

SDAS应用案例-SDAS_beta.2

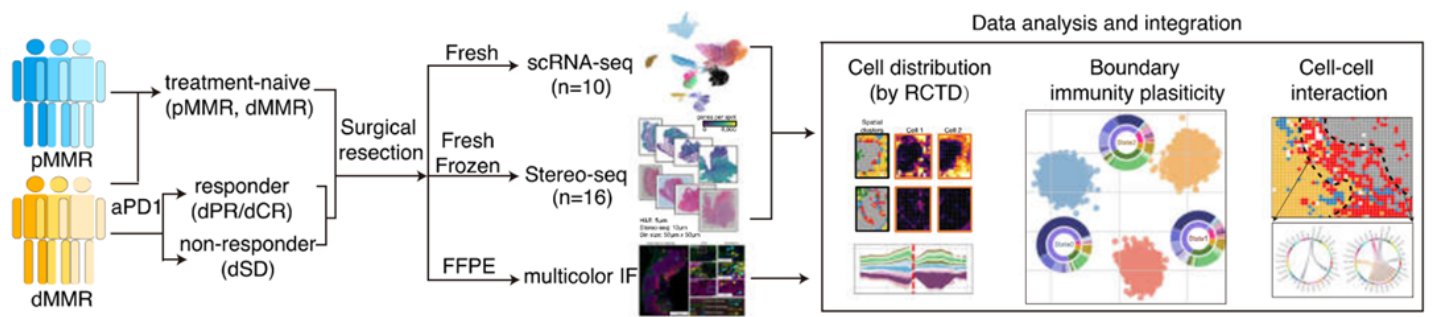
1. 文献说明

文章标题: Spatially organized tumor-stroma boundary determines the efficacy of immunotherapy in colorectal cancer patients

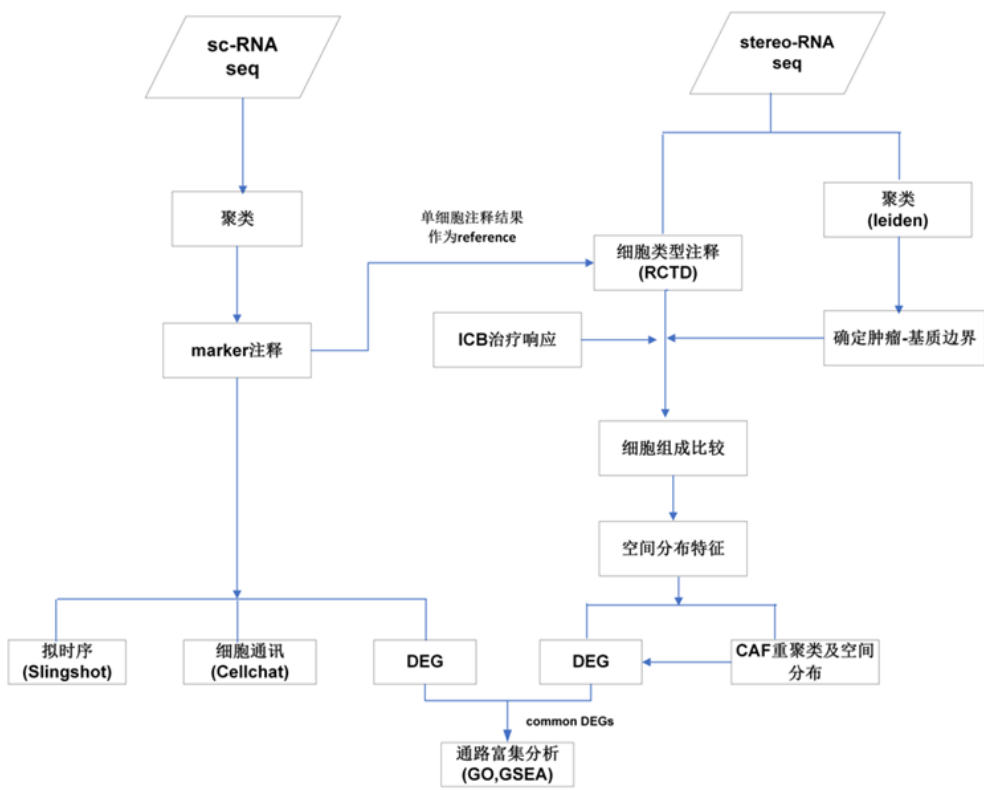
期刊: Nature Communications **影响因子:** 14.7

摘要: 本研究应用Stereo-seq和scRNA-seq，绘制CRC患者的高分辨率TME（肿瘤微环境）空间图谱，重点分析肿瘤基质边界的细胞与分子特征及其在ICB（免疫检查点抑制剂）响应中的作用机制。

样本分组: 1) 未做治疗：pMMR vs. dMMR 2) ICB治疗(dMMR)：dPR/dCR vs. dSD



下图为该文献的生信分析流程图，其中scRNA-seq的聚类 and 细胞类型注释直接使用文章提供的数据，其他分析均进行了复现。



2. 分析准备



SDAS软件包: SDAS_beta.2.tar.gz

文章数据: ./Application_cases/Application_test_data

single_cell.h5ad: 单细胞数据, 包含注释信息

stereo_bin100.h5ad: 16个样本的stereo-seq的bin100数据

运行脚本和Jupyter代码: ./Application_cases/Application_scripts

3. 分析步骤

脚本参数说明

- 1 `${tool_dir}`为软件解压后的文件夹路径
- 2 `${output_dir}`为输出结果的文件夹路径
- 3 `${single_cell_h5ad}`为文章的单细胞数据, `single_cell.h5ad`的绝对路径
- 4 `${stereo_h5ad}`为文章的stereo数据, `stereo_bin100.h5ad`的绝对路径
- 5 以下分析中标蓝色背景为需要通过脚本或jupyter notebook实现, 未包含在SDAS软件中

3.1 CCA去批次和leiden聚类

对stereo-seq数据进行多片CCA去批次, leiden聚类, 根据marker定义tumor-stroma boundary bins和其他大类。该功能未包含在SDAS中, 尝试使用SAW aggr的harmony去批次无法复现(不同去批次的算法效果有差异)。因此该部分通过jupyter notebook实现。未来会考虑将SAW aggr的多片去批次算法进行扩展。

- 1 *#Step1. 将stereo_bin100.h5ad转成rds格式, 输出文件为stereo_bin100.rds*
- 2 `${tool_dir}/SDAS dataProcess h5ad2rds -i ${stereo_h5ad} --run_mode stRNA -o ${output_dir}`
- 3
- 4 *#Step2. 在seurat环境中, 进行CCA去批次, 存储integrated后的表达矩阵和meta.data信息 (运行时间: 12h, 最大内存: 324G)*
- 5 *#环境可以用SDAS的数据处理模块的环境: `${tool_dir}/data_process/anaconda/bin/Rscript`*
- 6 `Rscript ./Application_cases/Application_scripts/Figure1_CCA_in_seurat.r`
- 7
- 8 *#Step3. 在scanpy环境中, 读入CCA后的表达矩阵和meta.data信息创建h5ad对象 (运行时间: 1h17m, 最大内存: 169G)*
- 9 *#环境可以用SDAS的数据处理模块的环境: `${tool_dir}/data_process/anaconda/bin/python`*

```

10 python
   ./Application_cases/Application_scripts/Figure1_generate_h5ad_after_CCA.py
11
12 #Step4. 在scanpy环境中, 进行leiden聚类, 可自行调整resolution参数; 根据文献提供的
   marker list, 绘制表达heatmap图, 并对cluster进行细胞类型定义
13 #环境可以用SDAS的数据处理模块的环境: ${tool_dir}/data_process/anaconda/bin/python
14 ./Application_cases/Application_scripts/Figure1_clustering_in_scanpy.ipynb

```

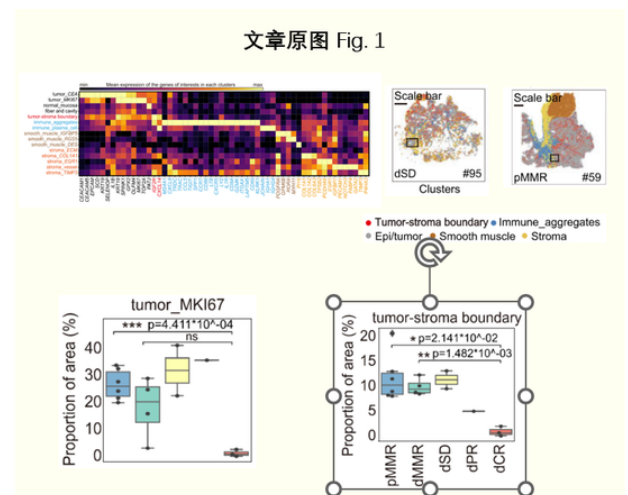
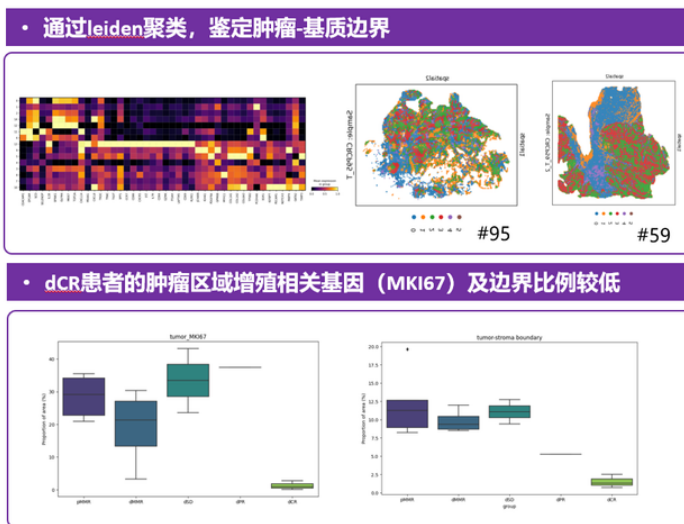
结果解读:

- dCR患者表现出肿瘤区域增殖相关基因（如MKI67）的表达显著减少，表明ICB治疗可能抑制了肿瘤细胞增殖
- dCR和dPR的肿瘤边界比例降低，可能与更高的免疫浸润能力相关

时空组学
STOmics

结果1: CRC空间转录图谱构建, leiden聚类, 确定肿瘤-基质边界

dCR患者表现出肿瘤区域增殖相关基因（如MKI67）的表达显著减少，表明ICB治疗可能抑制肿瘤细胞增殖
dCR和dPR的肿瘤边界比例降低，可能与更高的免疫浸润能力相关



说明：由于算法中的随机部分（CCA选取的高变基因可能有差异）和细节的参数设置，聚类结果与原文存在一定的差异，定义为boundary的bins不完全一致，为不影响其他模块的复现结果，后续的分析 and 解读均使用文章定义的boundary信息。

3.2 RCTD细胞类型注释

对stereo-seq数据进行RCTD细胞类型注释，分析boundary区域的细胞分布及特征，比较不同分组的免疫细胞频率

```

1 #reference为文章的单细胞数据single_cell.h5ad, 注释结果文件为
   stereo_bin100_anno_rctd.h5ad(运行时间: 7h, 最大内存40.5G)

```

```

2  ${tool_dir}/SDAS cellAnnotation rctd -i ${stereo_bin100.h5ad} -o $output_dir -
--reference ${single_cell_h5ad} --label_key annotation2 --bin_size 100 --
filter_rare_cell 0 --input_gene_symbol_key _index --slice_key id
3
4  #分析边界区域随距离变化的免疫细胞的分布曲线，不同分组的免疫细胞频率比较
5  ./Application_cases/Application_scripts/stereo_plot.ipynb

```

结果解读：

1) 免疫细胞在边界的动态分布

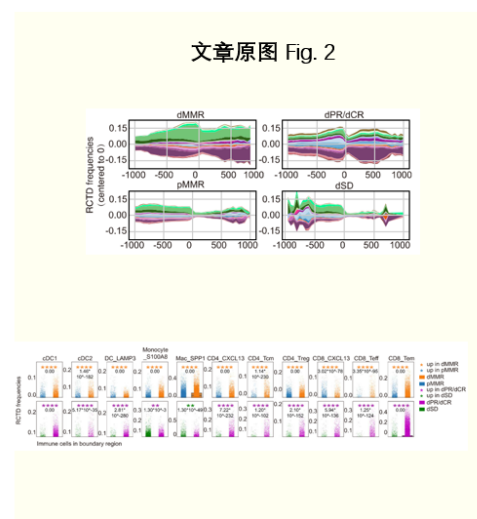
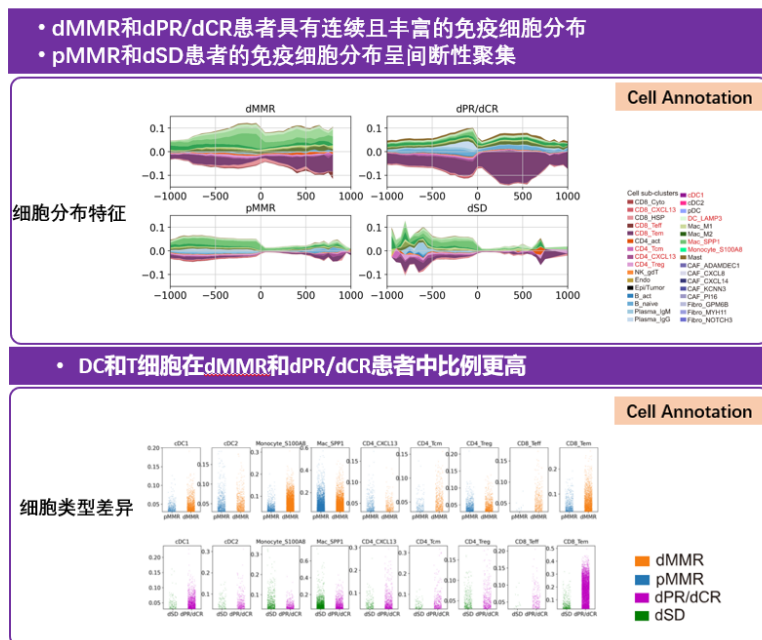
- 沿边界的免疫细胞分布曲线显示，pMMR患者的免疫细胞分布呈间断性聚集，而dMMR患者具有连续且丰富的免疫细胞分布，尤其在边界区域（ $0 \pm 150\mu\text{m}$ ）内

2) 免疫细胞富集与ICB疗效相关

- dMMR患者中CD8+效应T细胞（CD8_Teff）、记忆T细胞（CD8_Tem）以及DC_LAMP3显著富集于边界区域，局部免疫细胞聚集可能增强抗肿瘤免疫反应

■ 结果2 肿瘤-基质边界的免疫细胞特征

不同患者对ICB治疗的预后效果差异可能是由于肿瘤-基质边界的空间特征导致，局部免疫细胞聚集可能增强抗肿瘤免疫反应



3.3 boundary的DEG和富集分析

对stereo-seq数据提取boundary的bins，对dMMR和pMMR两个分组进行DEG和GSEA分析

```

1  #Step1. 提取通过聚类定义为boundary的bin
2  ${tool_dir}/SDAS dataProcess subsetAdata -i ${stereo_bin100.h5ad} --label_key
level1 --list_include boundary -o $output_dir
3

```

```

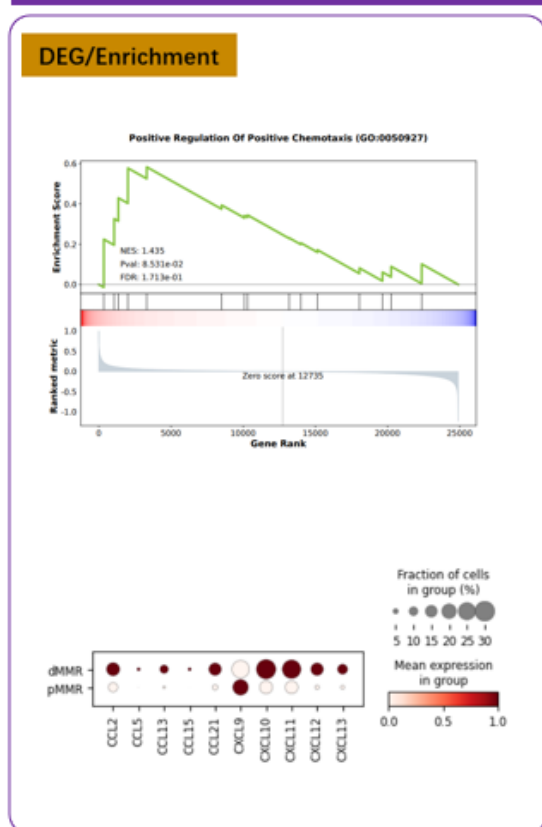
4 #Step2. 对dMMR和pMMR两个分组进行DEG分析
5 ${tool_dir}/SDAS DEG -i ${stereo_bin100_subset.h5ad} --diff_plan
boundary_deg_plan.csv -o $output_dir
6
7 #Step3. GSEA分析(运行时间11min, 最大内存14.3G), 绘制文章指定通路的GSEA图
8 ${tool_dir}/SDAS geneSetEnrichment gsea -i ${stereo_bin100_subset.h5ad} --
gsea_plan boundary_gsea_plan.csv --species human --gmt
${tool_dir}/sdas_deg_enrichment/lib/GSEADB/GO_Biological_Process_2023.gmt --
pathways boundary_pathway.txt -o $output_dir

```

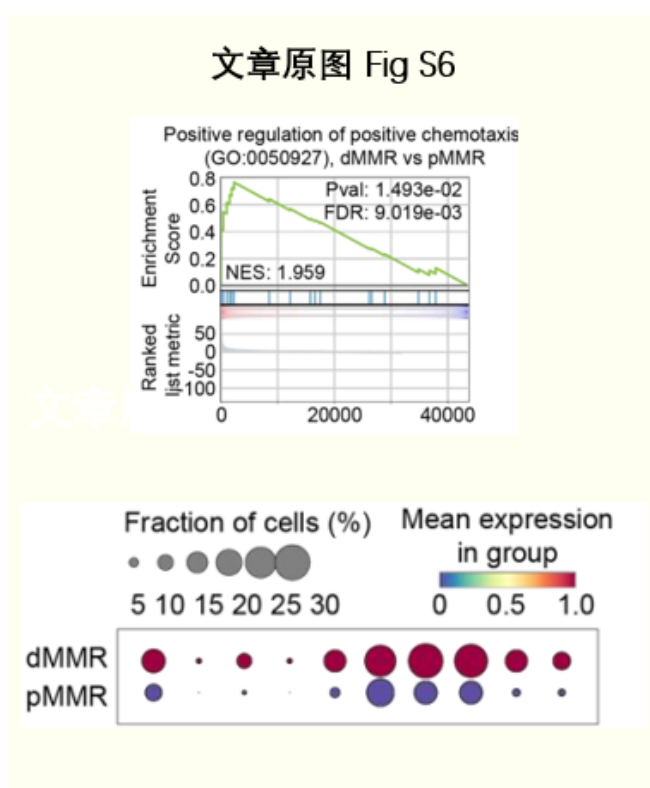
结果解读：

对dMMR和pMMR的边界区域进行GSEA分析，dMMR在正向趋化性（GO：0050927）的正向调控通路显著富集，且与pMMR相比，趋化相关基因CCL2/5/13/15/21 和 CXCL9/10/11/12/13 的表达明显较高。同时在不同细胞类型的比例比较中发现树突状细胞（DC）和T细胞在dMMR中较高，因此推测DC和T细胞可能通过趋化途径被募集到ICB治疗响应者的肿瘤基质边界。

dMMR vs. pMMR 边界区域的 DEG及通路富集



文章原图 Fig S6



3.4 细胞通讯分析

对单细胞数据提取样本类型为dMMR的数据进行细胞通讯分析，尝试找出DC和T细胞在ICB治疗的响应机制。


```

2  ${tool_dir}/SDAS dataProcess subsetAdata -i ${single_cell_h5ad} -o
   $output_dir --label_key annotation2 --list_include
   CAF_ADAMDEC1,CAF_CXCL8,CAF_CXCL14,CAF_KCNN3,CAF_PI16
3
4  #Step2. 过滤基因数<350的细胞, 使用BBKNN去批次, leiden聚类后生成h5ad文件
   ${single_cell_subset_umap.h5ad}
5  #环境可以用SDAS的数据处理模块的环境: ${tool_dir}/data_process/anaconda/bin/python
6  ./Application_cases/Application_scripts/CAF_umap.ipynb
7
8  #Step3. 将聚类后的h5ad转成rds
9  ${tool_dir}/SDAS dataProcess h5ad2rds --run_mode scRNA --layer counts -i
   ${single_cell_subset_umap.h5ad} -o ${output_dir}
10
11 #Step4. 进行拟时序分析, UMAP和cluster使用rds文件中自带的UMAP和cluster结果 (运行时间
    7min, 最大内存1.4G)
12 ${tool_dir}/SDAS trajectory monocle3 --input ${single_cell_subset_umap.rds} --
    root_key annotation2 --root CAF_PI16 --use_existing_umap_cluster --umap_key
    umap --cluster_key leiden --run_mode scRNA --gene_symbol_key _index --output
    $output_dir

```

- CXCL14+CAF与CXCL8+CAF的分化路径分析表明, CXCL14+CAF通过增强细胞外基质 (ECM) 组织化, 抑制T细胞浸润, 从而形成免疫排斥的边界结构。
- ICB治疗无响应的患者与肿瘤-基质边界区域内的CAF (尤其是CXCL14+CAF) 比例升高密切相关

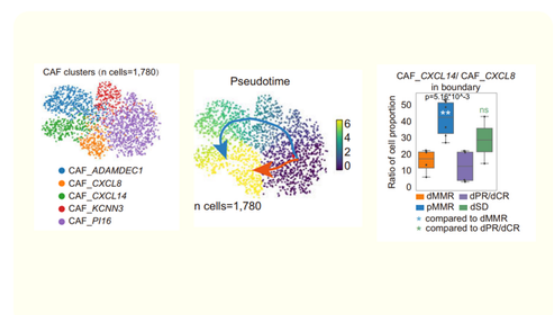
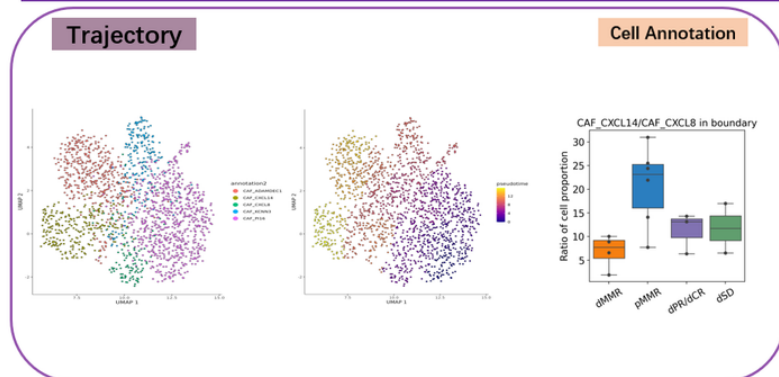
■ 结果4 CAF可塑性与免疫排斥的关联

CAF的分化路径为以CAF_PI16为起点, 分化成CXCL14+CAF与CXCL8+CAF

ICB治疗无响应的患者与肿瘤-基质边界区域内的CAF (尤其是CXCL14+CAF) 比例升高密切相关

CAF 重聚类+拟时序分析: 以CAF_PI16为起点, 分化成CXCL14+CAF与CXCL8+CAF
pMMR和dSD患者的CAF_CXCL14占比显著更高

文章原图 Fig. 4



3.6 CAF的DEG和富集分析

通过DEG和功能富集分析阐述不同CAF亚型具体的功能, 与ICB治疗响应的关系。

```

1 #Step1. 提取单细胞数据中CAF的细胞类型
2 ${tool_dir}/SDAS dataProcess subsetAdata -i ${single_cell_h5ad} -o
  $output_dir --label_key annotation2 --list_include
  CAF_ADAMDEC1,CAF_CXCL8,CAF_CXCL14,CAF_KCNN3,CAF_PI16
3
4 #Step2. 对CAF_CXCL14和CAF_CXCL8CAF两组细胞类型进行DEG分析
5 ${tool_dir}/SDAS DEG -i ${single_cell_subset.h5ad} --diff_plan
  CAF_deg_plan.csv -o $output_dir
6
7 #Step3. 对DEG进行GO富集
8 ${tool_dir}/SDAS geneSetEnrichment enrichr -i
  ${output_dir}/CAF_CXCL14.vs.CAF_CXCL8.deg_filtered.csv --species human --gmt
  ${tool_dir}/sdas_deg_enrichment/lib/GSEADB/GO_Biological_Process_2023.gmt --
  cut_off 0.05 -o $output_dir
9
10 #Step4. GSEA分析(运行时间10min, 最大内存14.7G), 绘制文章指定通路的GSEA图
11 ${tool_dir}/SDAS geneSetEnrichment gsea -i ${single_cell_subset.h5ad} --
  gsea_plan CAF_gsea_plan.csv --species human --gmt
  ${tool_dir}/sdas_deg_enrichment/lib/GSEADB/GO_Biological_Process_2023.gmt --
  pathways CAF_pathway.txt -o $output_dir

```

CAF在基质屏障形成中的作用

- CXCL14+CAF通过调节ECM结构，形成有序的基质屏障，排斥T细胞进入肿瘤核心区域
- ICB治疗响应低的患者（dSD）中有较高比例的CXCL14+CAF细胞类型，导致对免疫细胞的排斥，影响免疫治疗效果

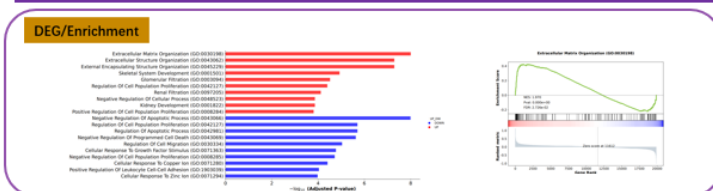
时空组学
STOmics

结果5 CAF在基质屏障形成中的作用，其中CAF_CXCL14与ECM形成相关

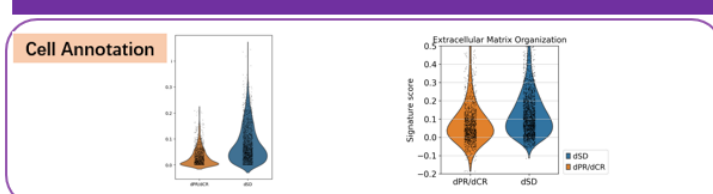
CXCL14+CAF通过调节ECM结构，形成有序的基质屏障，排斥T细胞进入肿瘤核心区域

ICB治疗响应低的患者（dSD）中有较高比例的CXCL14+CAF细胞类型，导致对免疫细胞的排斥，影响免疫治疗效果

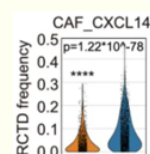
CXCL14+CAF vs. CXCL8+CAF DEG及富集分析，CXCL14+CAF主要富集在ECM形成通路



dSD患者的CXCL14+CAF细胞占比更高，ECM通路相关基因的评分更高



文章原图 Fig. 5



3.7 图表绘制

文章复现的图使用jupyter notebook进行绘制，数据均来源于上述功能模块的分析结果。第二阶段将考虑将图表绘制功能标准化，分享给用户使用。

```
1 #环境可以用SDAS的注释模块的环境：${tool_dir}/anno/anaconda/bin/python
2 ./Application_cases/Application_scripts/stereo_plot.ipynb
```

4. reference

Feng, Y., Ma, W., Zang, Y. *et al.* Spatially organized tumor-stroma boundary determines the efficacy of immunotherapy in colorectal cancer patients. *Nat Commun* **15**, 10259 (2024).
<https://doi.org/10.1038/s41467-024-54710-3>