The networks highlight potential interactions and communication patterns between CD8\_CM and other cell types. **c, e** Differential enrichment signaling pathway analysis between CD8\_CM and other cells

The results showed a significant association of RNF125 mutations with lung cancer risk (OR=0.5858, 95% CI: 0.3504–0.9793). Mutations in CD8B were associated with an increased lung cancer risk (OR=3.2477, 95% CI: 1.0978–9.6079), while mutations in TRGV9 were linked to a reduced risk (OR=0.3977, 95% CI: 0.2313–0.6837) (Fig. 8a). Additionally, the MR analysis for lung cancer and the TRGV9 gene showed a non-significant association (OR=1.0686, 95% CI: 0.8316–1.3731). We also conducted a reverse MR analysis, which, when considering lung cancer as the exposure and the TRGV9 gene as the outcome, did not show a significant association (Fig. 8b, c, d and e).

**Investigating the association between the TRGV9 gene and lung adenocarcinoma using Mendelian Randomization**

In this study, we employed the Mendelian randomization approach to probe the association between the TRGV9 gene and lung adenocarcinoma. Initially, we displayed the regional association plots for the TRGV9 gene eQTL

in tandem with the lung adenocarcinoma GWAS results. By contrasting their association strengths, we identified several single nucleotide polymorphisms (SNPs) demonstrating pronounced associations across both phenotypes. Notably, specific SNPs, such as rs23923593, associated with the TRGV9 gene eQTL, also exhibited significant correlation in the lung adenocarcinoma GWAS. This provided preliminary evidence hinting at a potential link between this gene and lung adenocarcinoma development (Fig. 9). We employed the Phenoscanner tool to further delve into these SNPs’ associations with other traits. The results revealed associations of some SNPs with diverse traits, suggesting their involvement in various biological processes. To ascertain the causative direction in our MR analysis, we performed the Steiger test, which discerns which trait lies closer to the gene and hence is more likely the true exposure. Our findings predominantly aligned with our preliminary hypothesis, suggesting that alterations in TRGV9 expression could modulate lung adenocarcinoma risk. Collectively, our analyses offer preliminary evidence



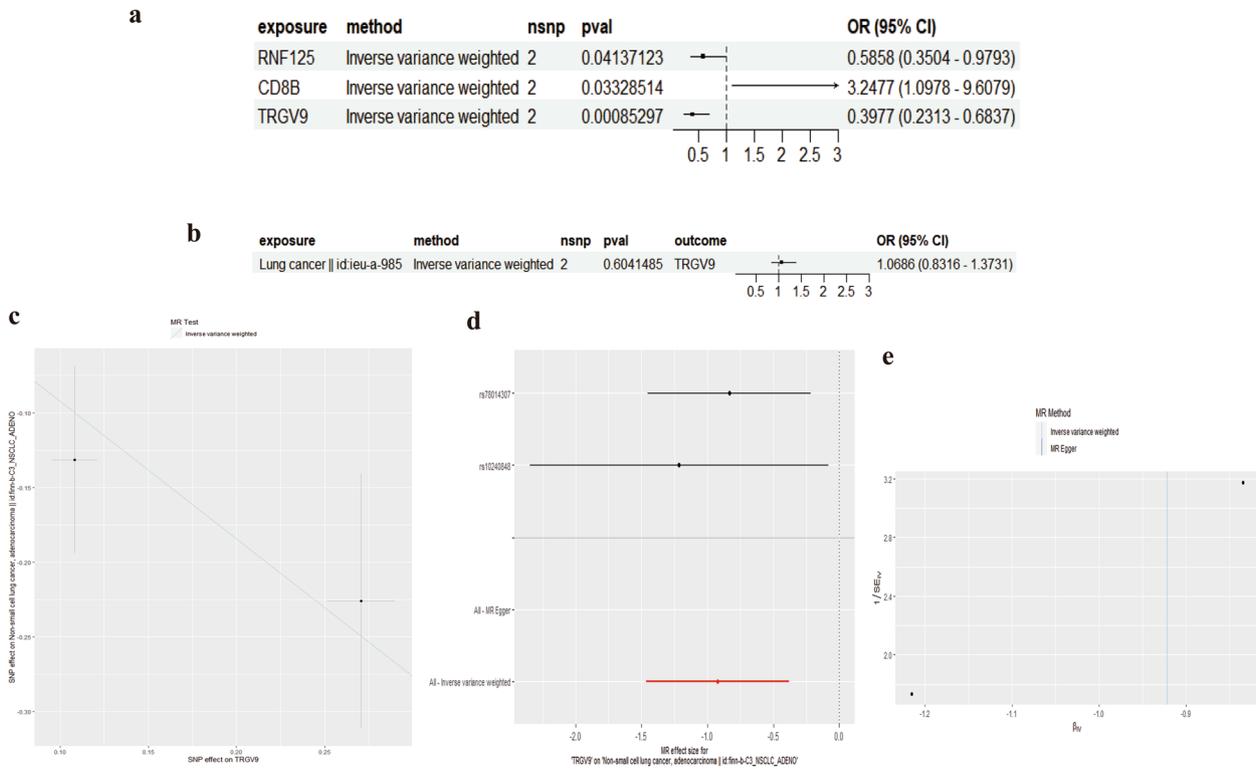
**Fig. 7** **a** Volcano plot illustrating the association between key genes and the risk of lung adenocarcinoma. **b** Forest plot depicting the association of key genes with the risk of lung adenocarcinoma. **c** Mendelian Randomization Heterogeneity Test This assesses the consistency of genetic associations across different SNPs. **d** Pleiotropy Test Evaluates the horizontal pleiotropy across various key genes. **e** Validity and Robustness of Mendelian Randomization. **f** Results Further validates the identified associations, indicating minimal bias and confounding

supporting a potential causal relationship between the TRGV9 gene and lung adenocarcinoma.

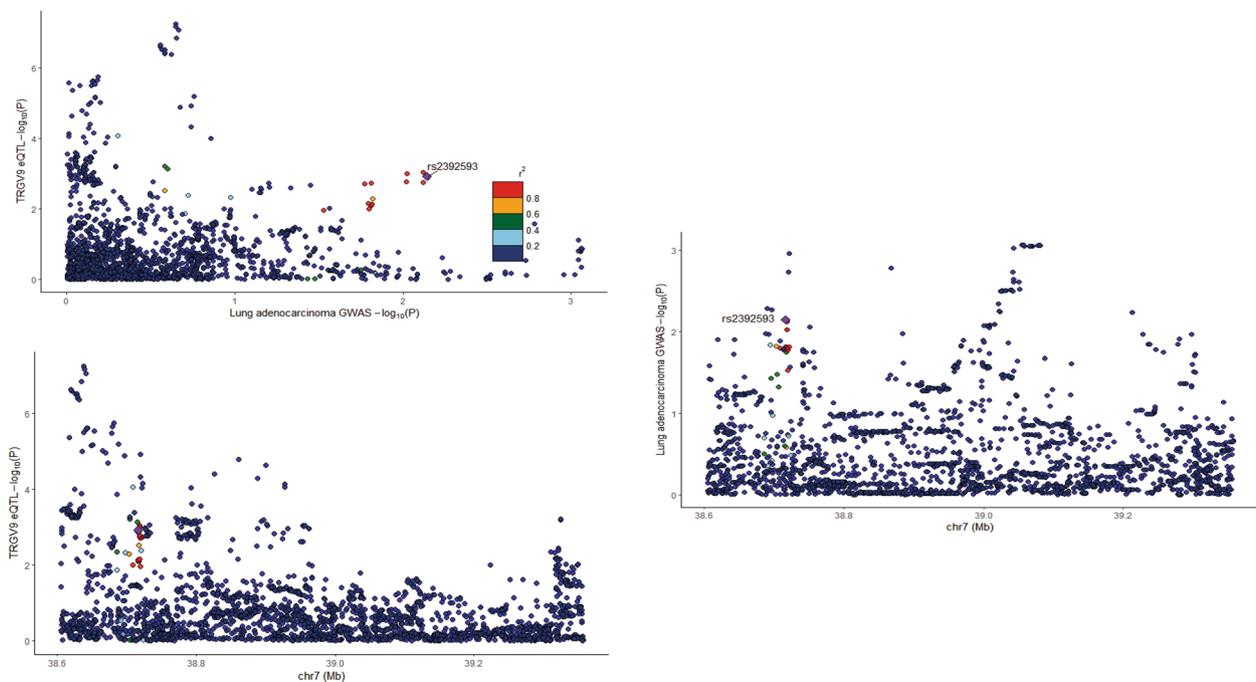
### Single-cell transcriptomic analysis reveals TRGV9’s crucial role in T-cell metabolism and function in LUAD

Utilizing the UMAP algorithm, we performed dimensionality reduction and visualization on single-cell transcriptomic data to identify distinct cellular subpopulations. We first visualized the expression patterns of three key marker genes across different cell clusters (Fig. 10a). Subsequently, we employed the feature plot function to specifically highlight the expression of TRGV9 and CD8B within individual cells (Fig. 10b and c). Moreover, we segmented the data based on tissue type and conducted a focused analysis of TRGV9 expression, revealing a lower level of expression in tumor tissues compared to normal tissues (Fig. 10d), thus providing clues for further functional studies. Employing the Slingshot tool, we performed cell trajectory analysis, observing the transition of T cells from a naive to a mature state along the trajectory (Fig. 10e), which suggests potential dynamic changes of T cells during disease progression. The distribution of gene expression is shown in Fig. 10f, where

most genes displayed low activity levels, while a minority exhibited relatively high expression. To explore changes in gene expression during cell development or transcriptional dynamics, we depicted the on/off states of various genes over pseudotime (Fig. 10g). Scatter plot analysis was used to depict the relationship between the expression of the 'TRGV9' gene and pseudotime. Statistical analysis indicated a moderate positive correlation (Pearson  $r=0.43$ ,  $p<0.001$ ), suggesting that 'TRGV9' expression might increase over pseudotime (Fig. 10h). Intercellular communication analysis, conducted with the CellChat tool, showed differences between the TRGV9+CD8\_CM and TRGV9-CD8\_CM cell subgroups (Fig. 10i), with the TRGV9+CD8\_CM subgroup demonstrating more ligand-receptor interactions and stronger cell signaling activities. Furthermore, by comparing enriched signaling pathways between TRGV9+CD8\_CM cells and other cells, we noted that the MIF—(CD74+CD44) and MIF—(CD74+CXCR4) signaling pathways were involved in cell communication in LUAD, paralleling our previous findings (Fig. 10j). Metabolic pathway analysis revealed distinct states of T cell activation; using the scMetabolism tool, we assessed metabolic activities within the



**Fig. 8** Validation Queue, Bidirectional Mendel. **a** Mendelian randomization analysis showing associations of RNF125, CD8B, and TRGV9 gene variants with lung cancer risk. **b** Reverse Mendelian randomization depicting the relationship between lung cancer and the TRGV9 gene. **c** Scatter plot examining heterogeneity in the Mendelian randomization analysis. **d** Forest plot illustrating the effect sizes for the associations. **e** Funnel plot assessing potential pleiotropy in the Mendelian randomization results



**Fig. 9** Regional association map

cells. Notably, key metabolic pathways differed significantly between the TRGV9+CD8\_CM and TRGV9-CD8\_CM subgroups. Specifically, the TRGV9+CD8\_CM subgroup exhibited enhanced uric acid synthesis capabilities, potentially linked to rapid T cell division and energy requirements. In the tricarboxylic acid (TCA) cycle, TRGV9+CD8\_CM cells showed increased activity, suggesting higher oxidative phosphorylation and ATP production. Compared to the TRGV9-CD8\_CM subgroup, TRGV9+CD8\_CM cells displayed reduced fatty acid oxidation activity, likely reflecting a preference for glucose over fatty acids as an energy source. Additionally, the TRGV9+CD8\_CM subgroup showed increased nucleotide synthesis activity, consistent with rapid proliferation and DNA synthesis demands. These differences could be related to T cell activation, proliferation, and immune functions (Fig. 10k), providing important insights for further research. Differential gene expression analysis between TRGV9+CD8\_CM and TRGV9-CD8\_CM cell subgroups revealed several genes differentially expressed in association with LUAD. Notably, the TRGV9 gene exhibited significant differences between LUAD and healthy groups. Similar gene expression patterns were observed in the overall transcriptomic data (Fig. 10l), further validating the accuracy and reliability of our single-cell analysis. Additionally, partial biological validation was performed, with Western blotting analysis indicating higher expression of TRGV9 in a normal cell line compared to three other lung adenocarcinoma cell lines (Fig. 10m), corroborating the results from our dataset analysis.

### Trajectory inference of T cell subsets

We conducted cell trajectory analysis to infer the differentiation status of T-cell subtypes, understanding the evolutionary process of cells in LUAD, while simultaneously demonstrating the trajectories of different differential genes. (Fig. S4a) illustrates pseudo-temporal trajectories of four T-cell subtypes. The specific trajectory directions are depicted in (Fig. S4b), where the analysis reveals that CD4\_Naïve serves as the pseudo-temporal starting point, while CD8\_CM represents the

endpoint of cell pseudo-temporality. Additionally, we present the expression status of the top 10 differential genes in different cell subtypes (Fig. S4c), as well as the pseudo-temporal expression changes of the top 10 genes across different cells (Fig. S4d). Finally, we identified 240 co-expression gene modules (Fig. S4e).

### Correlation analysis of TRGV9 expression differences with clinical features

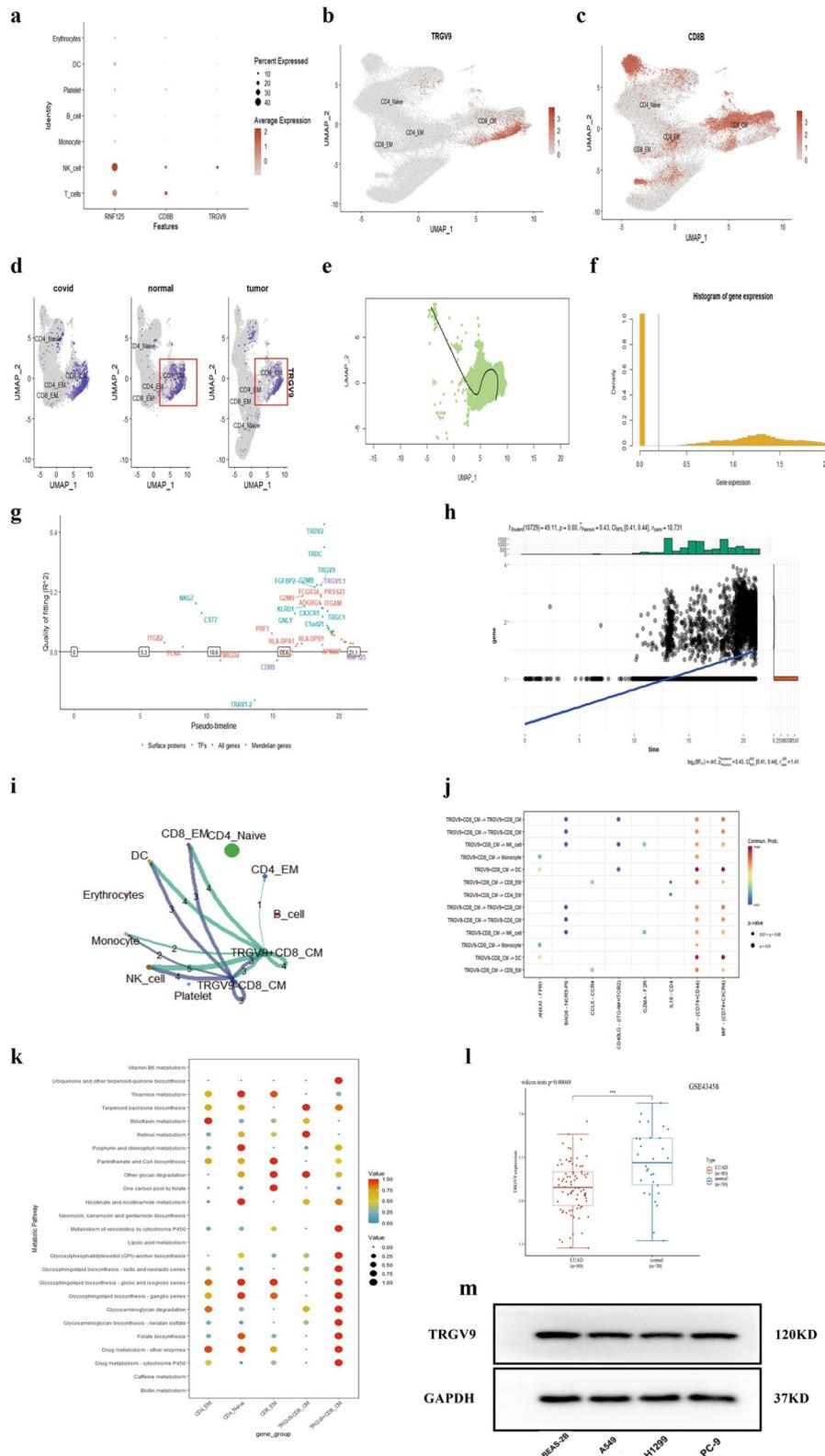
Initially, utilizing the GeneCards website [27] (<https://www.genecards.org/>), we identified TRGV9, also known as TRGC1. Through an analysis of TCGA-LUAD and its associated clinical information, we observed higher expression of TRGV9 in normal tissues compared to tumor tissues (Fig. S5a). Additionally, TRGV9 exhibited correlation with prognosis ( $p=0.001$ ) (Fig. S5b). Subsequently, we investigated the correlation of TRGV9 with clinical features. We assessed the correlation between TRGV9 expression and clinical features using a heatmap. The results revealed significant differences in the distribution of N staging, tumor grading, and T staging between high and low expression groups (Fig. S5c). Specifically, significant differences were observed between N0 and N1 ( $p=0.024$ ) and N0 and N1 (0.00069) (Fig. S5d). Furthermore, significant differences existed between stage I and stage III ( $p=0.001$ ) (Fig. S5e). Lastly, we identified significant differences between T1 and T2 ( $p=0.0021$ ) as well as between T1 and T3 ( $p=0.038$ ) within the T staging (Fig. S5f). In summary, the expression of TRGV9 appears to be closely associated with the clinical prognosis of patients with lung adenocarcinoma.

### Discussion

The advent of single-cell RNA sequencing (scRNA-seq) technology has furnished a unique opportunity to elucidate transcriptional heterogeneity among diverse cell types under various disease conditions, which is instrumental for a deeper understanding of cellular functions and interactions. In this study, scRNA-seq analysis was employed to explore the disparities among cell sub-populations

(See figure on next page.)

**Fig. 10** Single-cell transcriptomic analysis reveals the pivotal role of TRGV9 in T-cell metabolism and function in LUAD: **a** Expression pattern of key genes across different cell subpopulations. **b** UMAP dimensionality reduction visualizing cell subgroups with differential expressions of specific key genes. **c** Expression distribution of the CD8B gene across cells plotted using UMAP. **d** UMAP plots showing the distribution of cells in conditions of COVID, normal, and tumor, highlighting the expression of TRGV9 and CD8B. **e** Cell trajectory analysis from primary to mature T-cell states using Slingshot. **f** Histogram representing the distribution of gene expression across cells. **g** Gene on/off status in pseudo-time, reflecting gene expression dynamics during cellular developmental or transcriptional processes. **h** Scatter plot depicting the relationship between the expression of the 'TRGV9' gene and pseudo-time. **i** Intercellular communication analysis between TRGV9+CD8\_CM and TRGV9-CD8\_CM cell subgroups using the CellChat tool. **j** Differential signaling pathways enriched in communications involving TRGV9+CD8\_CM cells compared to others. **k** Metabolic pathway analysis showing the activation states of T cells, revealing differences in metabolic activities between TRGV9+CD8\_CM and TRGV9-CD8\_CM cell subgroups. **l** Differential gene expression analysis between TRGV9+CD8\_CM and TRGV9-CD8\_CM cell subpopulations, with a focus on the significant difference in TRGV9 expression between LUAD and healthy groups. **m** Expression of TRGV9 in normal cells and three adenocarcinoma cell types



**Fig. 10** (See legend on previous page.)

between COVID-19, lung adenocarcinoma (LUAD), and healthy controls, with a particular focus on T cells.

CD8<sub>CM</sub> cells, a sub-group within the CD8+T cell lineage, are memory T cells formed post-antigen activation [28]. Upon re-encounter with the corresponding antigen, CD8<sub>CM</sub> cells proliferate rapidly and differentiate into effector T cells, playing a vital role in secondary immune responses. They are predominantly found in peripheral blood and lymphatic tissues [29], exhibiting both memory and effector functionalities, thereby constituting a crucial component of cellular immune memory. They play a pivotal role in antiviral and anti-tumor immunity [30, 31]. CD8+T cells, especially the CD8<sub>CM</sub> subset, are primary tumor-infiltrating immune cells responsible for delivering antitumor responses. Immunotherapeutic strategies aiming to restore the effector function of CD8+T cells have provided support for the current successful cancer immunotherapy [32, 33]. Long-lived memory CD8+T cells play an important role in tumor immunity, inclusive of central, effector, stem-like, and tissue-resident memory CD8+T cell subsets [34]. Lung cancer induces functional defects in CD8+T cells, which correlates with the clinical response to immunotherapy. A deeper understanding of the impact of lung cancer on CD8+T cells might aid in the development of new therapeutic approaches [35]. Research indicates that SARS-CoV-2 mRNA vaccines can offer protection against severe disease as early as ten days post-vaccination, a time when neutralizing antibodies are barely detectable, suggesting that vaccine-induced CD8+T cells might be the main mediators of protection during this early stage [36]. CD8+T cells are essential for protective immune responses, directly participating in viral clearance, thereby playing a significant role in viral infections, including SARS-CoV-2 [30]. It was discovered that CD8+T cell responses are robustly activated one week post bnt162b2 mRNA vaccination, at a time when circulating CD4+T cells and neutralizing antibodies are only weakly detectable [36]. CD8+T cells might be the principal protective mediators induced by mRNA vaccines, expanding in the early protection window post prime vaccination, preceding the maturation of other effector arms of vaccine-induced immunity, and are stably maintained post boost vaccination [36]. Particularly, CD8<sub>CM</sub> cells play a significant role in immune surveillance and antitumor immune responses in lung adenocarcinoma [37]. In patients with lung adenocarcinoma, defective CD8<sub>CM</sub> cells are present; although these cells exhibit a memory T cell phenotype, their functionality resembles effector T cells, manifesting weakened effector function and reduced proliferative capacity [38]. Additionally, the number of CD8<sub>CM</sub> cells in lung adenocarcinoma tissues is lower compared

to peripheral blood, and their cytokine secretion ability is diminished. This can be partly attributed to the tumor microenvironment inhibiting CD8<sub>CM</sub> cell activity through mechanisms like the PD-1/PD-L1 pathway [39]. Enhancing the infiltration and activity of CD8<sub>CM</sub> cells in lung adenocarcinoma tissues through immunotherapeutic means is a crucial approach to augment antitumor immune responses in patients with lung adenocarcinoma [40]. Some studies have shown that PD-1/PD-L1 inhibitors can partially restore the functionality of CD8<sub>CM</sub> cells in tumors [41–44]. The subset and functional state of CD8<sub>CM</sub> cells can serve as significant biomarkers for predicting and monitoring the therapeutic efficacy of immunotherapy for lung adenocarcinoma [45]. Optimizing the functionality of CD8<sub>CM</sub> cells will contribute to further advancements in immunotherapy for lung adenocarcinoma. Moreover, CD8<sub>CM</sub> cells are key effector cells against SARS-CoV-2 infection, rapidly proliferating and differentiating into cytotoxic T lymphocytes in COVID-19 patients, directly recognizing and killing virus-infected cells, thus inhibiting viral replication [46, 47]. The increase in CD8<sub>CM</sub> cell numbers in COVID-19 patients correlates with the severity of the disease, indicating their involvement in the body's antiviral immune responses. Monitoring the dynamic changes of CD8<sub>CM</sub> cells can assess the immune status of the body, providing a basis for evaluating the effects of subsequent immunotherapy and vaccination. In summary, the subset and functional state of CD8<sub>CM</sub> cells are crucial for immune responses, whether in tumor immunity or in combating SARS-CoV-2 and other viral infections. In-depth exploration of the biological characteristics and functions of CD8<sub>CM</sub> cells, along with optimizing their functionality through immunotherapeutic strategies, will offer significant insights for advancing both antitumor and antiviral immunotherapy.

This study systematically investigated the composition and functionality of immune cells in the peripheral blood of COVID-19 and LUAD patients, as well as healthy individuals, through scRNA-seq analysis. Our data initially validated the presence of T cells, B cells, NK cells, etc., in peripheral blood, consistent with existing literature [48]. Notably, compared to healthy controls, both COVID-19 and LUAD patient groups exhibited a varying degree of decrease in the proportion of mature effector CD8+T cells (CD8<sub>CM</sub>). This might be associated with weakened immune function under disease conditions. For instance, multiple research groups have validated the reduction in CD8+T cell numbers and cytotoxicity in the peripheral blood of COVID-19 patients [49, 50]. Moreover, in mouse tumor models, the tumor microenvironment has been shown to inhibit the proliferation of CD8+T cell [51]. Therefore, the variation in CD8<sub>CM</sub> cell proportions may

reflect their significant role in disease immune responses. Through functional enrichment analysis, we discovered characteristic pathways of different cell sub-groups correlating with known cellular functions, providing clues to understand the physiological functions of each cell type. Additionally, through single-cell trajectory analysis, we unveiled the dynamic differentiation process of T cells from early to mature stages, offering a new perspective for studying T cell development and activation. Furthermore, cell-cell communication network analysis revealed enhanced connections between CD8\_CM T cells and B cells, NK cells, etc., suggesting their potential involvement in regulating other immune cells during disease progression. This study not only identified disease-related immune cell alterations but also proposed potential molecular mechanisms through multi-omics analysis. For example, we found that the reduction in TRGV9 expression might be one of the causes leading to CD8+ T cell functional decline. TRGV9 is a T cell receptor variable region gene involved in T cell antigen recognition [52]. In the tumor microenvironment, suppression of TRGV9 expression results in T cells failing to recognize tumor antigens, thereby reducing their effector functionality. In fact, in mouse tumor models, overexpression of TRGV9 enhances T cell cytotoxicity against tumor [53]. Hence, modulating TRGV9 expression levels might emerge as a new strategy to boost T cell antitumor effects. Additionally, RNF125, a ubiquitin ligase, also participates in various signaling pathway processes, and the specific mechanisms related to diseases require further investigation [54]. The expression level of TRGV9 in lung adenocarcinoma patients is closely associated with clinical features, where higher expression correlates with a better prognosis. The clinical relevance suggests significant differences in the pathological characteristics of lung adenocarcinoma, particularly in tumor staging and grading. These findings imply that TRGV9 could serve as a potential prognostic marker, offering valuable leads for further in-depth research and clinical applications in the future.

It's imperative to acknowledge certain limitations in this study. Firstly, due to the limited sample size, the results necessitate validation through an expanded sample scale. Moreover, subsequent studies are required to experimentally validate the functional roles of the predicted key genes in disease onset. Animal model studies should also be conducted to explore the molecular mechanisms and their regulatory modes. Besides, scRNA-seq technology has its inherent batch effects, which require caution in data interpretation. Technical noise during sample processing and library construction in scRNA-seq could lead to batch effects, hence careful interpretation of results is essential. Future research needs to increase the sample size and adopt more standardized procedures to validate the findings. Secondly, Mendelian

randomization analysis relies on the correlation between genetic markers and phenotypes to infer causal relationships. However, potential genetic pleiotropy might lead to false-positive results. Thus, the newly discovered candidate genes still require functional validation.

## Conclusion

This study preliminarily identified the association of immune cell composition and key gene expression with COVID-19 and LUAD, proposing some verifiable hypotheses. This lays the groundwork for exploring the molecular mechanisms of diseases, and also provides potential targets for the development of new preventive and therapeutic strategies. Subsequent research should employ a variety of experimental methods for validation and exercise caution in interpreting the biological significance of various omics data. This study offers valuable experience in bridging the gap between single-cell omics and genetics, and serves as a paradigm for precision medicine research in complex diseases.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41065-023-00307-7>.

**Additional file 1: Fig. s1.** Analysis of rRNA-sequencing data: a. Distribution of RNA features across different samples: nFeature\_RN: Number of unique RNA features in the samples. nCount\_RNA: Total count of RNA molecules in the samples, percent\_mt: Percentage of mitochondrial genes in the samples, percent\_HB: Percentage of HB genes in the samples. b. Cell distribution visualized through tSNE and UMAP algorithms within 9 samples before harmony. c. Cell distribution visualized through tSNE and UMAP algorithms within 9 samples after harmony. d. The relationship between "harmony" and "Standard Deviation".

**Additional file 2: Fig. s2.** For validation and to ensure the reliability of the annotations, manual annotations were also performed. The expression patterns of specific marker genes for each cell type are represented through violin plots and feature plots.

**Additional file 3: Fig. s3.** a. T\_cell distribution visualized through tSNE and UMAP algorithms within 9 samples before harmony. b. The relationship between "harmony" and "Standard Deviation" before harmony. c. T\_cell distribution visualized through tSNE and UMAP algorithms within 9 samples after harmony. d. The relationship between "harmony" and "Standard Deviation" after harmony.

**Additional file 4.**

**Additional file 5.**

**Additional file 6.**

**Additional file 7: Fig. s4.** Simulated time-series analysis a. Time trajectories of CD4\_Naive, CD4\_EM, CD8\_CM, and CD8\_EM b. Developmental direction of cells over time c. Feature plot of 10 genes across 4 T cell subgroups d. Simulated time-series trajectories of differential genes. **Fig. s5.** Correlation analysis between differential TRGV9 expression and clinical features a. Analysis of TRGV9 expression differences between normal tissues and LUAD b. Survival analysis of TRGV9 in LUAD c. Heatmap illustrating the association of TRGV9 expression with age, gender, stage, T, and N of LUAD patients d. Correlation of TRGV9 expression with N staging e. Correlation of TRGV9 expression with stage f. Correlation of TRGV9 expression with T staging. **Supplement Table 1.** Impact of RNF125, CD8B, and TRGV9 on lung adenocarcinoma.

### Acknowledgements

Thanks to all staff involved in the design and maintenance of TCGA and GEO database systems. We thank these workers for providing researchers with a publicly shared sequencing database for further analysis.

### Authors' contributions

Huijuan Xu and Jintao Wu contributed to the concept and design of this article. Yulong Ji, Zhao Li, Xiaohua Liu, Xianxin Yang and Junying Mao participated in data collection, data analysis and contributed equally to this work. Jintao Wu and Xiaocheng Mao participated in writing the manuscript and interpreting the data. Tianzhu Lu participated in revising the data and manuscript. All authors contributed to the article and approved the submitted version.

### Funding

This work was funded by the Traditional Chinese Medicine Science Research Fund of Jiangxi Province (grant no. 2022B928).

### Availability of data and materials

This study analyzed publicly available data sets. All original analytical datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Nanchang University Jiangxi Medical College, Nanchang, Jiangxi, China. <sup>2</sup>Departments of Blood Transfusion, Institute of Transfusion, Jiangxi Key Laboratory of Transfusion, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China. <sup>3</sup>Key Laboratory of Jiangxi Province for Transfusion Medicine, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China. <sup>4</sup>The First People's Hospital of Wenling, Affiliated Wenling Hospital, Wenzhou Medical University, Taizhou, Zhejiang, China. <sup>5</sup>The Fifth Affiliated Hospital of Jinan University, Heyuan, Guangdong, China. <sup>6</sup>The Fifth Affiliated Hospital of Shantou University, Shantou, Guangdong, China. <sup>7</sup>Department of Thoracic Surgery, Jiangxi Cancer Hospital, Nanchang, Jiangxi Province, China. <sup>8</sup>Department of Radiation Oncology, Jiangxi Cancer Hospital of Nanchang University, Nanchang, Jiangxi, People's Republic of China 330006. <sup>9</sup>NHC Key Laboratory of Personalized Diagnosis and Treatment of Nasopharyngeal Carcinoma, Jiangxi Cancer Hospital of Nanchang University, Nanchang, Jiangxi 330006, People's Republic of China. <sup>10</sup>Jiangxi Key Laboratory of Translational Cancer Research, Jiangxi Cancer Hospital, Jiangxi Province, China. <sup>11</sup>Department of Clinical Laboratory, Hangzhou TCM Hospital Affiliated to Zhejiang Chinese Medical University, Hangzhou, China.

Received: 6 November 2023 Accepted: 14 December 2023

Published online: 31 January 2024

### References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71:209–49.
- Sorin M, Rezanejad M, Karimi E, et al. Single-cell spatial landscapes of the lung tumour immune microenvironment. *Nature*. 2023;614:548–54.
- Nacer DF, Liljedahl H, Karlsson A, Lindgren D, Staaf J. Pan-cancer application of a lung-adenocarcinoma-derived gene-expression-based prognostic predictor. *Brief Bioinform*. 2021;22:bbab154.
- Hirano S, Kojima A, Nakayama Y, Takeda T, Kishimoto T, Takahashi T, Kuwbara S, Mori M. A case report of neuromyelitis optica spectrum disorder induced by pembrolizumab treatment for lung adenocarcinoma: a clinical and immunohistochemical study. *BMC Neurol*. 2022;22:483.
- Isla D, de Castro J, García-Campelo R, Lianes P, Felipe E, Garrido P, Paz-Ares L, Trigo JM. Treatment options beyond immunotherapy in patients with wild-type lung adenocarcinoma: a Delphi consensus. *Clin Transl Oncol*. 2020;22:759–71.
- Liu Z, Liu L, Weng S, et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat Commun*. 2022;13:816.
- Solari JIG, Filippi-Chiela E, Pilar ES, Nunes V, Gonzalez EA, Figueiró F, Andrade CF, Klamt F. Damage-associated molecular patterns (DAMPs) related to immunogenic cell death are differentially triggered by clinically relevant chemotherapeutics in lung adenocarcinoma cells. *BMC Cancer*. 2020;20:474.
- Docherty AB, Harrison EM, Green CA, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO clinical characterisation protocol: prospective observational cohort study. *BMJ*. 2020;369:m1985.
- Gaziano L, Giambartolomei C, Pereira AC, et al. Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities related to COVID-19. *Nat Med*. 2021;27:668–76.
- Fang E, Liu X, Li M, Zhang Z, Song L, Zhu B, Wu X, Liu J, Zhao D, Li Y. Advances in COVID-19 mRNA vaccine development. *Signal Transduct Target Ther*. 2022;7:94.
- COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature*. 2021;600:472–7.
- SeyedAlinaghi S, Mehrtak M, MohsseniPour M, et al. Genetic susceptibility of COVID-19: a systematic review of current evidence. *Eur J Med Res*. 2021;26:46.
- Fricke-Galindo I, Falfán-Valencia R. Genetics insight for COVID-19 susceptibility and severity: a review. *Front Immunol*. 2021;12:622176.
- Anastassopoulou C, Gkizarioti Z, Patrinos GP, Tsakris A. Human genetic factors associated with susceptibility to SARS-CoV-2 infection and COVID-19 disease severity. *Hum Genomics*. 2020;14:40.
- Velavan TP, Pallerla SR, Rüter J, Augustin Y, Kremsner PG, Krishna S, Meyer CG. Host genetic factors determining COVID-19 susceptibility and severity. *EBioMedicine*. 2021;72:103629.
- Baysoy A, Bai Z, Satija R, Fan R. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol*. 2023;24:1–19.
- Wen L, Tang F. Recent advances in single-cell sequencing technologies. *Precis Clin Med*. 2022;5:pbac002.
- Tang X, Huang Y, Lei J, Luo H, Zhu X. The single-cell sequencing: new developments and medical applications. *Cell Biosci*. 2019;9:53.
- Deng Y, Tu D, O'Callaghan CJ, Liu G, Xu W. Two-stage multivariate Mendelian randomization on multiple outcomes with mixed distributions. *Stat Methods Med Res*. 2023;32:1543–58.
- Liu D, Yang J, Feng B, Lu W, Zhao C, Li L. Mendelian randomization analysis identified genes pleiotropically associated with the risk and prognosis of COVID-19. *J Infect*. 2021;82:126–32.
- Billingsley KJ, Barbosa IA, Bandrés-Ciga S, et al. Mitochondria function associated genes contribute to Parkinson's Disease risk and later age at onset. *NPJ Parkinsons Dis*. 2019;5:8.
- Mountjoy E, Schmidt EM, Carmona M, et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat Genet*. 2021;53:1527–33.
- Long E, Patel H, Byun J, Amos CI, Choi J. Functional studies of lung cancer GWAS beyond association. *Hum Mol Genet*. 2022;31:R22–36.
- Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods*. 2019;16:1289–96. <https://doi.org/10.1038/s41592-019-0619-0>.
- Kolat D, Kalužińska Ž, Orzechowska M, et al. Functional genomics of AP-2α and AP-2γ in cancers: in silico study. *BMC Med Genomics*. 2020;13:174. <https://doi.org/10.1186/s12920-020-00823-9>.
- Tran B, Tran D, Nguyen H, et al. scCAN: single-cell clustering using autoencoder and network fusion. *Sci Rep*. 2022;12:10267. <https://doi.org/10.1038/s41598-022-14218-6>.
- Safran M, Dalah I, Alexander J, et al. GeneCards version 3: the human gene integrator. *Database (Oxford)*. 2010;2010:baq020. <https://doi.org/10.1093/database/baq020>.
- Ex vivo culture of chimeric antigen receptor T cells generates functional CD8+ T cells with effector and central memory-like phenotype - PubMed. <https://pubmed.ncbi.nlm.nih.gov/20428216/>. Accessed 23 Oct 2023

29. Larsen SE, Voss K, Laing ED, Snow AL. Differential cytokine withdrawal-induced death sensitivity of effector T cells derived from distinct human CD8<sup>+</sup> memory subsets. *Cell Death Discov.* 2017;3:17031.
30. Demberg T, Boyer JD, Malkevich N, et al. Sequential priming with simian immunodeficiency virus (SIV) DNA vaccines, with or without encoded cytokines, and a replicating adenovirus-SIV recombinant followed by protein boosting does not control a pathogenic SIVmac251 mucosal challenge. *J Virol.* 2008;82:10911–21.
31. Wallace J, Narasipura SD, Sha BE, French AL, Al-Harhi L. Canonical Wnts mediate CD8<sup>+</sup> T cell non-cytolytic anti-HIV-1 activity and correlate with HIV-1 clinical status. *J Immunol.* 2020;205:2046–55.
32. Kumar S, Singh SK, Rana B, Rana A. Tumor-infiltrating CD8<sup>+</sup> T cell anti-tumor efficacy and exhaustion: molecular insights. *Drug Discov Today.* 2021;26:951–67.
33. Raskov H, Orhan A, Christensen JP, Gögenur I. Cytotoxic CD8<sup>+</sup> T cells in cancer and cancer immunotherapy. *Br J Cancer.* 2021;124:359–67.
34. Han J, Khatwani N, Searles TG, Turk MJ, Angeles CV. Memory CD8<sup>+</sup> T cell responses to cancer. *Semin Immunol.* 2020;49:101435.
35. Prado-García H, Romero-García S, Aguilar-Cazares D, Meneses-Flores M, Lopez-Gonzalez JS. Tumor-induced CD8<sup>+</sup> T-cell dysfunction in lung cancer patients. *Clin Dev Immunol.* 2012;2012:741741.
36. Oberhardt V, Luxemburger H, Kemming J, et al. Rapid and stable mobilization of CD8<sup>+</sup> T cells by SARS-CoV-2 mRNA vaccine. *Nature.* 2021;597:268–73.
37. Yang L, He Y-T, Dong S, et al. Single-cell transcriptome analysis revealed a suppressive tumor immune microenvironment in EGFR mutant lung adenocarcinoma. *J Immunother Cancer.* 2022;10:e003534.
38. Kim H, Park S, Han K-Y, et al. Clonal expansion of resident memory T cells in peripheral blood of patients with non-small cell lung cancer during immune checkpoint inhibitor treatment. *J Immunother Cancer.* 2023;11:e005509.
39. Karaki S, Blanc C, Tran T, et al. CXCR6 deficiency impairs cancer vaccine efficacy and CD8<sup>+</sup> resident memory T-cell recruitment in head and neck and lung tumors. *J Immunother Cancer.* 2021;9:e001948.
40. Zhang C, Yin K, Liu S-Y, Yan L-X, Su J, Wu Y-L, Zhang X-C, Zhong W-Z, Yang X-N. Multiomics analysis reveals a distinct response mechanism in multiple primary lung adenocarcinoma after neoadjuvant immunotherapy. *J Immunother Cancer.* 2021;9:e002312.
41. Morisada M, Clavijo PE, Moore E, Sun L, Chamberlin M, Van Waes C, Hodge JW, Mitchell JB, Friedman J, Allen CT. PD-1 blockade reverses adaptive immune resistance induced by high-dose hypofractionated but not low-dose daily fractionated radiation. *Oncoimmunology.* 2017;7:e1395996.
42. Lee J, Ahn E, Kissick HT, Ahmed R. Reinvigorating exhausted T cells by blockade of the PD-1 pathway. *For Immunopathol Dis Therap.* 2015;6:7–17.
43. Sakuishi K, Apetoh L, Sullivan JM, Blazar BR, Kuchroo VK, Anderson AC. Targeting Tim-3 and PD-1 pathways to reverse T cell exhaustion and restore anti-tumor immunity. *J Exp Med.* 2010;207:2187–94.
44. Leen AM, Sukumaran S, Watanabe N, et al. Reversal of tumor immune inhibition using a chimeric cytokine receptor. *Mol Ther.* 2014;22:1211–20.
45. Yamauchi T, Hoki T, Oba T, et al. T-cell CX3CR1 expression as a dynamic blood-based biomarker of response to immune checkpoint inhibitors. *Nat Commun.* 2021;12:1402.
46. Kared H, Redd AD, Bloch EM, et al. SARS-CoV-2-specific CD8<sup>+</sup> T cell responses in convalescent COVID-19 individuals. *J Clin Invest.* 2021;131(e145476):145476.
47. Rha M-S, Kim AR, Shin E-C. SARS-CoV-2-Specific T cell responses in patients with COVID-19 and unexposed individuals. *Immune Netw.* 2021;21:e2.
48. Boldt A, Borte S, Fricke S, Kentouche K, Emmrich F, Borte M, Kahlenberg F, Sack U. Eight-color immunophenotyping of T-, B-, and NK-cell subpopulations for characterization of chronic immunodeficiencies. *Cytometry B Clin Cytom.* 2014;86:191–206.
49. Zheng M, Gao Y, Wang G, Song G, Liu S, Sun D, Xu Y, Tian Z. Functional exhaustion of antiviral lymphocytes in COVID-19 patients. *Cell Mol Immunol.* 2020;17:533–5.
50. Mahmoodpoor A, Hosseini M, Soltani-Zangbar S, et al. Reduction and exhausted features of T lymphocytes under serological changes, and prognostic factors in COVID-19 progression. *Mol Immunol.* 2021;138:121–7.
51. Yan W, Liu X, Ma H, Zhang H, Song X, Gao L, Liang X, Ma C. Tim-3 fosters HCC development by enhancing TGF- $\beta$ -mediated alternative activation of macrophages. *Gut.* 2015;64:1593–604.
52. Fichtner AS, Karunakaran MM, Gu S, Boughter CT, Borowska MT, Starick L, Nöhren A, Göbel TW, Adams EJ, Herrmann T. Alpaca (*Vicugna pacos*), the first nonprimate species with a phosphoantigen-reactive V $\gamma$ 9V $\delta$ 2 T cell subset. *Proc Natl Acad Sci U S A.* 2020;117:6697–707.
53. Oberg HH, Kellner C, Gonnermann D, Sebens S, Bauerschlag D, Gramatzki M, Kabelitz D, Peipp M, Wesch D. Tribody [(HER2)2xCD16] is more effective than Trastuzumab in enhancing  $\gamma\delta$  T cell and natural killer cell cytotoxicity against HER2-expressing cancer cells. *Front Immunol.* 2018;9:814.
54. Hu J, Ding R, Liu S, Wang J, Li J, Shang Y. Hypermethylation of RNF125 promotes autophagy-induced oxidative stress in asthma by increasing HMGB1 stability. *iScience.* 2023;26:107503.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.