基于CiteFuse的肺腺癌单细胞多组学数据整合细胞类型识别*

徐昭阳[1,2] 刘改琴[2] 王 彤[1,2] 田雅昕[1,2] 曾 平[3]
房瑞玲[2,4] 罗艳虹[2,4] 余红梅[2,4] 张岩波[2,4] 曹红艳[2,4△]

【提 要】 **目的** 探讨CiteFuse方法在肺腺癌单细胞多组学数据细胞类型识别中的应用，发现潜在生物标志物与治疗靶点。**方法** 采用CiteFuse整合肺腺癌转录组和表位的细胞索引测序技术（cellular indexing of transcriptomes and epitopes by sequencing，CITE-seq）数据中的转录组和蛋白组信息，识别肺腺癌细胞类型，进一步对细胞类型结果进行细胞注释、RNA和抗体衍生标签（antibody-derived tags，ADT）的差异表达分析、GO富集分析、ADT-RNA网络构建、细胞通讯分析和ADT重要性评估。**结果** CiteFuse将肺腺癌细胞分为8类，鉴定出与各细胞类型显著相关的RNA和ADT分别为75个和25个，GO富集分析得到356个富集项，构建出36个ADT-RNA对。此外，配体-受体分析进一步揭示了细胞间通讯机制，得到316个重要的配体-受体对，77条信号通路，并分析VEGF信号通路网络中细胞类型间的相互作用。通过对ADT的评估得到46个重要蛋白。**结论** CiteFuse可以准确地识别细胞类型，筛选出与肺腺癌进展相关的潜在生物标志物及关键信号通路，为进一步探索肺腺癌的分子机制及针对性治疗策略提供重要依据。

【关键词】 单细胞多组学 CiteFuse 相似网络融合 细胞类型识别 肺腺癌
【中图分类号】 Q78；R734.2 【文献标志码】 A **DOI** 10.11783/j.issn.1002-3674.2026.03.006

Cell Type Identification in Lung Adenocarcinoma Single-cell Multi-omics Data Integration based on CiteFuse

Xu Zhaoyang，Liu Gaiqin，Wang Tong，et al（*Academy of Medical Sciences*，*Shanxi Medical University*，*Taiyuan* 030001）

【**Abstract**】 **Objective** To evaluate the application of CiteFuse for cell type identification in single-cell multi-omics data from lung adenocarcinoma and to explore potential biomarkers and therapeutic targets. **Methods** This study integrates transcriptomic and proteomic data from cellular indexing of transcriptomes and epitopes by sequencing（CITE-seq）of lung adenocarcinoma using CiteFuse to identify cell types. Further analyses include cell type annotation，differential expression analysis of RNA and antibody-derived tags（ADT），GO enrichment analysis，construction of the ADT-RNA network，cell communication analysis，and evaluation of ADT importance. **Results** CiteFuse identified eight distinct cell types in lung adenocarcinoma samples. A total of 75 RNA markers and 25 ADT markers were significantly associated with specific cell populations，and 356 enriched GO terms were identified. In addition，36 ADT-RNA pairs were detected. Cell-cell communication analysis revealed 316 significant ligand-receptor pairs and 77 signaling pathways，highlighting complex intercellular interactions，particularly within the VEGF signaling network. Furthermore，46 important proteins were identified through ADT feature evaluation. **Conclusion** CiteFuse enables more accurate identification of cell types and the selection of potential biomarkers and key signaling pathways，providing crucial insights for further exploration of the molecular mechanisms of lung adenocarcinoma and targeted therapeutic strategies.

【**Key words**】 Single-cell multi-omics；CiteFuse；Similarity network fusion；Cell type identification；Lung adenocarcinoma

肺癌是全球第一大常见癌症，也是全球癌症死亡的首位原因[1]。肺癌的五年生存率仅为22.9%[2]。基于临床和组织病理学，肺癌可分为非小细胞肺癌和小细胞肺癌，其中非小细胞肺癌占比超过80%，肺腺癌为非小细胞肺癌中最常见类型，大多数肺腺癌患者在诊断时已为晚期，预后不良[3-4]。随着靶点发现和靶向药物的研发，部分患者通过精准匹配靶点接受靶向治疗，预后和生存期有所改善，然而更多患者由于难以匹配靶点无法接受相应治疗。肺癌细胞异质性是影响肺癌靶向药物治疗效果的重要原因[5]。传统批量测序难以解析肺癌不同细胞群体的异质性特征，而单细胞测序能捕捉到肺癌组织细胞水平的特异性变化，开展肺癌细胞类型识别研究，进而确定特异性标志物和治疗靶点，为制定肺癌个体化治疗策略提供了新的方向和契机[6]。

在单细胞多组学测序技术中，转录组和表位的细胞索引测序技术（cellular indexing of transcriptomes and epitopes by sequencing，CITE-seq）[7]以及RNA表达和蛋白质测序技术（RNA expression and protein

*基金项目：国家自然科学基金（82473739；82273742）；山西省基础研究计划（202303021211130）；山西省回国留学人员科研资助项目（2024-081）；山西省高等教育“百亿工程”科技引导专项；煤炭环境致病与防治教育部重点实验室（山西医科大学）开放课题基金（MEKLCEPP/SXMU-202415）
1.山西医科大学医学科学院（030001）
2.山西医科大学公共卫生学院卫生统计教研室，重大疾病风险评估山西省重点实验室
3.徐州医科大学公共卫生学院流行病与卫生统计学系
4.煤炭环境致病与防治教育部重点实验室
△通信作者：曹红艳，E-mail：caohy@sxmu.edu.cn

sequencing, REAP-seq)[8], 通过结合scRNA-seq和抗体衍生标签(antibody-derived tags, ADT)检测, 可以在单个细胞中同时定量分析肺癌全局基因表达和高度多重化的细胞表面蛋白。整合单细胞多组学数据识别细胞类型, 和单一组学相比, 能够捕获不同组学间关联和潜在因果关系, 形成信息互补[9], 提高细胞类型识别的准确性。

现有单细胞多组学数据整合方法可以分为三类: 基于矩阵分解的方法、基于神经网络的方法和基于相似网络的方法[10]。基于矩阵分解的方法, 对单细胞的多个组学数据同时进行分解并提取多组学之间的共同信息, 如MOFA+[11]、scAI[12]等。基于矩阵分解的方法简单易行, 可解释性强, 但是难以捕获特征间的非线性关系。基于神经网络的方法, 通过对输入数据采用编码器进行数据压缩, 将其映射到低维空间, 使用解码器实现对输入数据的重构, 能有效地捕获特征间的非线性关系, 如scMVAE[13]、DeepMAPS[14]等, 但计算复杂。基于相似网络的方法, 利用核函数将数据的原始特征映射到高维特征空间, 构建细胞相似网络, 能够很好地处理非线性问题, 且关注邻近细胞的相似性和相互作用, 能够捕捉细胞之间的局部关系, 如CiteFuse[15]、Joint diffusion[16]等, 其中CiteFuse将批量测序的经典多组学数据整合方法相似网络融合(similarity network fusion, SNF)[17]拓展到单细胞CITE-seq的mRNA和蛋白数据整合, 采用信息传递的方法将多个相似网络迭代融合为一个相似网络, 采用谱聚类或Louvain聚类等方法进行细胞类型识别, 融合后的细胞相似网络包含了不同网络的共享和互补信息, 在细胞类型识别研究中独具优势。

因此, 本研究针对肺腺癌CITE-seq数据, 采用CiteFuse方法整合转录组和蛋白组数据, 识别肺腺癌不同细胞类型, 进一步筛选出差异基因, 结合GO功能富集分析探讨相关生物学功能, 构建RNA-ADT网络并评估ADT重要性, 通过细胞通讯分析揭示细胞间动态信号传递, 对深入理解肺腺癌的分子机制并推动个体化治疗策略的制定具有重要意义。

数据和方法

1. 数据来源

从figshare平台(https://doi.org/10.6084/m9.figshare.c.5018987.v1)下载肺腺癌单细胞RNA和ADT组学数据, 使用R包Seurat[18]中的CreateSeuratObject函数进行基因和细胞筛选, 每个基因在至少5个细胞中有非零表达值, 每个细胞至少表达60个基因(特征), 进而筛选出RNA总数小于40000的细胞(nCount_RNA<40000)及线粒体基因占比低于15%的细胞(percent.mt<15), 预处理后肺腺癌样本包含3657个细胞, 15778个基因, 52个ADT。

2. CiteFuse方法

CiteFuse将批量测序的SNF方法拓展到单细胞CITE-seq数据整合, 采用指数核构建每个组学的细胞相似网络, 进一步利用信息传递方法进行细胞相似网络融合, 在融合网络基础上采用谱聚类识别细胞类型。

设n个细胞的基因表达和ADT计数的特征数分别为p和q, CiteFuse对归一化的基因表达矩阵和ADT计数矩阵分别采用Pearson相关系数以及perb相似性度量构建相关系数矩阵[19], 进一步将相关系数矩阵转化为距离矩阵D, 基于RNA和ADT的距离矩阵D, 采用指数核分别构建细胞相似网络$G=(V,E)$, 顶点V对应细胞$\{x_1, x_2, \cdots, x_n\}$, 边$E$根据细胞$x_i$和$x_j$的相似度$W(i,j)$进行加权:

$$W(i,j)=\exp\left(-\frac{(D(x_i,x_j))^2}{\mu\epsilon_{i,j}}\right) \tag{1}$$

其中, $D(x_i,x_j)$表示为细胞x_i和x_j之间的距离, μ为超参数, 设置为$\mu=0.5$, $\epsilon_{i,j}=\frac{\text{mean}(D(x_i,N_i))+\text{mean}(D(x_j,N_j))+D(x_i,x_j))}{3}$用于消除缩放比例问题, N_i表示细胞x_i的邻居集合。$\text{mean}(D(x_i,N_i))$是细胞x_i与其邻近细胞距离的平均值。

在细胞相似网络W基础上, 定义全局核矩阵P和局部核矩阵S, 其中, 矩阵P包含每个细胞与其他所有细胞的相似信息, 而矩阵S仅保留细胞与其近邻的相似信息, 计算公式如下:

$$P(i,j)=\begin{cases}\dfrac{W(i,j)}{2\Sigma_{k\neq i}W(i,k)}, & j\neq i\\ \dfrac{1}{2}, & j=i\end{cases} \tag{2}$$

$$S(i,j)=\begin{cases}\dfrac{W(i,j)}{\Sigma_{k\in N_i}W(i,k)}, & j\in N_i\\ 0, & \text{otherwise}\end{cases} \tag{3}$$

对RNA和ADT的细胞相似网络采用信息传递的方法进行网络融合, 得到最终的融合网络P_{final}:

$$P^v=S^v\times\left(\frac{\sum_{k\neq v}P^{(k)}}{r-1}\right)\times(S^{(v)})^T \tag{4}$$

$$P_{final}=\frac{\sum_{v=1}^{r}P^v}{v}, v=1,2,\cdots,m \tag{5}$$

融合后的相似网络P_{final}包含了RNA和ADT两个细胞相似网络的共享和互补信息。在融合网络P_{final}基础上, 采用谱聚类方法[20]识别细胞类型, 并将使两个连续特征值之差最大化的值作为最佳聚类数。

3. 细胞注释

通过R包SingleR[21]进行细胞注释, 参考基因列表为human primary cell atlas(HPCA)数据集, 包含36种

主要的细胞类型以及 157 种细胞亚型，对细胞进行主要细胞类型和细化亚型的双重细胞注释。

4. 差异表达分析与 GO 富集分析

采用 Wilcoxon 秩和检验筛选不同细胞类型间 RNA 和 ADT 差异特征，采用 Benjamini 和 Hochberg 方法[22]进行校正。进一步按以下条件过滤每个细胞类型的 RNA 和 ADT 标志物：①P_{adj} < 0. 05；②RNA 和 ADT 平均表达量高于其他所有细胞类型；③RNA 和 ADT 的细胞比例至少高于其他细胞类型的 10%。采用 R 包 scRNAtoolVis 对 RNA 和 ADT 的差异表达进行可视化分析。

将过滤后的差异基因输入 metascape 数据库[23]，设置为默认阈值，进行 GO 功能注释分析，获得差异基因在细胞组分（cellular component，CC）、分子功能（molecular function，MF）和生物学过程（biological process，BP）中的参与情况。

5. ADT−RNA 相关网络构建

通过计算 RNA 和 ADT 表达之间的 Pearson 相关系数构建 ADT−RNA 共表达网络，提取相关性较高的 ADT−RNA 对（相关系数大于 0. 6），使用 R 包 igraph[24] 的 visNetwork 函数可视化网络。

6. 细胞通讯分析

使用 R 包 CellChat[25] 识别细胞通讯和配体−受体分析，配体−受体库为 CellChatDB. human 数据集。

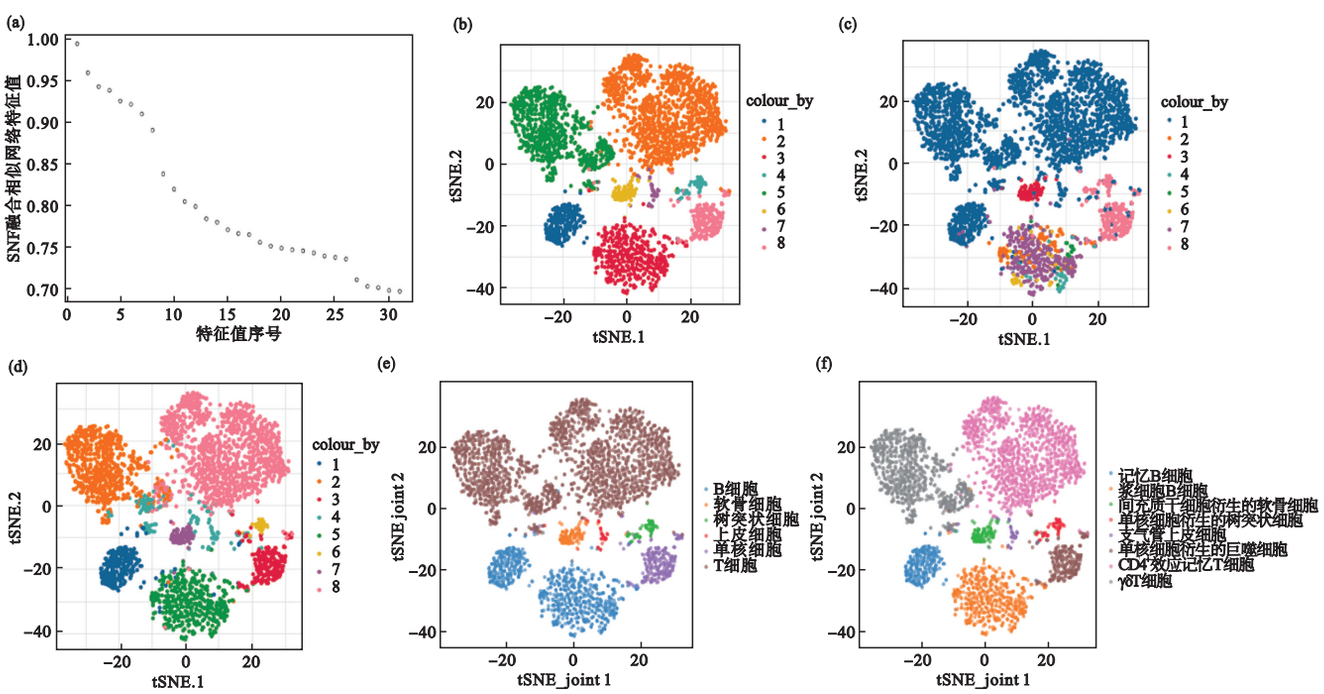
7. ADT 重要性评估

采用 R 包 CiteFuse[15] 的 importanceADT 函数评估每个 ADT 对细胞类型识别的重要性，保留重要性分数大于 5 的 ADT，对保留的重要 ADT 集与完整 ADT 集的细胞类型结果采用调整后的兰德指数（adjusted Rand index，ARI）进行一致性评估，ARI 取值范围为 0~1，ARI 越大，一致性越强。

结　果

1. 肺腺癌细胞类型识别

CiteFuse 整合的最佳聚类数为 8 类，如图 1（a）。CiteFuse 整合以及 RNA 和 ADT 单个组学的细胞类型识别结果，如图 1（b~d），RNA 单组学的细胞类型识别无法有效区分细胞类型，CiteFuse 整合的细胞类型识别结果优于 RNA 单组学。



（a）特征值分布图；（b）基于 CiteFuse 整合的细胞类型；（c）RNA 单组学细胞类型；（d）ADT 单组学细胞类型；（e）主要细胞类型注释；（f）细分亚型注释。

图 1　肺腺癌单细胞多组学数据细胞类型识别及细胞注释

对 CiteFuse 整合的细胞类型进行细胞注释，将其划分为 6 种主要细胞类型如图 1（e），包括 B 细胞、软骨细胞、树突状细胞、上皮细胞、单核细胞和 T 细胞。进一步细分为 8 种细胞亚型如图 1（f）：记忆 B 细胞、浆细胞 B 细胞、间充质干细胞衍生的软骨细胞、单核细胞衍生的树突状细胞、CD4$^+$效应记忆 T 细胞、γδT 细胞、支气管上皮细胞和单核细胞衍生的巨噬细胞，其中 T 细胞占比最高，其次为 B 细胞，见表 1。

2. 差异表达分析与 GO 富集可视化

过滤后共得到 75 个差异表达基因和 25 个差异表达 ADT，如图 2（a~b），其中，*MS4A1*、*CD37* 等基因在浆细胞 B 细胞差异高表达，*CD79A*、*TENT5C* 等基因在 γδT 细胞中高表达；*CD45RA*、*CD70* 等为浆细胞 B 细胞中的高表达蛋白标志物，*CD19*、*CD31* 等为 γδT 细胞中的高表达蛋白标志物，*CD39* 蛋白在 CD4$^+$效应记忆 T 细胞中高表达。图 2（c）为 *CD8A* 的 RNA 表达和 *CD8*

表1 不同细胞类型数量及比例

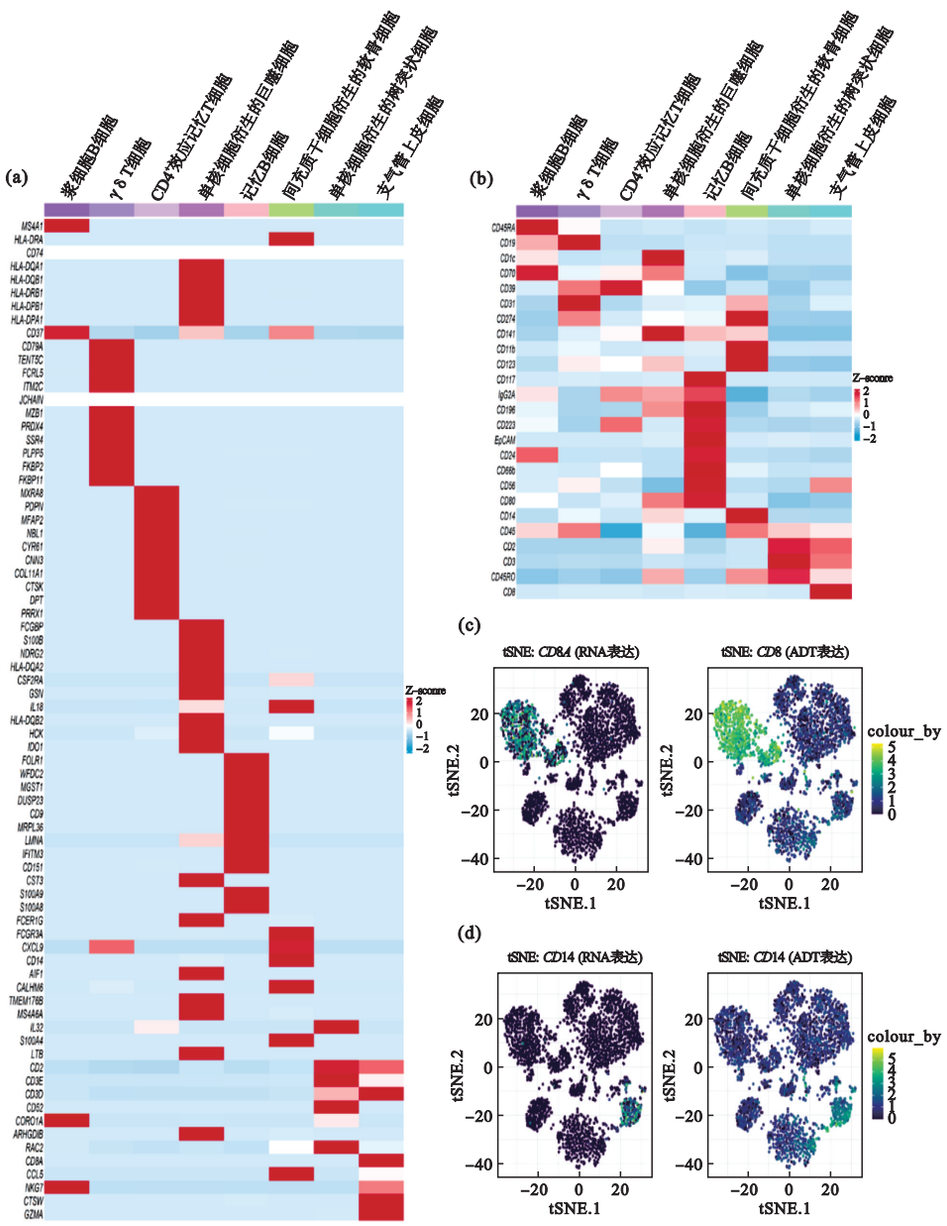
细胞类型	细胞数量	比例(%)
CD4⁺效应记忆T细胞	1333	36.45
γδT细胞	889	24.31
浆细胞B细胞	634	17.34
记忆B细胞	281	7.68
单核细胞衍生的巨噬细胞	265	7.25
间充质干细胞衍生的软骨细胞	116	3.17
单核细胞衍生的树突状细胞	80	2.19
支气管上皮细胞	59	1.61

的蛋白表达，图2(d)为*CD*14的RNA表达和ADT表达，RNA和ADT表达模式总体一致，反映了各细胞类型间的异质性。

将过滤后的差异基因进行GO功能注释共得到356个富集项，最显著的前10个集群如图3，桑基图展示富集项相关的基因，气泡图中气泡的位置为GeneRatio值，气泡的大小展示了富集到该项下的基因数量，气泡的颜色展示了该富集项的−log10(*P*)值，主要涉及膜侧(GO:0098552: side of membrane)、MHC蛋白复合物结合(GO:0023023: MHC protein complex binding)相关和参与细胞激活(GO:0001775: cell activation)过程等。

3. ADT−RNA相关网络构建

对差异表达基因和蛋白质构建相互作用网络，如图4，共识别到36个RNA−ADT对，节点表示蛋白质和RNA。网络中的中心节点CD19蛋白可能与多种基因相互作用，CD19与*IGHGP*、*IGLL5*、*HERPUD*1、*IGHG*1、*SSR*4、*XBP*1等基因通过红线相连，表示可能存在正向的相互作用，与*ACTB*基因通过蓝线相连，表示可能存在负向的相互作用。



(a)差异表达基因热图；(b)差异ADT表达热图；(c)*CD8A*的RNA表达和*CD8*的ADT表达；(d)*CD*14的RNA和ADT表达。

图2 差异表达RNA和ADT可视化

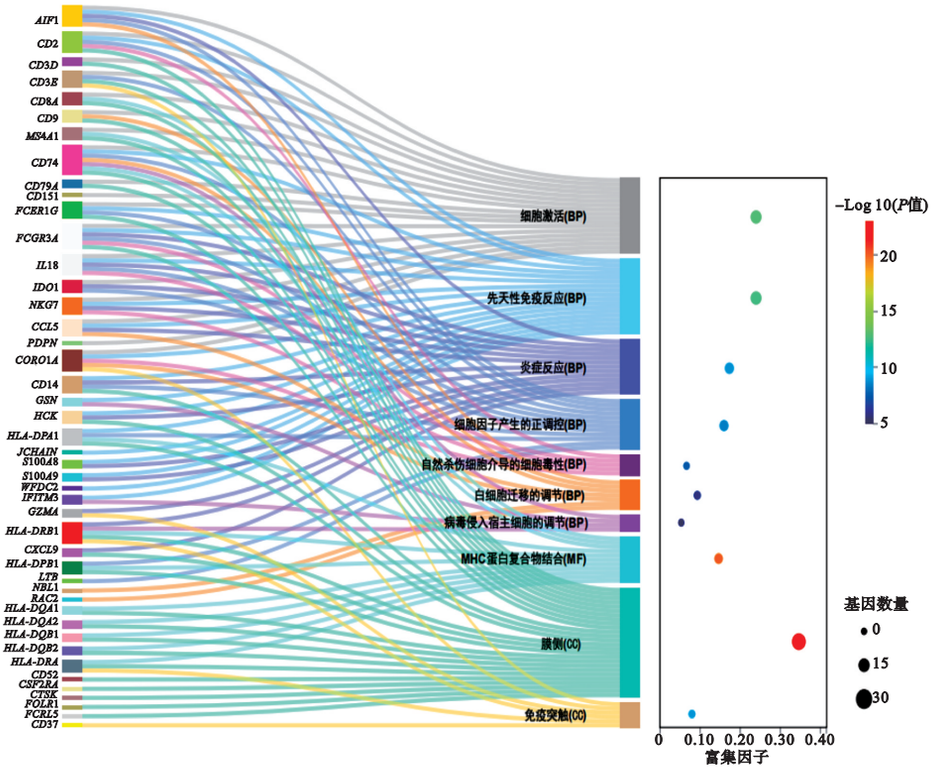


图 3　GO 富集分析桑基气泡图

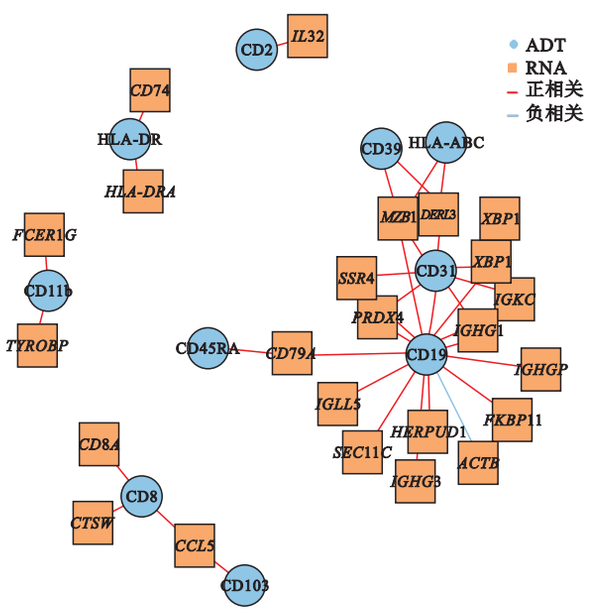


图 4　ADT−RNA 共表达网络

4. 细胞通讯分析

图 5 为细胞−细胞间通信网络图，每个颜色代表不同的细胞，结合箭头代表作用方向，线的粗细代表数量/权重，其中 CD4⁺效应记忆 T 细胞的细胞数最多，单核细胞衍生的树突状细胞与巨噬细胞的互作强度最高。

在 8 个细胞类型中检测到 316 个重要的配体−受体对，涉及 77 条信号通路，包括 MHC−Ⅱ、COLLAGEN、MHC−Ⅰ、MIF、FN1、LAMININ 和 VEGF 等通路。图 6 为 VEGF 信号网络，图 6(a)显示 B 细胞和软骨细胞类型之间的自分泌和旁分泌信号，图 6(b)为其余细胞类型。边宽度表示通信概率。

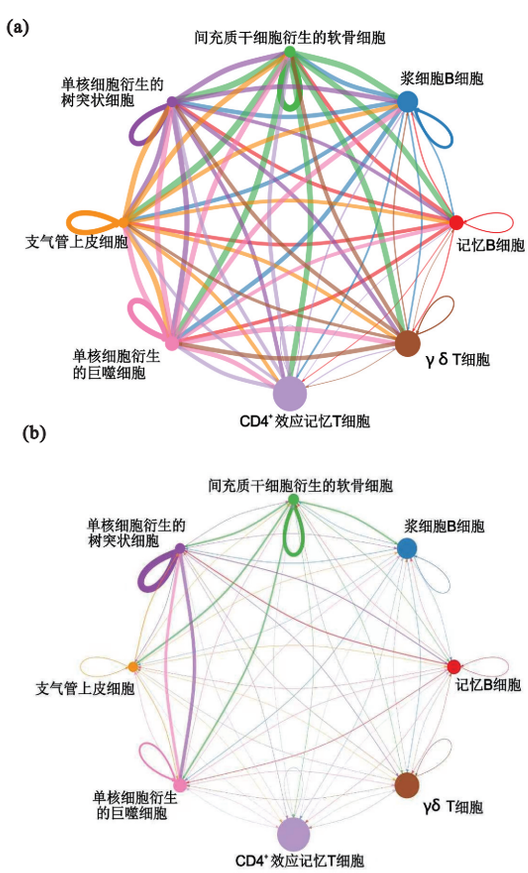
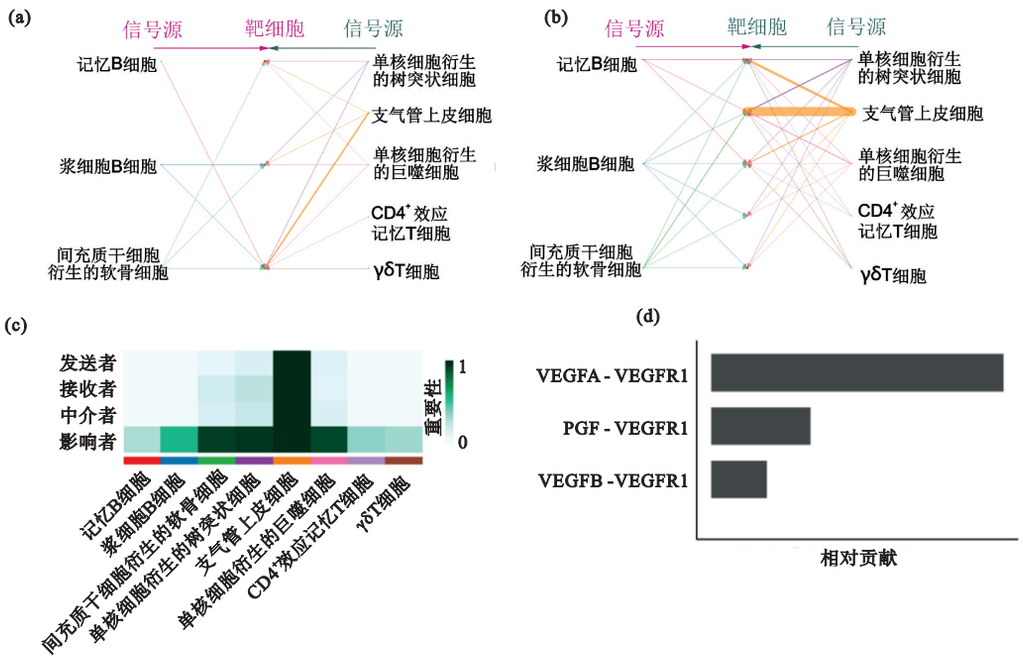


图 5　肺腺癌中细胞亚群间细胞通讯的数目(a)以及强度(b)

VEGF 信号网络表现复杂且高度冗余，多个配体来源共同作用于目标细胞。图 6(c)为 VEGF 信号网络中不同细胞类型的功能角色热图，大多数 VEGF 信号在细胞类型的相互作用为旁分泌，只有少数细胞类型(如

支气管上皮细胞）表现自分泌信号。图6（d）为VEGF信号网络中不同配体–受体对（L–R pair）的相对贡献，VEGF信号由VEGFA配体及VEGFR1受体主导。



注：（a）靶细胞介导的细胞间通讯网络；（b）发送细胞介导的细胞间通讯网络；（c）不同细胞类型的功能角色热图；（d）不同配体–受体对的相对贡献。

图6　VEGF信号网络可视化

5. ADT重要性评估

ADT重要性评估结果见图7，CD4和CD8是肺腺癌中的前两种区分蛋白，重要性分数大于5的ADT共46个，与采用完整ADT集的细胞类型识别结果比较，ARI为0.99，表明根据重要ADT识别出的细胞类型结果和采用完整ADT集识别的结果高度一致。

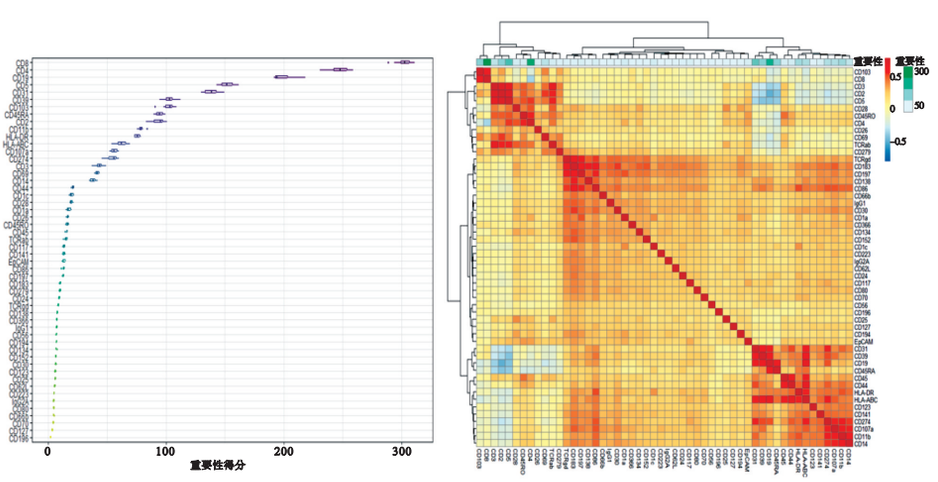


图7　ADT重要性得分图

讨　论

本文针对肺腺癌单细胞多组学数据，采用CiteFuse方法进行整合与细胞类型识别，共识别出6种主要细胞类型以及8种细胞亚型。基于细胞亚型得到的潜在生物标记物（基因、蛋白或信号通路），将为肺腺癌患者个体化治疗提供重要的参考依据。

识别的细胞亚型包括两类T细胞，即$CD4^+$效应记忆T细胞和γδT细胞；两类B细胞，即记忆B细胞和浆细胞B细胞以及间充质干细胞衍生的软骨细胞、单核细胞衍生的树突状细胞、支气管上皮细胞和单核细胞衍生的巨噬细胞，这与其他研究基本一致[14]。鉴定出差异表达的基因有*MS4A1*、*CD37*、*CD79A*、*TENT5C*等，差异表达的ADT有*CD45RA*、*CD70*、*CD19*、*CD39*等。其中，*MS4A*基因家族在细胞分化、信号传导和细胞周期控制中起着至关重要的作用[26]，既往研究表明*MS4A1*是肺腺癌的*MS4A*家族中突变的重要基因[27]，与正常组织相比，*MS4A1*表达上调。*CD79A*和*CD37*有助于监测特定的T–B细胞相互作用，还可以作为新的生物标志物，帮助评估肿瘤发生过程中癌症微环境

的变化[28]。*CD*37 是肺癌脑转移的有效生物标志物[29]。*TENT*5*C* 肿瘤抑制功能在肝癌中起作用[30]，还可用作结直肠癌的生物标志物[31]，在肺腺癌中的确切生物学作用需进一步阐明。*CD*45*RA* 在 $CD4^+$ 和 $CD8^+$T 细胞上的重新表达可能导致外周免疫活性降低[32]。*CD*70 过表达在局部晚期非小细胞肺癌患者的肿瘤样本中已得到证实，并在小细胞肺癌中高表达，*CD*70 的表达与 $CD3^+$ 和 $CD8^+$T 细胞浸润减少以及 T 细胞耗竭标志物增加相关[33]。*CD*19 信号通过控制 B 细胞浸润到肺组织，与肺纤维化的发展有关[34]。在 ADT−RNA 相关网络中，*CD*19 与 *IG* 家族基因（如 *IGHG*1）相关，有研究提示 *CD*19 可能在 B 细胞的免疫反应中与表面免疫球蛋白共同促进信号转导[35]。

在富集分析结果中，膜侧、与 MHC 蛋白复合物结合功能、细胞激活过程等可能与肺腺癌的免疫微环境或免疫反应相关。膜侧的变化可能会影响细胞间的信号传递和免疫细胞的相互作用。MHC Ⅱ类分子能够将抗原肽呈递给 $CD4^+$T 细胞，激活特异性免疫反应[36]。MHC Ⅰ类相关肽表位是比过表达蛋白更相关的真实肿瘤抗原来源[37]，并且鉴定出的肽可作为治疗性疫苗的候选抗原，用于肺癌的免疫治疗。

本研究基于细胞通讯分析探索各细胞类型间的相互作用及其信号通路网络。其中，血管内皮生长因子（vascular endothelial growth factor，VEGF）为血管生成刺激因子，在 VEGF 信号通路中，VEGF 与 VEGFR2 结合后，可通过激活 PI3K 在 PI3K−Akt 信号通路中调控细胞迁移、生存和通透性[38]，进一步阐明 VEGF 通路的病理功能将推动肺腺癌在预后、诊断和治疗方面的进展。VEGF−VEGFR1 信号通路对肿瘤转移至关重要，阻断 VEGF−VEGFR1 诱导的转移可能为预防和治疗肿瘤转移提供新的方法[39]。有研究表明肺泡巨噬细胞和肺上皮细胞作为肺部“守门人”，是宿主防御和先天免疫的第一道防线，两者相互作用以维持肺稳态和气体交换[40-41]，关于肺腺癌的更多细胞间相互作用与机制需进一步探索。此外，CD4 和 CD8 作为重要区分蛋白，它们表达的差异有助于区分不同类型的 T 细胞，还能够反映免疫系统的活性和功能状态[42]。

综上所述，本文基于 CiteFuse 方法对肺腺癌单细胞多组学数据的整合分析，能够有效识别细胞亚型，不同细胞类型在基因表达、细胞通讯和调控方面均存在差异，有助于进一步理解肺腺癌发生发展的潜在分子机制。此外，筛选出的生物标志物将为肺腺癌个体化治疗提供一定的理论依据和临床指导。

参 考 文 献

[1] Li W, Liang H, Wang W, et al. Global cancer statistics for adolescents and young adults: population based study [J]. Journal of Hematology & Oncology, 2024, 17(1): 99.

[2] Wood DE, Kazerooni EA, Aberle D, et al. NCCN guidelines insights: lung cancer screening, version 1. 2022: featured updates to the NCCN guidelines[J]. Journal of the National Comprehensive Cancer Network, 2022, 20(7): 754-764.

[3] 刘夏,徐昭阳,刘改琴,等. 肺癌空间转录组学数据空间表达基因的识别研究[J]. 中国卫生统计,2026,43(2):202-207.

[4] 高梦瑶,穆华夏,卜伟晓,等. 基于TCGA数据库肺腺癌的风险预测模型的构建与验证[J]. 中国卫生统计,2025,42(2):191-196.

[5] 张凌谱,杨佳臻,唐东升,等. 靶向肿瘤进化的创新纳米药物[J]. 中国科学:化学,2022,52(12):2121-2155.

[6] 刘亚飞,叶贯超,黄岚,等. 单细胞转录组在晚期肺腺癌患者个体化治疗中的应用[J]. 中华实验外科杂志,2020,37(5):948-950.

[7] Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells [J]. Nature Methods, 2017, 14(9): 865-868.

[8] Peterson VM, Zhang KX, Kumar N, et al. Multiplexed quantification of proteins and transcripts in single cells [J]. Nature Biotechnology, 2017, 35(10): 936-939.

[9] Lee J, Hyeon DY, Hwang D. Single-cell multiomics: technologies and data analysis methods [J]. Experimental & Molecular Medicine, 2020, 52(9): 1428-1442.

[10] Stanojevic S, Li Y, Ristivojevic A, et al. Computational methods for single-cell multi-omics integration and alignment [J]. Genomics, Proteomics and Bioinformatics, 2022, 20(5): 836-849.

[11] Argelaguet R, Arnol D, Bredikhin D, et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data [J]. Genome Biology, 2020, 21(1): 111.

[12] Lopez R, Regier J, Cole MB, et al. Deep generative modeling for single-cell transcriptomics [J]. Nature Methods, 2018, 15(12): 1053-1058.

[13] Zuo C, Chen L. Deep-joint-learning analysis model of single cell transcriptome and open chromatin accessibility data [J]. Briefings in Bioinformatics, 2021, 22(4): bbaa287.

[14] Ma A, Wang X, Wang C, et al. Deepmaps: Single-cell biological network inference using heterogeneous graph transformer [EB/OL]. (2021-12-11) [2025-06-24]. https://www. biorxiv. org/content/10. 1101/2021. 10. 31. 466658v2.

[15] Kim HJ, Lin Y, Geddes TA, et al. CiteFuse enables multi-modal analysis of CITE-seq data [J]. Bioinformatics, 2020, 36(14): 4137-4143.

[16] Kuchroo M, Godavarthi A, Tong A, et al. Multimodal data visualization and denoising with integrated diffusion[C]//2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2021: 1-6.

[17] Wang B, Mezlini AM, Demir F, et al. Similarity network fusion for aggregating data types on a genomic scale [J]. Nature Methods, 2014, 11(3): 333-337.

[18] Satija R, Farrell JA, Gennert D, et al. Spatial reconstruction of single-cell gene expression data [J]. Nature Biotechnology, 2015, 33(5): 495-502.

[19] Quinn TP, Richardson MF, Lovell D, et al. Propr: an R-package for identifying proportionally abundant features using compositional data analysis [J]. Scientific Reports, 2017, 7(1): 16252.

[20] Ng A, Jordan M, Weiss Y. On spectral clustering: Analysis and an algorithm [C]. Cambridge: MIT Press, 2001.
[21] Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage [J]. Nature Immunology, 2019, 20(2): 163-172.
[22] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing [J]. Journal of the Royal Statistical Society: Series B (Methodological), 1995, 57 (1): 289-300.
[23] Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets [J]. Nature Communications, 2019, 10(1): 1523.
[24] Csardi MG. Package 'igraph' [J]. Last Accessed, 2013, 3(9): 2013.
[25] Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat [J]. Nature Communications, 2021, 12(1): 1088.
[26] Silva-Gomes R, Mapelli SN, Boutet MA, et al. Differential expression and regulation of MS4A family members in myeloid cells in physiological and pathological conditions [J]. Journal of Leukocyte Biology, 2022, 111(4): 817-836.
[27] Zheng Z, Li H, Yang R, et al. Role of the membrane-spanning 4A gene family in lung adenocarcinoma [J]. Frontiers in Genetics, 2023, 14: 1162787.
[28] Huang GH, Zhang YH, Chen L, et al. Identifying lung cancer cell markers with machine learning methods and single-cell rna-seq data [J]. Life, 2021, 11(9): 940.
[29] Dohm A, Su J, McTyre ER, et al. Identification of CD37, cystatin A, and IL-23A gene expression in association with brain metastasis: analysis of a prospective trial [J]. The International Journal of Biological Markers, 2019, 34(1): 90-97.
[30] Min X, Lin F, Zhao X, et al. TENT5A mediates the cancer-inhibiting effects of EGR1 by suppressing the protein stability of RPL35 in hepatocellular carcinoma [J]. Cellular Oncology, 2024, 47 (6): 2247-2264.
[31] Han W, Li C, Wang Y, et al. Heme Metabolism-Related Gene TENT5C is a Prognostic Marker and Investigating Its Immunological Role in Colon Cancer [J]. Pharmacogenomics and Personalized Medicine, 2023, 16: 1127-1143.
[32] d'Alessandro M, Bergantini L, Gangi S, et al. Imbalance of Lymphocyte Subsets and CD45RA-Expressing Cells in Intrathoracic Lymph Nodes, Alveolar Compartment and Bloodstream of Pulmonary Sarcoidosis Patients [J]. International Journal of Molecular Sciences, 2023, 24(12): 10344.
[33] Ortiz-Cuaran S, Swalduz A, Foy JP, et al. Epithelial-to-mesenchymal transition promotes immune escape by inducing CD70 in non-small cell lung cancer [J]. European Journal of Cancer, 2022, 169: 106-122.
[34] Komura K, Yanaba K, Horikawa M, et al. CD19 regulates the development of bleomycin-induced pulmonary fibrosis in a mouse model [J]. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology, 2008, 58(11): 3574-3584.
[35] Pesando JM, Bouchard LS, McMaster BE. CD19 is functionally and physically associated with surface immunoglobulin [J]. The Journal of experimental medicine, 1989, 170(6): 2159-2164.
[36] Roche PA, Furuta K. The ins and outs of MHC class Ⅱ-mediated antigen processing and presentation [J]. Nature Reviews Immunology, 2015, 15(4): 203-216.
[37] Shetty V, Sinnathamby G, Nickens Z, et al. MHC class Ⅰ-presented lung cancer-associated tumor antigens identified by immunoproteomics analysis are targets for cancer-specific T cell response [J]. Journal of Proteomics, 2011, 74(5): 728-743.
[38] Yan Z, Ping Z. Reviews and prospectives of signaling pathway analysis in idiopathic pulmonary fibrosis [J]. Autoimmunity Reviews, 2014, 13(10): 1020-1025.
[39] Liu W, Xu J, Wang M, et al. Tumor-derived vascular endothelial growth factor (VEGF)-a facilitates tumor metastasis through the VEGF-VEGFR1 signaling pathway [J]. International Journal of Oncology, 2011, 39(5): 1213-1220.
[40] Tao F, Kobzik L. Lung macrophage–epithelial cell interactions amplify particle-mediated cytokine release [J]. American Journal of Respiratory Cell and Molecular Biology, 2002, 26(4): 499-505.
[41] Bissonnette EY, Lauzon-Joset JF, Debley JS, et al. Cross-talk between alveolar macrophages and lung epithelial cells is essential to maintain lung homeostasis [J]. Frontiers in Immunology, 2020, 11: 583042.
[42] Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion [J]. Science, 2011, 331(6024): 1565-1570.

（责任编辑：邓　妍）