

# HNRNPH1、Wnt 与瘢痕增生的关联性挖掘

2024-05-13

LiChuang Huang



@ 立效研究院

# Contents

<b>1 摘要</b>	<b>1</b>
1.1 需求 . . . . .	1
1.2 结果 . . . . .	1
<b>2 前言</b>	<b>1</b>
<b>3 材料和方法</b>	<b>1</b>
3.1 材料 . . . . .	1
3.2 方法 . . . . .	2
<b>4 分析结果</b>	<b>2</b>
4.1 HNRNPH1、Wnt 与 PPI 网络分析 . . . . .	2
4.2 HNRNPH1、Wnt 与斑痕的 scRNA-seq 分析 . . . . .	3
4.3 HNRNPH1、Wnt 与姜黄素 . . . . .	8
<b>5 结论</b>	<b>10</b>
<b>6 附：分析流程</b>	<b>10</b>
6.1 关于源代码的说明 . . . . .	10
6.1.1 生成该 PDF 文档的源代码 . . . . .	10
6.1.2 环境变量配置 . . . . .	10
6.1.3 R 包 . . . . .	11
6.2 关于本文档源代码中涉及的文件 . . . . .	11
6.3 HNRNPH1、Wnt 与 PPI 网络分析 . . . . .	12
6.3.1 姜黄素的 mRNA-seq 数据 . . . . .	12
6.3.2 DEGs . . . . .	12
6.3.3 wnt 信号通路 . . . . .	15
6.3.4 构建 PPI 网络 . . . . .	16
6.3.4.1 DEGs PPI . . . . .	16
6.3.4.2 HNRNPH1 与 Wnt 通路 . . . . .	16
6.4 HNRNPH1、Wnt 与斑痕的 scRNA-seq 分析 . . . . .	18
6.4.1 数据来源 . . . . .	18
6.4.2 细胞聚类和鉴定 . . . . .	19
6.4.3 差异分析 . . . . .	21
6.4.3.1 HNRNPH1 的表达 . . . . .	22
6.4.3.2 Wnt 通路基因的表达 . . . . .	23
6.4.4 拟时分析 . . . . .	25
6.4.4.1 拟时终点与 APCDD1 . . . . .	25
6.4.4.2 APCDD1 主要在 Scar 中高表达 . . . . .	27
6.4.4.3 Fibroblast 拟时轨迹下的差异基因 . . . . .	30
6.4.4.4 姜黄素有调控作用的靶点 (所有的差异基因中) . . . . .	32

6.4.5	细胞通讯 . . . . .	33
6.4.5.1	总体通讯 . . . . .	33
6.4.5.2	Fibroblast 分支与免疫细胞 . . . . .	36
6.4.5.3	差异通讯 . . . . .	37
6.4.5.4	Others . . . . .	38
6.4.5.5	蛋白互作 (PPI) . . . . .	39
6.5	HNRNPH1、Wnt 与姜黄素 . . . . .	41
6.5.1	分子对接结果 . . . . .	41
6.5.2	可视化 . . . . .	42
6.6	Session Info . . . . .	44
<b>Reference</b>		<b>46</b>

## List of Figures

1	MAIN Fig 1 . . . . .	3
2	MAIN Fig 2 . . . . .	5
3	MAIN Fig 3 . . . . .	6
4	MAIN Fig 4 . . . . .	7
5	MAIN Fig 5 . . . . .	8
6	MAIN Fig 6 . . . . .	9
7	Treat vs control DEGs . . . . .	13
8	DEG hsa04310 visualization . . . . .	15
9	PPI HNRNPH1 and Wnt . . . . .	17
10	UMAP Clustering . . . . .	20
11	SCSA Cell type annotation . . . . .	21
12	Violin plot of expression level of the HNRNPH1 . . . . .	23
13	Heatmap show the Wnt DEGs of Curcumin affected . . . . .	24
14	Violin plot of Wnt DEGs of Curcumin affected . . . . .	25
15	Principal points . . . . .	26
16	Dimension plot of expression level of the Wnt Degr . . . . .	27
17	Pseudotime . . . . .	28
18	APCDD1 pseudotime curve . . . . .	29
19	APCDD1 pseudotime density . . . . .	30
20	PSEU GO enrichment . . . . .	31
21	Pseudotime heatmap of genes . . . . .	32
22	Intersection of FB pseu DEGs with Treat DEGs . . . . .	33
23	Overall communication count . . . . .	35
24	Overall communication weight . . . . .	36
25	Diff path COLLAGEN . . . . .	38
26	PPI Wnt LR . . . . .	40
27	Overall combining Affinity . . . . .	42

28	Docking 969516 into APCDD1 . . . . .	43
29	Docking 969516 into APCDD1 detail . . . . .	44

## List of Tables

1	Data treat vs control DEGs . . . . .	14
2	Genes Wnt Curcumin affected . . . . .	16
3	DEGs of the contrasts . . . . .	21
4	Wnt DEGs of Curcumin affected . . . . .	24
5	Graph test significant results . . . . .	29
6	LR information . . . . .	40

# 1 摘要

## 1.1 需求

1. 客户的 RNA-seq 数据集, 以 DEGs 建立 PPI 网络, 试分析 HNRNPH1 的作用, 以及 wnt 通路。
2. scRNA-seq (可能需要两组数据, 瘢痕增生 (SH) 和正常组织), HNRNPH1 的作用, 免疫细胞的行为, 免疫细胞的 DEGs。
  - 拟时分析, HNRNPH1 的拟时表达变化等
  - 细胞通讯, 巨噬细胞等的通讯, Wnt 通路相关基因的表达和通讯
3. 姜黄素对 HNRNPH1 的作用 (直接作用还是间接, 是否可以结合, 可以尝试分子对接, 或者从转录因子角度出发)
4. 视结果整理, 可做一些新的分析, 或探究一些新的方法。
5. 提供分析源代码

## 1.2 结果

1. 差异分析和 PPI 发现, 姜黄素可对 HNRNPH1 和 Wnt 通路的基因具调控作用, 且 HNRNPH1 和 Wnt 以 TP53 存在直接互作联系。
2. scRNA-seq 数据分析, 未发现 HNRNPH1 的差异表达; 发现了以 APCDD1 为代表的 Wnt 通路基因的表达量变化, 且关联斑痕增生。拟时分析表明, 拟时末期 APCDD1 在 Fibroblast 中表达量显著上升 (Top 2)。而姜黄素的 RNA-seq 数据集中, APCDD1 表达量下调。APCDD1 表现为经典 Wnt 通路抑制作用。以上表明, Curcumin 可通过下调 APCDD1, 激活经典 Wnt 通路, 改善斑痕增生。
3. 分子对接进一步发现了, Curcumin 与 APCDD1 蛋白的优异结合能, 说明 Curcumin 可能通过直接结合 APCDD1 发挥调控作用。
4. 额外的分析已包含在上述各部分中。
5. 本文档提供了与图表一一对应的分析源代码。如果客户需要根据源代码重现分析, 请注意, 所有的源代码和分析均实现于 Linux 系统 (Pop!\_OS 22.04 LTS)。更多系统和 R 配置信息请参考 6.6。此外, 本分析涉及的软件和代码的简要说明见 6.1

详见 4

# 2 前言

## 3 材料和方法

### 3.1 材料

All used GEO expression data and their design:

- **GSE156326:** Single-cell transcriptome of human hypertrophic scars and human skin, and 6 and 8 weeks old mouse scars »> Raw data are unavailable due to patient privacy concerns «<

### 3.2 方法

Mainly used method:

- R package `CellChat` used for cell communication analysis<sup>1</sup>.
- R package `STEINGdb` used for PPI network construction<sup>2,3</sup>.
- R package `ClusterProfiler` used for gene enrichment analysis<sup>4</sup>.
- R package `ClusterProfiler` used for GSEA enrichment<sup>4</sup>.
- GEO <https://www.ncbi.nlm.nih.gov/geo/> used for expression dataset acquisition.
- R package `Limma` and `edgeR` used for differential expression analysis<sup>5,6</sup>.
- R package `Monocle3` used for cell pseudotime analysis<sup>7,8</sup>.
- The R package `Seurat` used for scRNA-seq processing<sup>9,10</sup>.
- The CLI tools of `AutoDock vina` and `ADFR` software used for auto molecular docking<sup>11-15</sup>.
- R package `pathview` used for KEGG pathways visualization<sup>16</sup>.
- The MCC score was calculated referring to algorithm of `CytoHubba`<sup>3</sup>.
- `SCSA` (python) used for cell type annotation<sup>17</sup>.
- R version 4.4.0 (2024-04-24); Other R packages (eg., `dplyr` and `ggplot2`) used for statistic analysis or data visualization.

## 4 分析结果

### 4.1 HNRNPH1、Wnt 与 PPI 网络分析

对姜黄素 (Curcumin) 的 mRNA 数据集进行差异分析, Fig. 1a, Wnt 通路为 Fig. 1c, 所示, Curcumin 可以调控 Wnt 中的多个基因。将 HNRNPH1 与 Wnt 的各个调控基因建立 PPI 网络 (Physical 网络)。发现 HNRNPH1 与 TP53 存在直接作用。而 TP53 与其它 Wnt 蛋白存在互作。

R input

```
fig1 <- cls(  
  cl("./Figure+Table/Treat-vs-control-DEGs.pdf",  
    "./Figure+Table/PPI-HNRNPH1-and-Wnt.pdf"),  
  cl("./Figure+Table/DEG-hsa04310-visualization.png"))  
)  
render(fig1)
```

Figure 1 (下方图) 为图 MAIN Fig 1 概览。

(对应文件为 `./Figure+Table/fig1.pdf`)

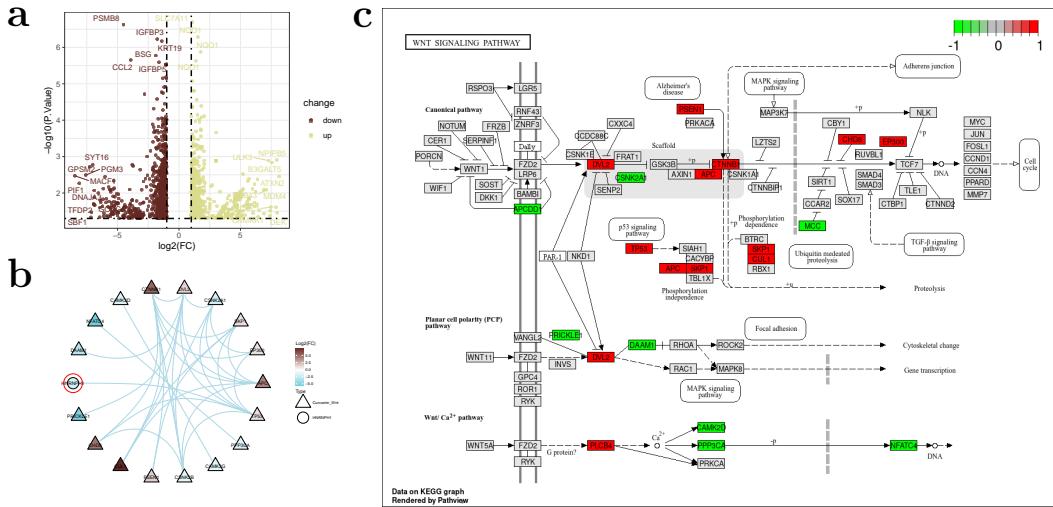


Figure 1: MAIN Fig 1

## 4.2 HNRNPH1、Wnt 与瘢痕的 scRNA-seq 分析

分析 GSE156326 的两组数据 (Scar 和 Skin)，以 Seurat 初步分析，细胞聚类后注释如 Fig. 11a 所示。对 Scar 和 Skin 的各类细胞进行差异分析，见 Tab. 3，探究 HNRNPH1 和 Wnt 通路各个基因的表达，发现 HNRNPH1 为非差异表达基因 (Fig. 2b)。而在 Fibroblast 细胞和 pericyte 细胞中，共有 6 个差异表达基因 (Fig. 2c)。再对这 6 个基因的进一步考察中 (Fibroblast 细胞)，发现 APCDD1 集中表达于特定区域 (Fig. 3a)，而其它基因不具备此特点。

提取 Fibroblast 细胞并重聚类，以 Monocle3 进行拟时分析。这里将 APCDD1 高表达的区域定为拟时终点，随后我们发现 Fibroblast 可以主要分为 2 大 Branch，向拟时终点变化 (Fig. 3b)。绘制 APCDD1 沿拟时轨迹的表达量变化可以发现，APCDD1 在拟时末期时呈上升趋势。进一步探究其来源可以发现，APCDD1 在拟时末期时，主要在 Scar 中表达量增加。而结合 Fig. 1c 可以知道，Curcumin 是可以下调 APCDD1 的表达量。可以推测，Curcumin 对 Fibroblast 细胞 APCDD1 的下调作用，可能改善瘢痕增生。

随后，根据 Fibroblast 拟时图 (Fig. 3b) 进行差异分析 (Graph test)。结果见 Tab. 5。其中，APCDD1 为 Top 2 的差异基因。将 Top 50 的差异基因分 2 个分支 (根据 Fig. 2) 绘制拟时热图 (Fig. 4)，并结合了这些基因在 GO 的富集和是否存在与 Wnt 通路以及 Curcumin 是否对其有调控作用。在这些基因中，APCDD1 和 JUN 基因属于 Wnt 通路，在 Curcumin 的数据中，仅 APCDD1 表现出被 Curcumin 调控表达量变化，为下调趋势。而其余差异基因与 Collagen、ECM 等相关。这些都可能是与瘢痕增生密切关联的通路。

结合 Fig. 3c，将 Fibroblast Pseudotime  $> 10$  的细胞分为 FB:ends,  $< 10$  的细胞分为 FB:begins, 和其他细胞做 CellChat 细胞通讯分析，以发现两部分细胞的差异点 (见 Fig. 5a 和 b，代表通讯数量和权重 (count, weight))。在这些细胞中，COLLAGEN 通路在输入和输出通路都为强度最高的通路 (Fig. 5c, d)。随后，我们对比了 FB:begins 和 FB:ends 和两种免疫细胞 Macrophage 和 Dendritic Cell 的通讯通路差异。这些通路包括：CD99, COLLAGEN, MIF, MK (COLLAGEN 的通讯见 Fig. 5e，其余可见 6.4.5.3)。

由于 APCDD1 为 Fibroblast 显著差异表达基因 (Top 2)，这里推测在 Scar 中上调的 APCDD1 所抑制的 Wnt 通路会对 Fibroblast 与免疫细胞之间的通讯带来调控效果。因此，我们将 CD99, COLLAGEN, MIF，

MK 所涉及的受体配体的基因，与 Wnt 通路的基因 (Curcumin 可调控的) 创建功能关联的 PPI 网络，并且将 Curcumin RNA-seq 数据集中，这些基因的表达量变化映射为  $\text{Log}_2(\text{FC})$ 。随后发现，Wnt 通路基因与上述这些受体配体蛋白存在诸多互作关系，且其中 CD44 可能受姜黄素调控影响。

R input

```
fig2 <- cl(
  rw("./Figure+Table/SCSA-Cell-type-annotation.pdf",
  "./Figure+Table/Violing-plot-of-expression-level-of-the-HNRNPH1.pdf"),
  rw("./Figure+Table/Violing-plot-of-Wnt-DEGs-of-Curcumin-affected.pdf")
)
render(fig2)
```

Figure 2 (下方图) 为图 MAIN Fig 2 概览。

(对应文件为 ./Figure+Table/fig2.pdf)

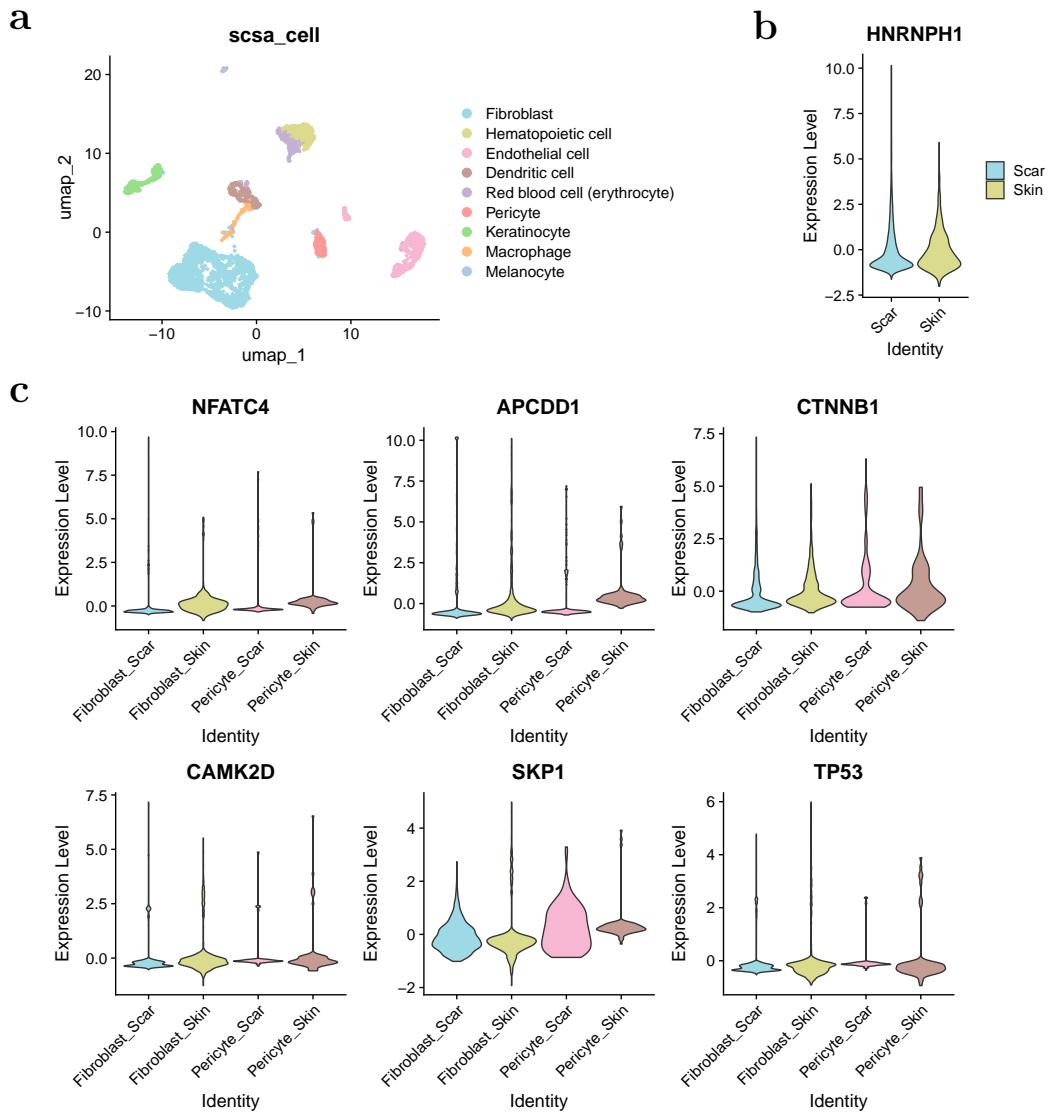


Figure 2: MAIN Fig 2

R input

```
fig3 <- cls(  
  cl("./Figure+Table/Dimension-plot-of-expression-level-of-the-Wnt-Degs.pdf"),  
  cl("./Figure+Table/Pseudotime.pdf",  
    "./Figure+Table/APCDD1-pseudotime-curve.pdf",  
    "./Figure+Table/APCDD1-pseudotime-density.pdf")  
)  
render(fig3)
```

Figure 3 (下方图) 为图 MAIN Fig 3 概览。

(对应文件为 ./Figure+Table/fig3.pdf)

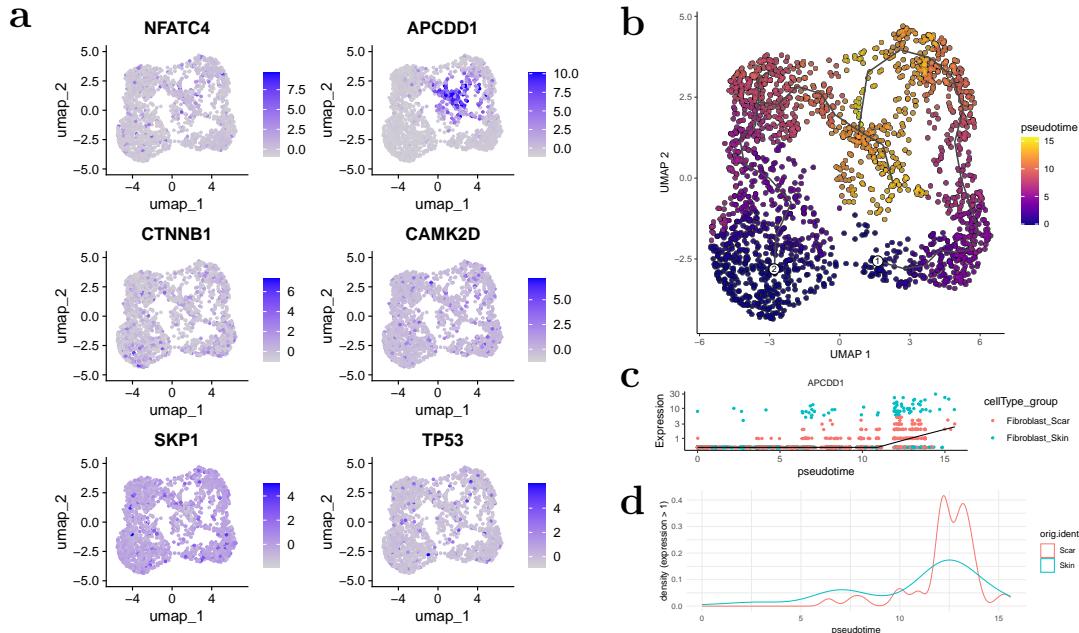


Figure 3: MAIN Fig 3

Figure 4 (下方图) 为图 MAIN Fig 4 概览。

(对应文件为 ./Figure+Table/Pseudotime-heatmap-of-genes.pdf)

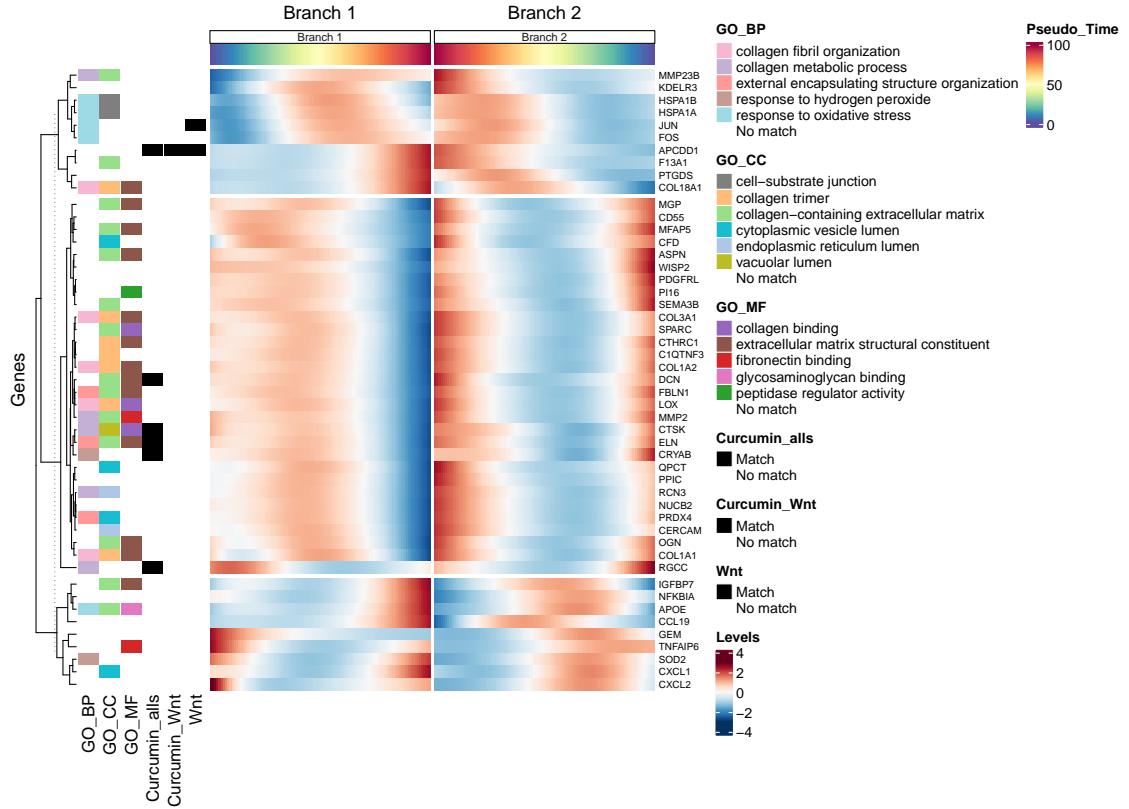


Figure 4: MAIN Fig 4



Figure 5 (下方图) 为图 MAIN Fig 5 概览。

(对应文件为 ./Figure+Table/fig5.pdf)

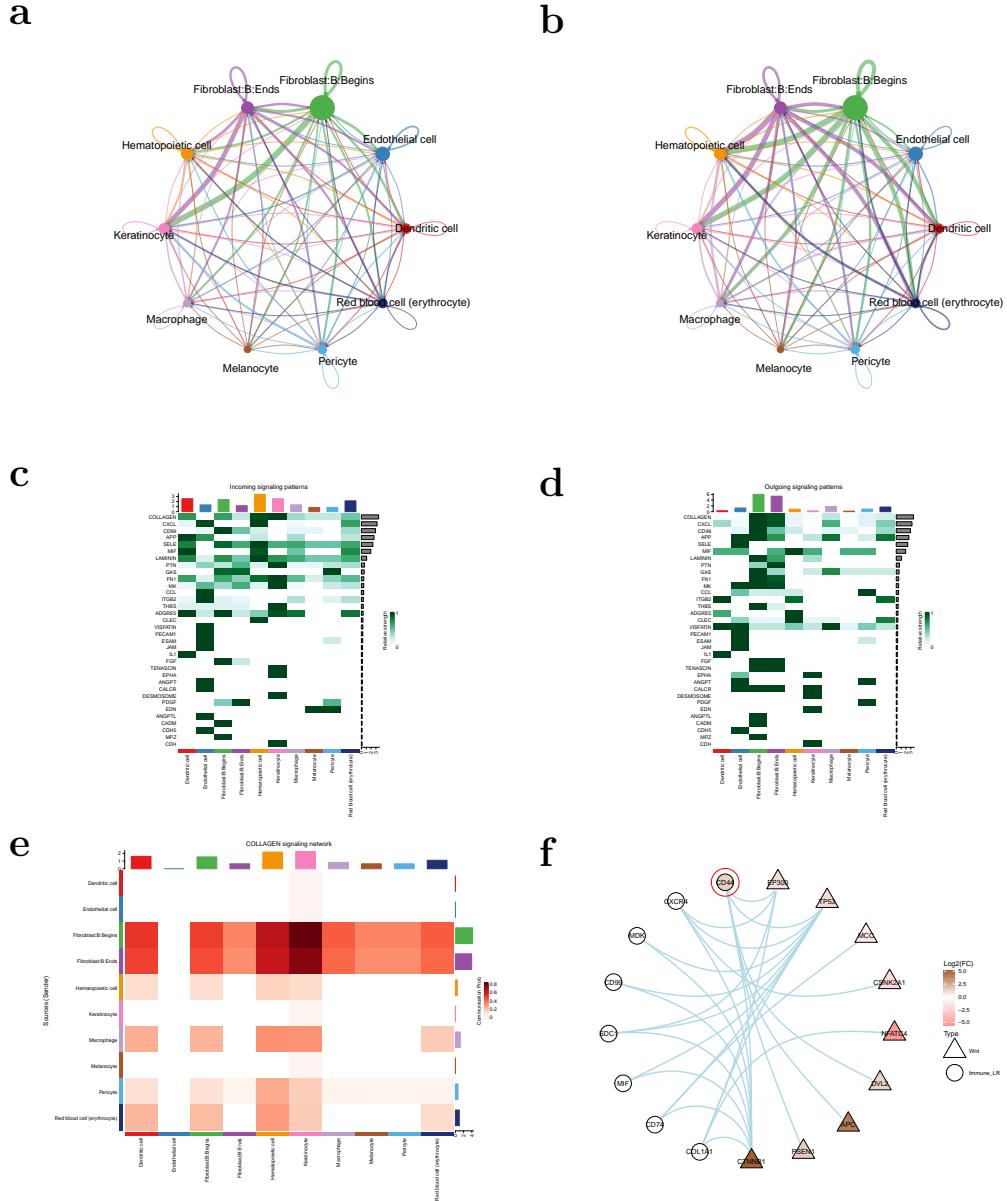


Figure 5: MAIN Fig 5

### 4.3 HNRNPH1、Wnt 与姜黄素

结合上述发现，Curcumin 可能作用 APCDD1 发挥改善瘢痕增生作用。这种作用既可能是直接结合，也可能是间接作用。分子对接可以用以探究药物直接结合蛋白的可能性。这里，我们将 Curcumin 与包括 APCDD1 在内的诸多 Wnt 通路的蛋白，以及 HNRNPH1 蛋白做分子对接。对接亲和能见 Fig. 6a 所示，APCDD1 有

着与 Curcumin 优异的亲和性，且在对接结果中排名最高（对接图见 Fig. 6b 和 c）。因此，Curcumin 将可能通过直接结合 APCDD1 蛋白发挥其表达量调控作用，进而改善斑痕增生。

```
R input
fig6 <- cls(
  # rw(),
  cl("./Figure+Table/Overall-combining-Affinity.pdf"),
  cl("./Figure+Table/Docking-969516-into-APCDD1.png",
  "./Figure+Table/Docking-969516-into-APCDD1-detail.png")
)
render(fig6)
```

Figure 6 (下方图) 为图 MAIN Fig 6 概览。

(对应文件为 ./Figure+Table/fig6.pdf)

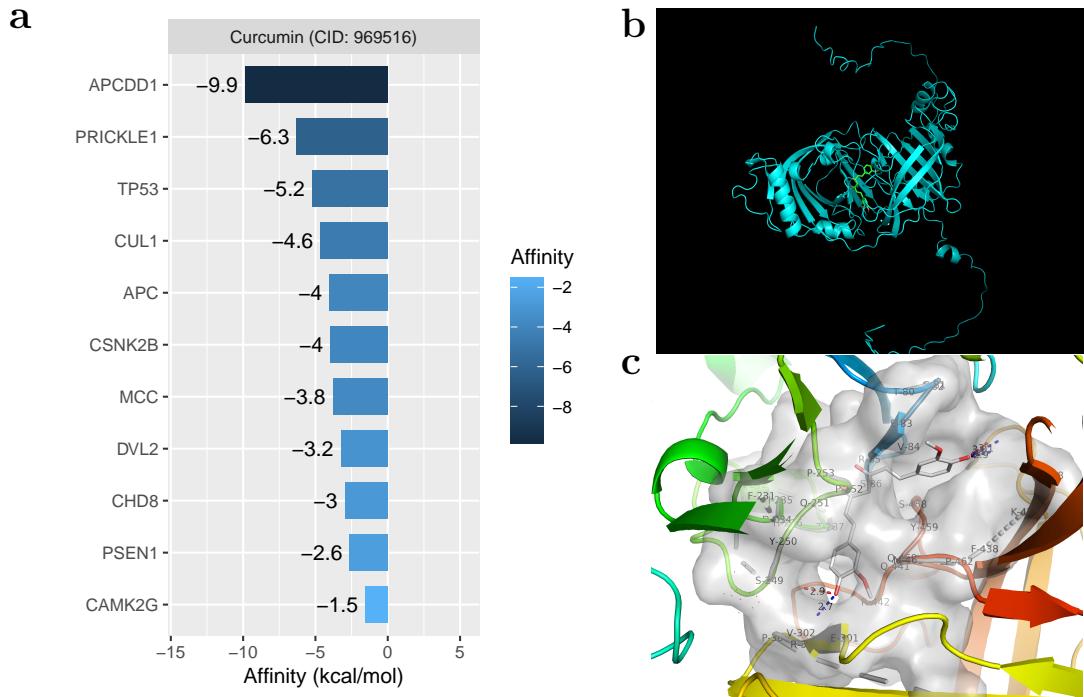


Figure 6: MAIN Fig 6

‘Tiff figures’ 数据已全部提供。

(对应文件为 ./Figure+Table/TIFF)

注：文件夹./Figure+Table/TIFF 共包含 5 个文件。

1. fig1.tiff
2. fig2.tiff
3. fig3.tiff
4. fig5.tiff
5. fig6.tiff

## 5 结论

## 6 附：分析流程

### 6.1 关于源代码的说明

#### 6.1.1 生成该 PDF 文档的源代码

注：请忽略其中包含 `include = F`, 或者函数 `set_cover`, `set_index` 等的 R 代码块。

‘Output RMarkdown’ 数据已提供。

(对应文件为 `output.Rmd`)

‘Output Tex’ 数据已提供。

(对应文件为 `output.tex`)

#### 6.1.2 环境变量配置

大部分程序为 R 代码，但少数 (SCSA 注释、分子对接工具组) 等涉及了其它工具。如果需要使用本文档提供的代码复现这些分析，请确保使用的是 Linux 系统，且以下程序可运行 (即，通过 R 的 `system` 命令可以成功运行它们，例如 `system("mk_prepare_ligand.py")`，你需要做的是安装这些程序，并配置到环境变量，例如 `export` 到 `.bashrc` 中。)：

```
R input
list(
  mk_prepare_ligand.py = "mk_prepare_ligand.py",
  prepare_gpf.py = "prepare_gpf.py",
  autogrid4 = "autogrid4",
  scsa = "python3 ~/SCSA/SCSA.py",
  pymol = "pymol",
  obgen = "obgen"
)
```

### 6.1.3 R 包

本文档使用的代码均为重新封装后的代码，而不是对应软件的原始代码。因此，要复现本文档中的源代码，你需要加载这些代码的函数。所有的函数提供在了 `utils.tool` (如没有明确的理由，请不要修改移动其中的任意文件，否则会出错) 中。

‘R package files’ 数据已全部提供。

(对应文件为 `./utils.tool`)

注：文件夹`./utils.tool` 共包含 6 个文件。

1. DESCRIPTION
2. LICENSE
3. LICENSE.md
4. NAMESPACE
5. R
6. ...

当有需要运行本文档的任意代码时，请先使用以下代码加载该 R 包：(若提示缺少依赖包，请自行安装它们，通过 `BiocManager::install` 或者 `devtools::install_github`)

R input

```
devtools::load_all("./utils.tool")
# 以下是某些文件设定的存储位置, 请修改到需要的位置
options(wd_prefix = "~/outline/lixiao/",
        db_prefix = "~/disk_sdb5/job_internal_data",
        op_prefix = "~/operation/")
```

## 6.2 关于本文档源代码中涉及的文件

‘External files’ 数据已全部提供。

(对应文件为 `./material`)

注：文件夹`./material` 共包含 5 个文件。

1. APCDD1.pdb
2. PLCB4.pdb
3. PRICKLE1
4. PRICKLE1.pdb
5. quant\_hg38\_mrna

## 6.3 HNRNPH1、Wnt 与 PPI 网络分析

### 6.3.1 姜黄素的 mRNA-seq 数据

R input

```
## file.copy("~/outline/lixiao/2023_07_07_eval/quant_hg38_mrna", "./material/", recursive = T)
lst_mrna <- read_kall_quant("../material/quant_hg38_mrna")

lst_mrna$metadata <- dplyr::mutate(lst_mrna$metadata,
  group = ifelse(grpl(sample, "CT"), "control", "treat"))
)
lst_mrna$metadata
lst_mrna$genes

mart <- new_biomart()
lst_mrna$genes <- filter_biomart(mart, generalAttrs(F, T),
  "ensembl_transcript_id", lst_mrna$counts$target_id
)
```

### 6.3.2 DEGs

R input

```
lm <- job_limma(new_dge(lst_mrna$metadata, lst_mrna$counts, lst_mrna$genes))
lm <- step1(lm)
lm <- step2(lm, treat - control, use = "P.Value", use.cut = .05, cut.fc = 1)
Tops <- lm@tables$step2$tops`treat - control`
dplyr::filter(Tops, hgnc_symbol == "HNRNPH1")[, -(1:4)]
Tops
```

Figure 7 (下方图) 为图 Treat vs control DEGs 概览。

(对应文件为 Figure+Table/Treat-vs-control-DEGs.pdf)

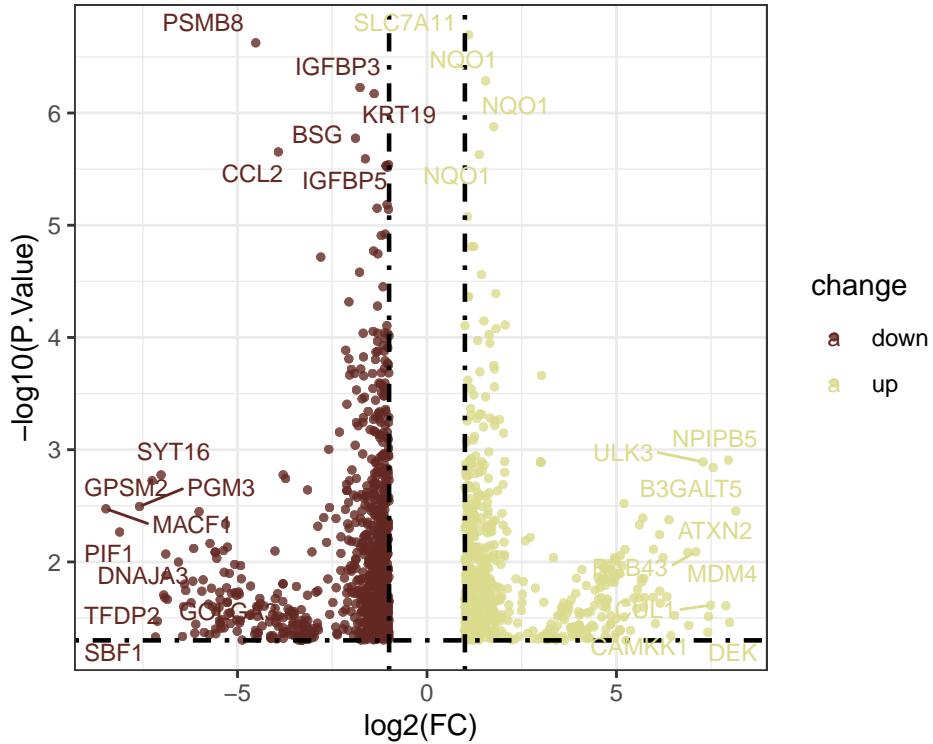


Figure 7: Treat vs control DEGs

P.Value cut-off :

0.05

Log2(FC) cut-off :

1

(上述信息框内容已保存至 Figure+Table/Treat-vs-control-DEGs-content)

Table 1 (下方表格) 为表格 Data treat vs control DEGs 概览。

(对应文件为 Figure+Table/Data-treat-vs-control-DEGs.xlsx)

注：表格共有 1598 行 16 列，以下预览的表格可能省略部分数据；含有 1598 个唯一 ‘rownames’；含有 1300 个唯一 ‘hgnc\_symbol’。

1. hgnc\_symbol: 基因名 (Human)
2. logFC: estimate of the log2-fold-change corresponding to the effect or contrast (for ‘topTableF’ there may be several columns of log-fold-changes)
3. AveExpr: average log2-expression for the probe over all arrays and channels, same as ‘Amean’ in the ‘MarrayLM’ object
4. t: moderated t-statistic (omitted for ‘topTableF’ )
5. P.Value: raw p-value
6. B: log-odds that the gene is differentially expressed (omitted for ‘topTreat’ )

Table 1: Data treat vs control DEGs

rownames	ensembl....2	ensembl....3	entrez...	hgnc_s...	refseq...	chromo...	start_...	end_...	po...	...
36302	ENST00...	ENSG00...	23657	SLC7A11	NM_014331	4	138164097	138242349	...	...
139079	ENST00...	ENSG00...	1728	NQO1	NM_000903	16	69706996	69726668	...	...
109115	ENST00...	ENSG00...	3486	IGFBP3		7	45912245	45921874	...	...
138968	ENST00...	ENSG00...	3880	KRT19	NM_002276	17	41523617	41528308	...	...
111841	ENST00...	ENSG00...	682	BSG	NM_198589	19	571277	583494	...	...
174056	ENST00...	ENSG00...	3488	IGFBP5	NM_000599	2	216672105	216695549	...	...
139083	ENST00...	ENSG00...	1728	NQO1		16	69706996	69726668	...	...
161824	ENST00...	ENSG00...	4176	MCM7	NM_005916	7	100092233	100101940	...	...
85389	ENST00...	ENSG00...	9537	TP53I11	NM_001...	11	44885903	44951306	...	...
6443	ENST00...	ENSG00...	128239	IQGAP3	NM_178229	HG2515...	83962	131161	...	...
158307	ENST00...	ENSG00...	128239	IQGAP3	NM_178229	1	156525405	156572604	...	...
94816	ENST00...	ENSG00...	3838	KPNA2	NM_002266	17	68035636	68047364	...	...
66013	ENST00...	ENSG00...	899	CCNF	NM_001...	16	2429394	2458854	...	...
139080	ENST00...	ENSG00...	1728	NQO1	NM_001...	16	69706996	69726668	...	...
68497	ENST00...	ENSG00...	2512	FTL	NM_000146	19	48965309	48966879	...	...
...	...	...	...	...	...	...	...	...	...	...

### 6.3.3 wnt 信号通路

R input

```

en.deg <- job_enrich(Tops$hgnc_symbol)
en.deg <- step1(en.deg)
en.deg <- step2(en.deg, "hsa04310",
  gene.level = dplyr::select(Tops, hgnc_symbol, logFC)
)
en.deg@plots$step2$p.pathviews$hsa04310

genes.wnt <- dplyr::filter(en.deg@tables$step1$res.kegg$ids, ID == "hsa04310")
genes.wnt <- dplyr::select(genes.wnt, ID, Description, geneName_list)
genes.wnt <- reframe_col(genes.wnt, "geneName_list", function(x) x[[1]])
genes.wnt

```

Figure 8 (下方图) 为图 DEG hsa04310 visualization 概览。

(对应文件为 Figure+Table/DEG-hsa04310-visualization.png)

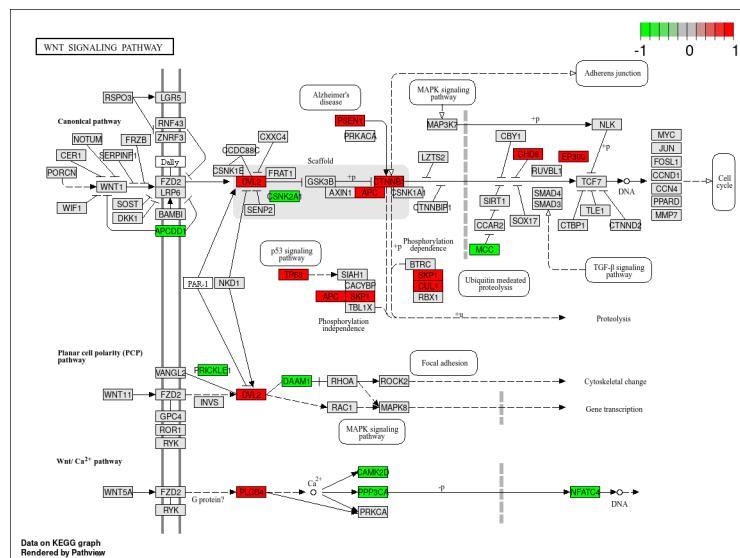


Figure 8: DEG hsa04310 visualization

Interactive figure :

<https://www.genome.jp/pathway/hsa04310>

Enriched genes :

NFATC4, PPP3CA, CAMK2D, PLCB4, DAAM1, DVL2, PRICKLE1, EP300, CHD8, CUL1, SKP1, APC, TP53, CTNNB1, PSEN1, CSNK2A1, APCDD1, MCC

Table 2 (下方表格) 为表格 Genes Wnt Curcumin affected 概览。

(对应文件为 **Figure+Table/Genes-Wnt-Curcumin-affected.csv**)

注：表格共有 20 行 3 列，以下预览的表格可能省略部分数据；含有 1 个唯一 ‘ID’。

Table 2: Genes Wnt Curcumin affected

ID	Description	geneName_list
hsa04310	Wnt signaling pathway	APC
hsa04310	Wnt signaling pathway	APCDD1
hsa04310	Wnt signaling pathway	CAMK2D
hsa04310	Wnt signaling pathway	CAMK2G
hsa04310	Wnt signaling pathway	CHD8
hsa04310	Wnt signaling pathway	CSNK2A1
hsa04310	Wnt signaling pathway	CSNK2B
hsa04310	Wnt signaling pathway	CTNNB1
hsa04310	Wnt signaling pathway	CUL1
hsa04310	Wnt signaling pathway	DAAM1
hsa04310	Wnt signaling pathway	DVL2
hsa04310	Wnt signaling pathway	EP300
hsa04310	Wnt signaling pathway	MCC
hsa04310	Wnt signaling pathway	NFATC4
hsa04310	Wnt signaling pathway	PLCB4
...	...	...

### 6.3.4 构建 PPI 网络

#### 6.3.4.1 DEGs PPI

R input

```
sdb.deg <- job_stringdb(Tops$hgnc_symbol)
sdb.deg <- step1(sdb.deg)
sdb.deg@plots$step1$p.ppi
```

#### 6.3.4.2 HNRNPH1 与 Wnt 通路

注：这里的 PPI 网络为 physical, 即 HNRNPH1 与 Wnt 蛋白之间的直接结合性。

R input

```
# filter the PPI network
lstPPI <- filter(sdb.deg, genes.wnt$geneName_list, "HNRNPH1",
  level.x = dplyr::select(Tops, hgnc_symbol, logFC),
  top = NULL, keep.ref = T, arrow = F, HLS = "HNRNPH1",
  label.shape = c(from = "Curcumin_Wnt", to = "HNRNPH1")
)

lstPPI$p.mcc
```

Figure 9 (下方图) 为图 PPI HNRNPH1 and Wnt 概览。

(对应文件为 Figure+Table/PPI-HNRNPH1-and-Wnt.pdf)

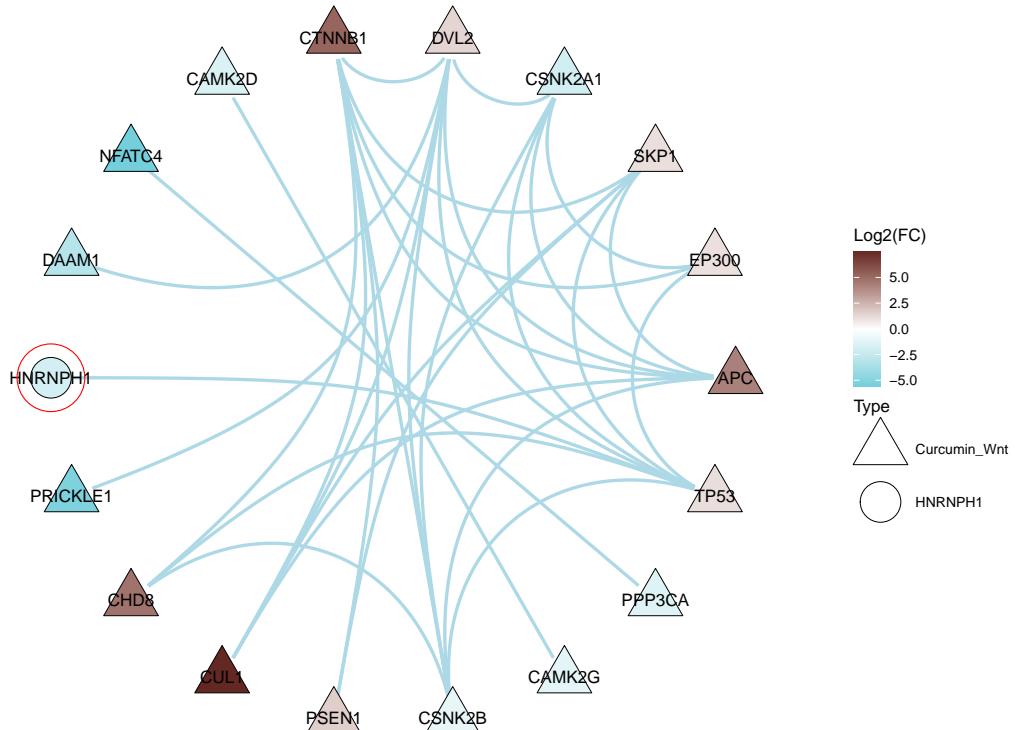


Figure 9: PPI HNRNPH1 and Wnt

## 6.4 HNRNPH1、Wnt 与瘢痕的 scRNA-seq 分析

### 6.4.1 数据来源

R input

```
# Download data from GEO
geo.sc <- job_geo("GSE156326")
geo.sc <- step1(geo.sc)
geo.sc@params$guess
geo.sc <- step2(geo.sc)
untar("./GSE156326/GSE156326_RAW.tar", exdir = "./GSE156326")
prepare_10x("./GSE156326/", "GSM4729097_human_skin_1")
prepare_10x("./GSE156326/", "GSM4729100_human_scar_1")
```

**Data Source ID :**

GSE156326

**data\_processing :**

Raw sequencing data were demultiplexed, aligned to a reference genome (GrCh38/mm10) and counted using Cell Ranger (version 3.0.2; 10x Genomics). Raw sequencing data were demultiplexed, aligned to a reference genome (GrCh38/mm10) and counted using Cell Ranger (version 3.0.2; 10x Genomics).

**data\_processing.1 :**

Genome\_build: GrCh38/mm10 Genome\_build: Genome\_build: GrCh38/mm10

**data\_processing.2 :**

Supplementary\_files\_format\_and\_content: mtx (count matrix in sparse matrix format), barcodes.tsv (barcode ids), features.tsv: (gene ids) Supplementary\_files\_format\_and\_content: Supplementary\_files\_format\_and\_content: mtx (count matrix in sparse matrix format), barcodes.tsv (barcode ids), features.tsv: (gene ids)

(上述信息框内容已保存至 Figure+Table/SC-GSE156326-content)

#### 6.4.2 细胞聚类和鉴定

R input

```
# sr.scar <- job_seurat("./GSE156326/GSM4729100_human_scar_1_barcodes")
# sr.scar <- step1(sr.scar)
# sr.scar@plots$step1$p.qc
# sr.scar <- step2(sr.scar, 0, 5000, 20)
#
# sr.skin <- job_seurat("./GSE156326/GSM4729097_human_skin_1_barcodes")
# sr.skin <- step1(sr.skin)
# sr.skin@plots$step1$p.qc
# rm(sr.skin, sr.scar)

sr <- job_seuratn(c("./GSE156326/GSM4729100_human_scar_1_barcodes",
                     "./GSE156326/GSM4729097_human_skin_1_barcodes"),
                     c("Scar", "Skin"))

sr <- step1(sr, 0, 5000, 20)
sr <- step2(sr)
sr@plots$step2$p.pca_rank
sr <- step3(sr, 1:15, 1.2)
sr@plots$step3$p.umap
sr <- step4(sr, "")
sr <- step5(sr)
# SCSA for cell type annotation
sr <- step6(sr, "Skin")
sr@plots$step6$p.map_scsa
```

Figure 10 (下方图) 为图 UMAP Clustering 概覽。

(对应文件为 Figure+Table/UMAP-Clustering.pdf)

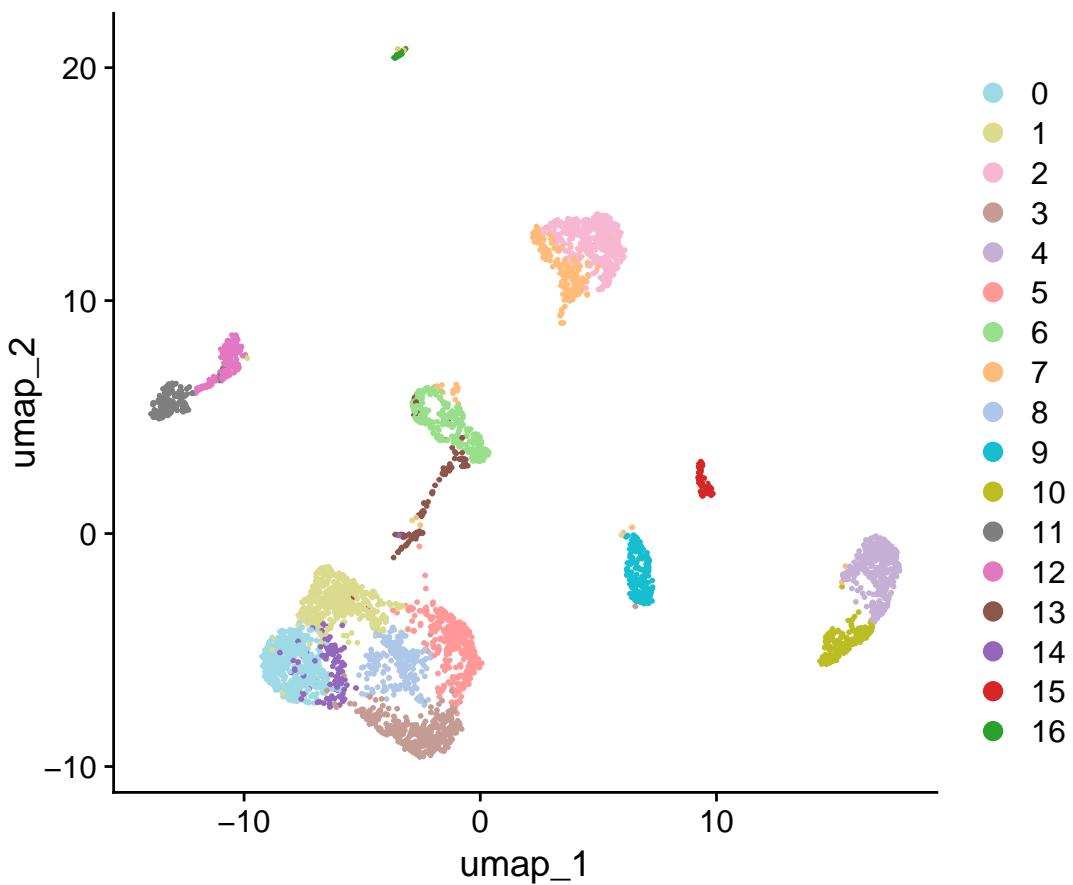


Figure 10: UMAP Clustering

Figure 11 (下方图) 为图 SCSA Cell type annotation 概览。

(对应文件为 Figure+Table/SCSA-Cell-type-annotation.pdf)

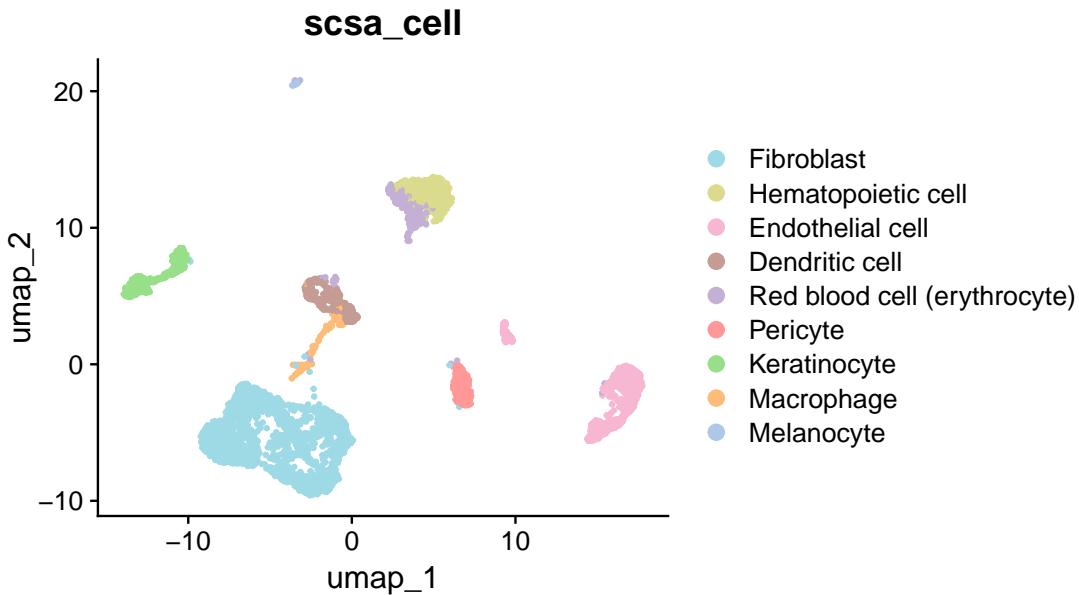


Figure 11: SCSA Cell type annotation

#### 6.4.3 差异分析

R input

```
sr <- mutate(sr, group_cellType = paste0(orig.ident, "_", make.names(scsa_cell)))
contrasts.sr <- lapply(make.names(ids(sr)), function(x) paste0(c("Scar", "Skin"), "_", x))
contrasts.sr

sr <- diff(sr, "group_cellType", contrasts.sr, name = "HN_group")
sr@params$HN_group
```

Table 3 (下方表格) 为表格 DEGs of the contrasts 概览。

(对应文件为 [Figure+Table/DEGs-of-the-contrasts.csv](#))

注：表格共有 4863 行 7 列，以下预览的表格可能省略部分数据；含有 9 个唯一 ‘contrast’。

Table 3: DEGs of the contrasts

contrast	p_val	avg_log2FC	pct.1	pct.2	p_val_adj	gene
Scar_Hemat...	3.26093311...	-0.2786380...	0.002	1	0.00097827...	ST6GALNAC6
Scar_Hemat...	4.83497164...	2.72042542...	0.01	0.222	0.00145049...	CCL19
Scar_Hemat...	4.97701805...	2.36267982...	0.01	0.889	0.00149310...	PPP1R12B
Scar_Hemat...	6.54317478...	-0.7858693...	0.015	0.111	0.00196295...	HLA-DMB
Scar_Hemat...	6.54317478...	7.81465650...	0.015	0.778	0.00196295...	SNX10

contrast	p_val	avg_log2FC	pct.1	pct.2	p_val_adj	gene
Scar_Hemat...	1.01658537...	3.50570583...	0.012	0.333	0.00304975...	HHEX
Scar_Hemat...	1.04560332...	-0.3003328...	0.024	1	0.00313680...	BCL9L
Scar_Hemat...	1.04564108...	1.40598233...	0.024	1	0.00313692...	CHSY1
Scar_Hemat...	1.10616882...	8.49062033...	0.022	0.667	0.00331850...	NTRK2
Scar_Hemat...	1.32705918...	2.57196817...	0.012	0.778	0.00398117...	TRIM25
Scar_Hemat...	1.34570508...	2.90265755...	0.022	0.556	0.00403711...	SLC40A1
Scar_Hemat...	1.56816847...	8.08654102...	0.032	0.889	0.00470450...	ITPR1
Scar_Hemat...	1.63472658...	4.54715445...	0.027	0.222	0.00490417...	TSPAN5
Scar_Hemat...	1.72770803...	8.32439693...	0.039	0.889	0.00518312...	CD70
Scar_Hemat...	1.72770803...	5.75361787...	0.034	1	0.00518312...	MTSS1
...	...	...	...	...	...	...

#### 6.4.3.1 HNRNPH1 的表达

HNRNPH1 在这批单细胞数据中，为非差异表达基因。

R input

```

p.mapHn_cell <- focus(sr, "HNRNPH1")
p.mapHn_group <- focus(sr, "HNRNPH1", "orig.ident")
wrap(p.mapHn_group$p.vln, 3, 4)

# No results
dplyr::filter(sr@params$HN_group, gene == "HNRNPH1")

```

Figure 12 (下方图) 为图 Violing plot of expression level of the HNRNPH1 概览。

(对应文件为 Figure+Table/Violing-plot-of-expression-level-of-the-HNRNPH1.pdf)

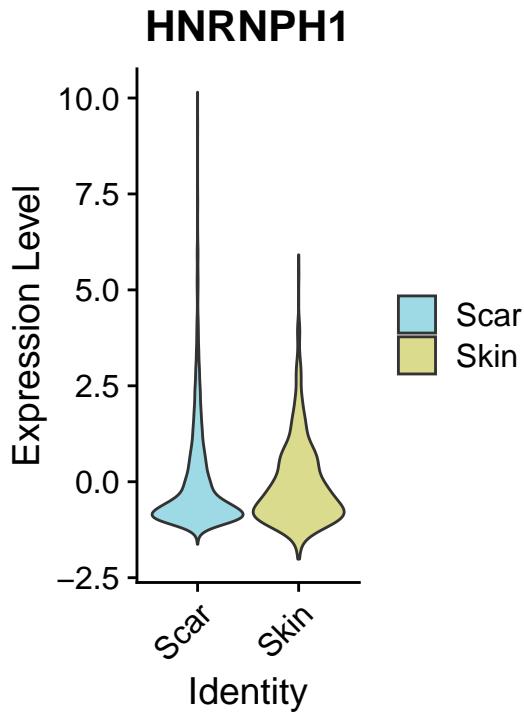


Figure 12: Violin plot of expression level of the HNRNPH1

#### 6.4.3.2 Wnt 通路基因的表达

- scRNA-seq, Scar vs Skin (Fibroblast, Pericyte), TP53 ↓, APCDD1 ↑
- RNA-seq, 姜黄素给药, TP53 ↑, APCDD1 ↓

```
R input
scDegs.wnt <- dplyr::filter(sr@params$HN_group, gene %in% genes.wnt$geneName_list)
scDegs.wnt

scCell.degWnt <- which(ids(sr, "scsa_cell", F) %in% c("Fibroblast", "Pericyte"))
scCell.degWnt

sr <- mutate(sr, cellType_group = gs(group_cellType, "^(\\w+)_\\1", "\\2_\\1"))
p.hpWnt <- map(sr, scDegs.wnt$gene, group_by = "cellType_group", cells = scCell.degWnt)
p.hpWnt

p.focScDegWnt <- focus(getsub(sr, cells = scCell.degWnt),
  scDegs.wnt$gene, group_by = "cellType_group"
)
p.focScDegWnt$p.vln
```

Table 4 (下方表格) 为表格 Wnt DEGs of Curcumin affected 概览。

(对应文件为 Figure+Table/Wnt-DEGs-of-Curcumin-affected.csv)

注：表格共有 8 行 7 列，以下预览的表格可能省略部分数据；含有 2 个唯一 ‘contrast’。

Table 4: Wnt DEGs of Curcumin affected

contrast	p_val	avg_log2FC	pct.1	pct.2	p_val_adj	gene
Scar_Fibro...	8.27560595...	3.85702595...	0.157	0.584	2.48268178...	NFATC4
Scar_Fibro...	3.73579114...	1.64972906...	0.245	0.317	1.12073734...	APCDD1
Scar_Fibro...	2.83791959...	1.52983006...	0.31	0.362	8.51375879...	CTNNB1
Scar_Fibro...	2.54894125...	0.60325628...	0.112	0.288	7.64682377...	CAMK2D
Scar_Fibro...	4.69293478...	-2.8547127...	0.382	0.133	1.40788043...	SKP1
Scar_Pericy...	1.03782691...	2.97367652...	0.112	0.863	3.11348075...	NFATC4
Scar_Pericy...	5.43911467...	2.00598608...	0.216	0.925	1.63173440...	APCDD1
Scar_Pericy...	5.30098924...	-2.3467874...	0.052	0.125	0.00015902...	TP53

Figure 13 (下方图) 为图 Heatmap show the Wnt DEGs of Curcumin affected 概览。

(对应文件为 Figure+Table/Heatmap-show-the-Wnt-DEGs-of-Curcumin-affected.pdf)

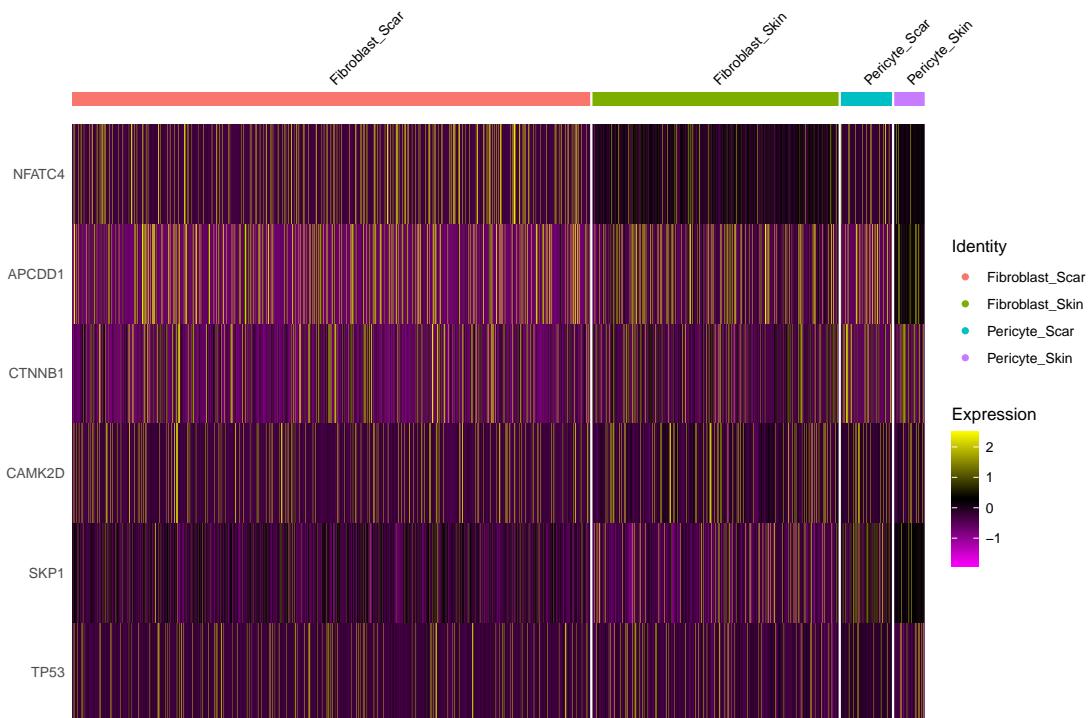


Figure 13: Heatmap show the Wnt DEGs of Curcumin affected

Figure 14 (下方图) 为图 Violing plot of Wnt DEGs of Curcumin affected 概览。

(对应文件为 Figure+Table/Violing-plot-of-Wnt-DEGs-of-Curcumin-affected.pdf)

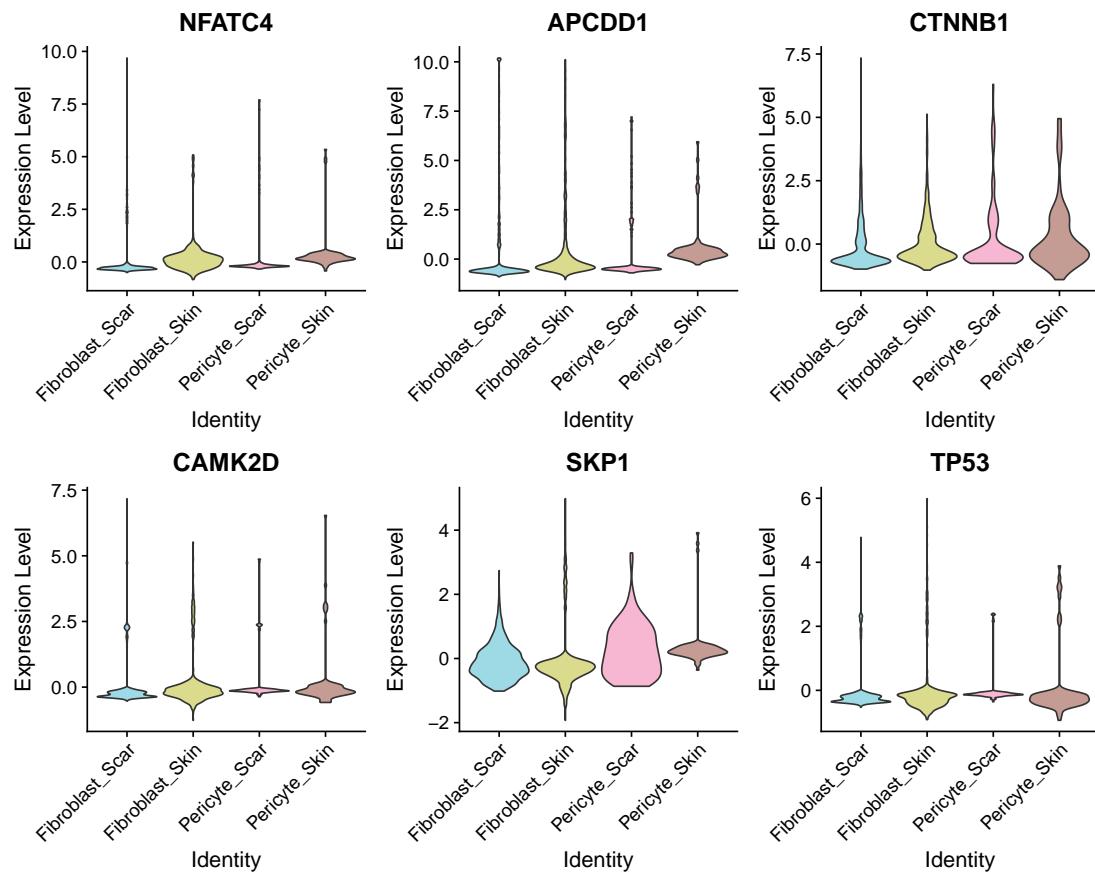


Figure 14: Violin plot of Wnt DEGs of Curcumin affected

#### 6.4.4 拟时分析

##### 6.4.4.1 拟时终点与 APCDD1

这里发现 Fig. 16 APCDD1 集中表达于一个区域，因此这里尝试将该区域选定为拟时终点。

```
R input
mn <- do_monocle(sr, "Fibroblast")
mn <- step1(mn, "cellType_group", pre = T)
wrap(mn@plots$step1$p.prin, 5, 4)

p.srSub_wnt <- focus(mn@params$sr_sub, scDegs.wnt$gene)
p.srSub_wnt$p.dim
```

Figure 15 (下方图) 为图 Principal points 概览。

(对应文件为 Figure+Table/Principal-points.pdf)

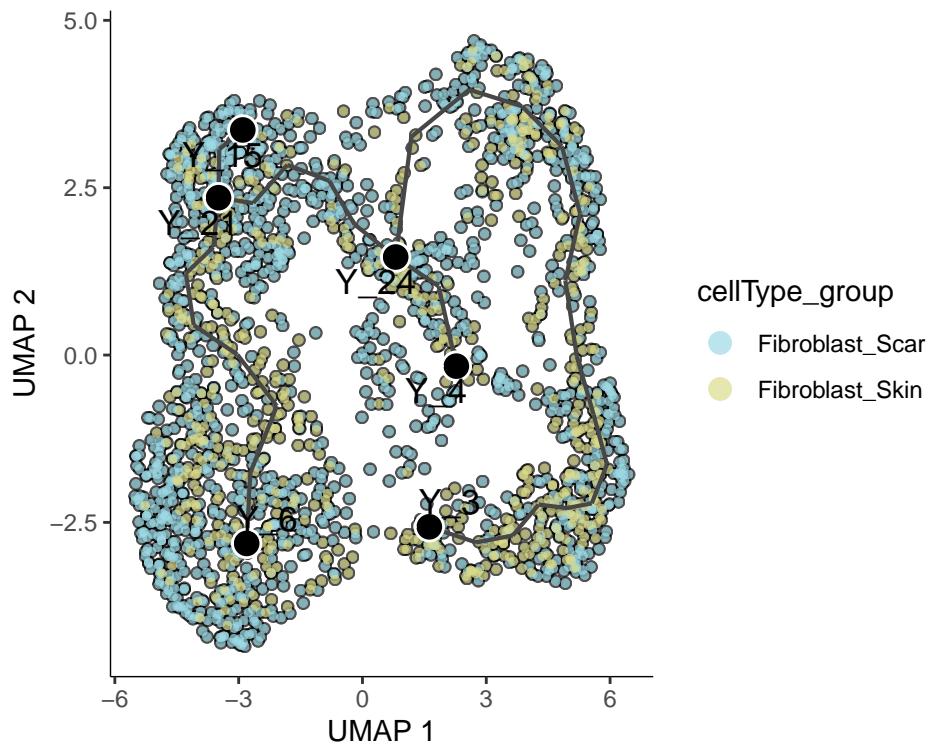


Figure 15: Principal points

Figure 16 (下方图) 为图 Dimension plot of expression level of the Wnt Degt 概览。

(对应文件为 [Figure+Table/Dimension-plot-of-expression-level-of-the-Wnt-Degs.pdf](#))

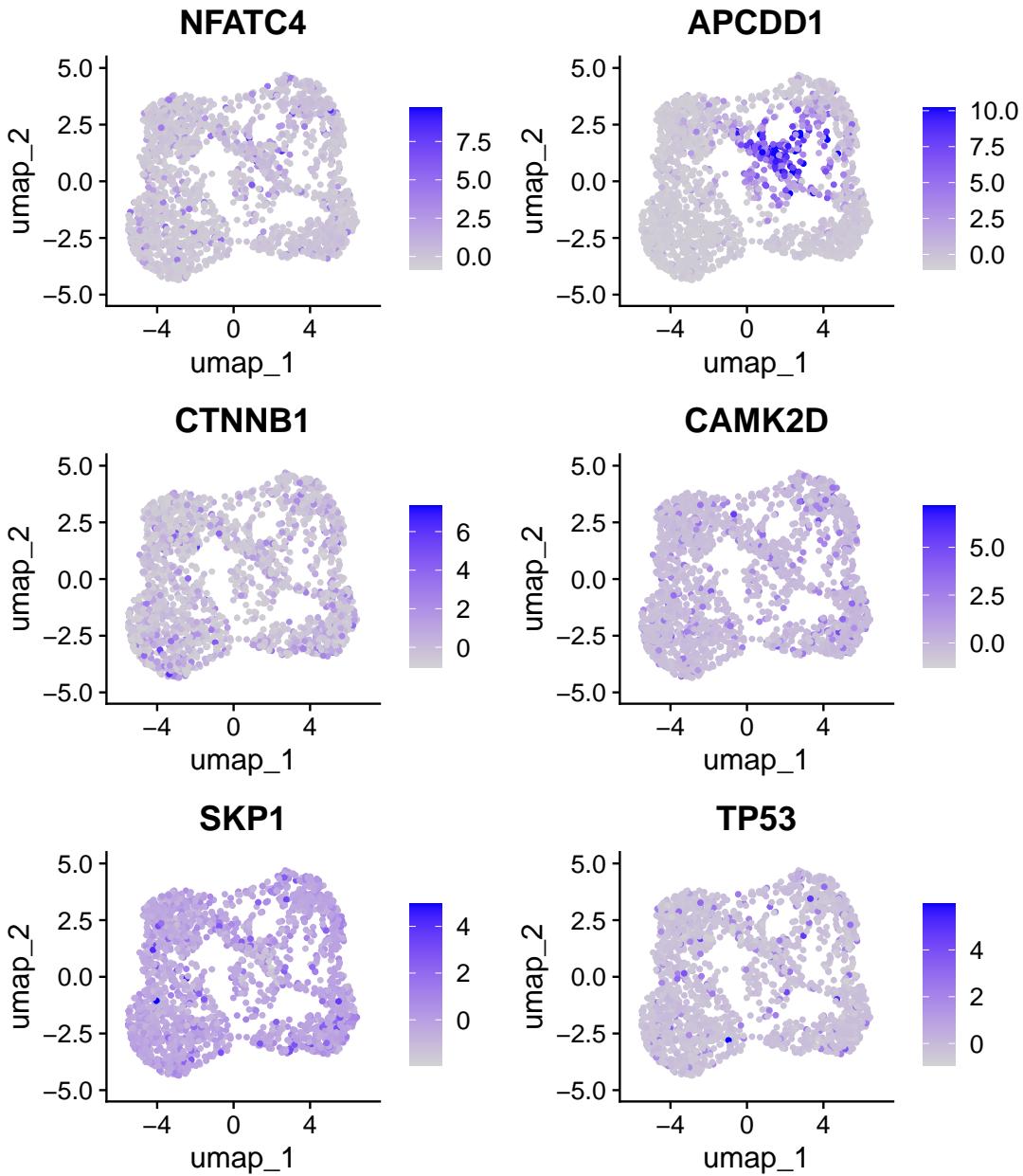


Figure 16: Dimension plot of expression level of the Wnt Degs

#### 6.4.4.2 APCDD1 主要在 Scar 中高表达

随后发现，APCDD1 的确在拟时末期高表达，而且是主要在 Scar 组织中高表达，见 Fig. 19

R input

```
mn <- step2(mn, c("Y_3", "Y_6"))
mn@plots$step2$p.pseu

mn <- step3(mn, group.by = "seurat_clusters")
mn <- step4(mn, ids(mn), "APCDD1", "cellType_group")

mn@tables$step3$graph_test.sig
mn@plots$step4$genes_in_pseudotime$pseudo1
mn@plots$step4$plot_density$pseudo1
```

Figure 17 (下方图) 为图 Pseudotime 概览。

(对应文件为 Figure+Table/Pseudotime.pdf)

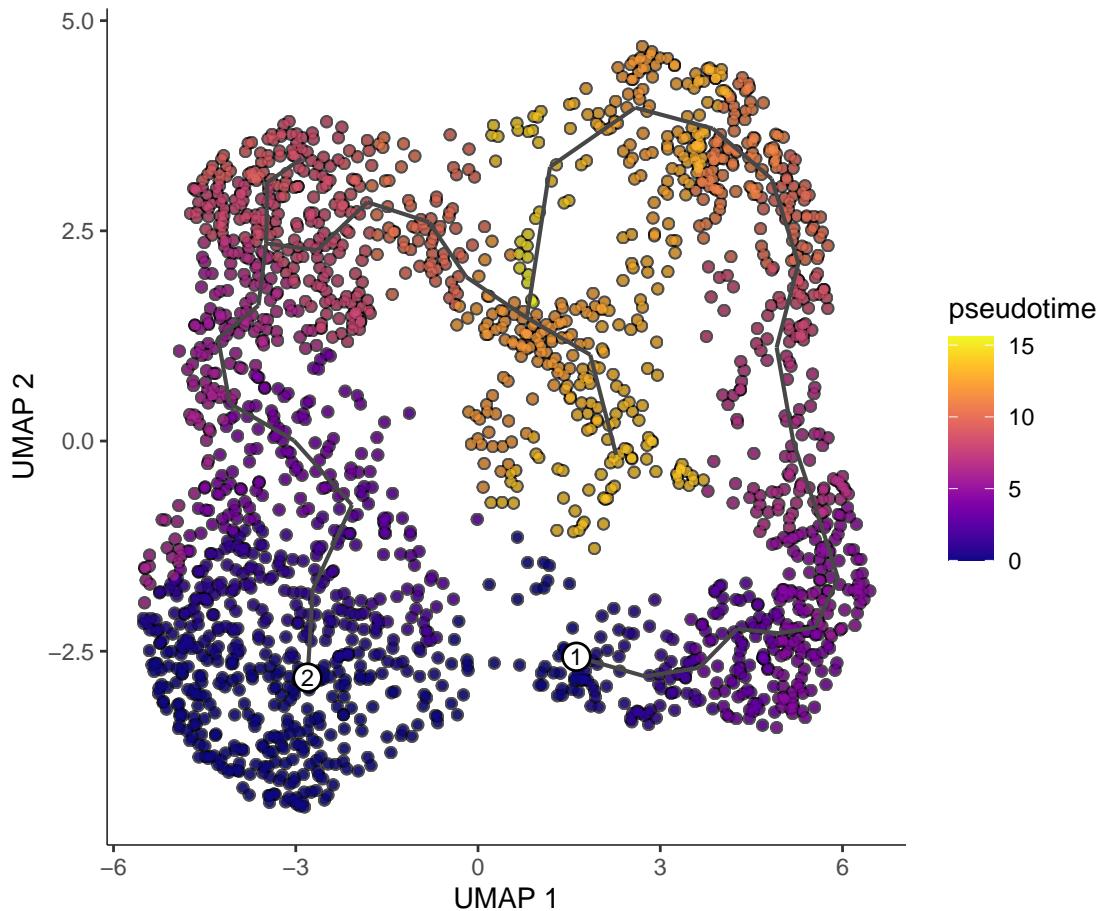


Figure 17: Pseudotime

Table 5 (下方表格) 为表格 graph test significant results 概览。

(对应文件为 Figure+Table/graph-test-significant-results.csv)

注：表格共有 5090 行 6 列，以下预览的表格可能省略部分数据；含有 5090 个唯一 ‘gene\_id’。

1. gene\_id: GENCODE/Ensembl gene ID

Table 5: Graph test significant results

gene_id	status	p_value	morans_tes...	morans_I	q_value
CTHRC1	OK	0	82.0068307...	0.49377906...	0
APCDD1	OK	0	81.8052541...	0.49183724...	0
PI16	OK	0	79.1167758...	0.47634067...	0
IGFBP7	OK	0	78.4677201...	0.47238376...	0
FOS	OK	0	78.0245837...	0.46974828...	0
WISP2	OK	0	77.1253098...	0.46433708...	0
PDGFRL	OK	0	74.1245797...	0.44626545...	0
C1QTNF3	OK	0	72.9964250...	0.43937764...	0
FBLN1	OK	0	69.0651449...	0.41554895...	0
ELN	OK	0	68.4719775...	0.41216524...	0
MFAP5	OK	0	68.3194338...	0.41124414...	0
MMP2	OK	0	68.3131206...	0.41117358...	0
SEMA3B	OK	0	64.3202720...	0.38704468...	0
SOD2	OK	0	64.0519950...	0.38551658...	0
APOE	OK	0	64.0171412...	0.38530148...	0
...	...	...	...	...	...

Figure 18 (下方图) 为图 APCDD1 pseudotime curve 概览。

(对应文件为 [Figure+Table/APCDD1-pseudotime-curve.pdf](#))

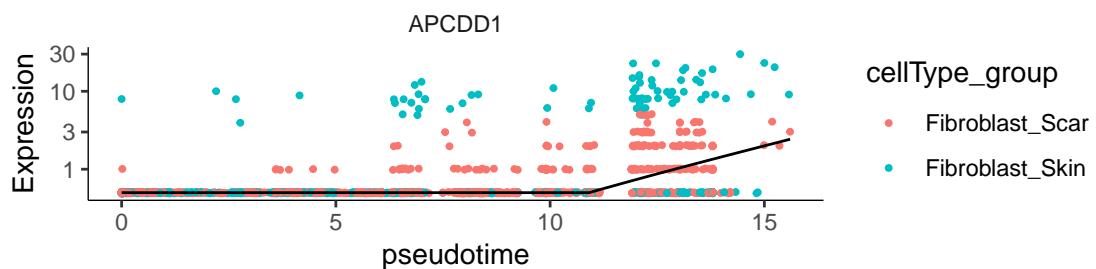


Figure 18: APCDD1 pseudotime curve

Figure 19 (下方图) 为图 APCDD1 pseudotime density 概览。

(对应文件为 [Figure+Table/APCDD1-pseudotime-density.pdf](#))

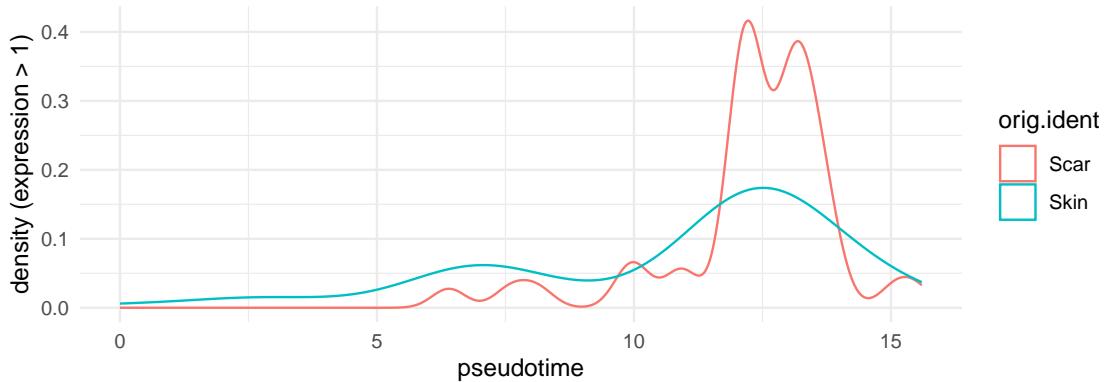


Figure 19: APCDD1 pseudotime density

#### 6.4.4.3 Fibroblast 拟时轨迹下的差异基因

- GO 富集表明，差异基因主要富集于和 Collagen 相关的通路。
- APCDD1 为排名第 2 的差异基因。
- 在两个主要的拟时分支中，APCDD1 均呈表达量上升趋势。
- APCDD1 是 Top 50 的差异基因中，唯一和 Wnt 相关且姜黄素对其有调控作用的基因见 Fig. 21。

R input

```
scDegs.pseu <- head(dplyr::filter(mn@tables$step3$graph_test.sig, q_value < .000001), 500)

en.pseu <- job_enrich(scDegs.pseu$gene_id)
en.pseu <- step1(en.pseu)
en.pseu@plots$step1$p.go
```

Figure 20 (下方图) 为图 PSEU GO enrichment 概览。

(对应文件为 Figure+Table/PSEU-GO-enrichment.pdf)

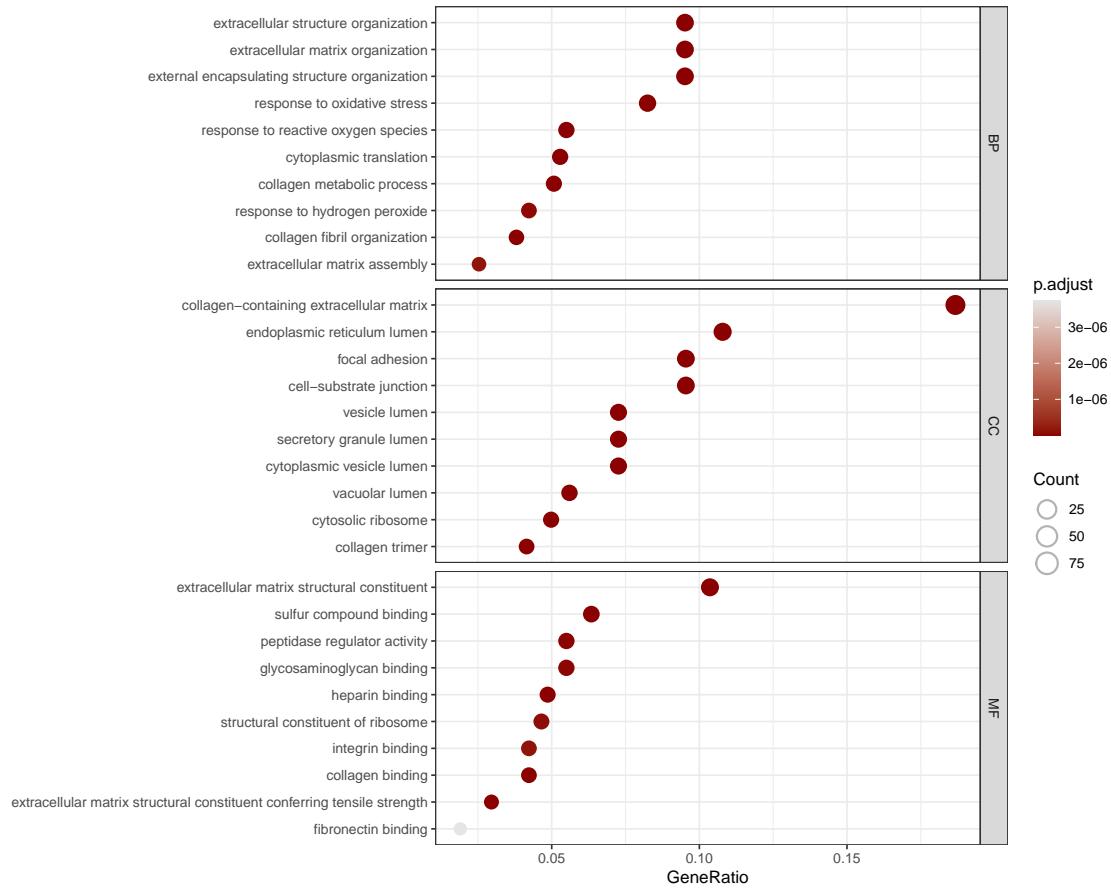


Figure 20: PSEU GO enrichment

R input

```

genes.allWnt <- get_genes.keggPath("hsa04310")

p.hpPseu <- map(mn, head(scDegs.pseu$gene_id, 50), enrich = en.pseu,
branches = list(c("Y_6", "Y_24"), c("Y_3", "Y_24")),
HLS = list(Wnt = genes.allWnt, Curcumin_Wnt = genes.wnt$geneName_list,
Curcumin_all = Tops$hgnnc_symbol)
)

p.hpPseu

```

Figure 21 (下方图) 为图 Pseudotime heatmap of genes 概览。

(对应文件为 Figure+Table/Pseudotime-heatmap-of-genes.pdf)

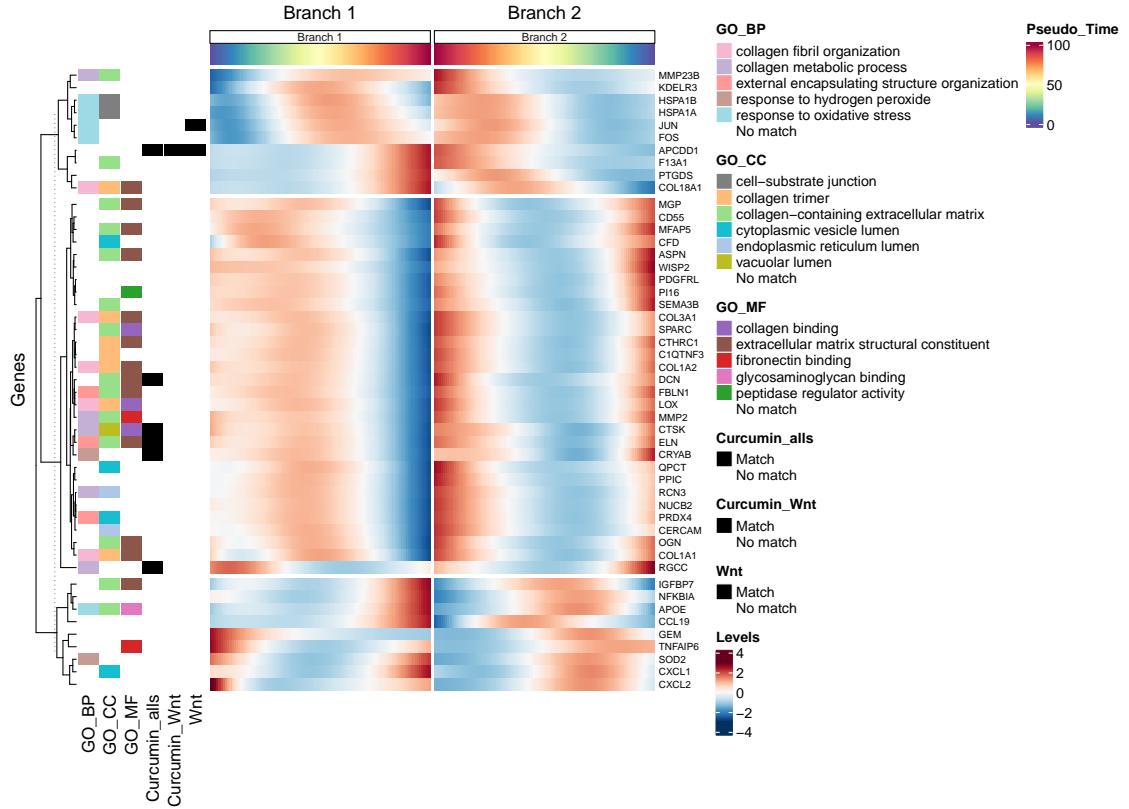


Figure 21: Pseudotime heatmap of genes

#### 6.4.4.4 姜黄素有调控作用的靶点 (所有的差异基因中)

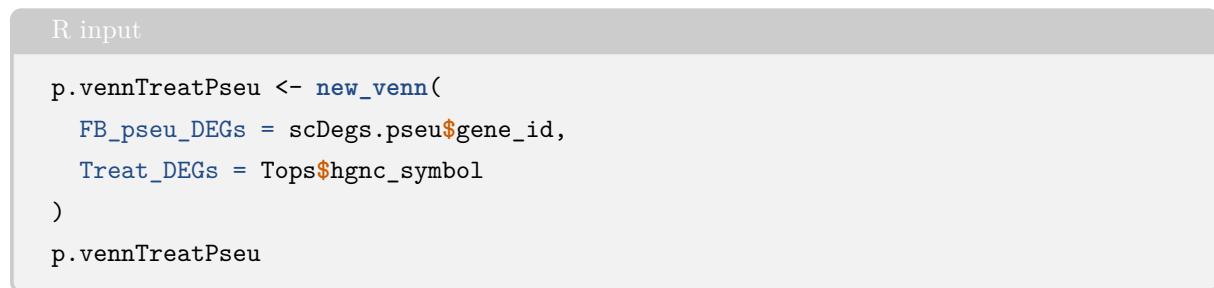


Figure 22 (下方图) 为图 Intersection of FB pseu DEGs with Treat DEGs 概览。

(对应文件为 Figure+Table/Intersection-of-FB-pseu-DEGs-with-Treat-DEGs.pdf)

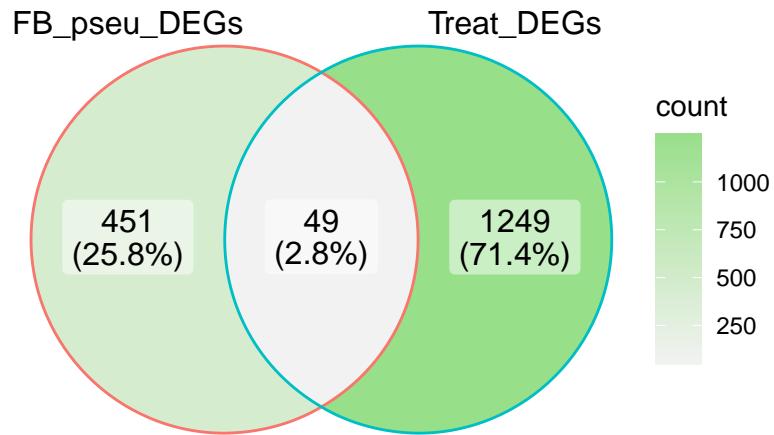


Figure 22: Intersection of FB pseu DEGs with Treat DEGs

**Intersection :**

APCDD1, ELN, CRYAB, DCN, CTSK, RGCC, CCL2, IGFBP5, CRABP2, AEBP1, PPIB, PLD3, POSTN, CTSC, ZFP36, BSG, BIRC3, TMEM258, FTL, ID3, FN1, NFE2L2, FKBP7, TNFAIP3, PYCR1, TIMP1, S100A10, CES1, OLFM2, KPNA2, FTH1, TMED10, DDX5, IL6, COL13A1, DPP7, RPL28, HTRA3, PDIA3, TNXB, C1QTNF2, CTSB, ASAHI, GLT8D1,...

(上述信息框内容已保存至 [Figure+Table/Intersection-of-FB-pseu-DEGs-with-Treat-DEGs-content](#))

## 6.4.5 细胞通讯

### 6.4.5.1 总体通讯

因为 Fig. 21 所示，末期的 APCDD1 表达量升高，这里尝试将 Fibroblast 细胞分为 Begins 和 Ends 两组，作为两种亚型，和其它细胞做细胞通讯分析。

### R input

```
mn <- add_anno(mn, branches = list(c("Y_6", "Y_24"), c("Y_3", "Y_24")))
sr <- map(sr, mn)
sr <- mutate(sr,
  branch_time = paste0("B:", ifelse(pseudotime > 10, "Ends", "Begins")),
  cellType_sub = as.character(scsa_cell),
  cellType_sub = ifelse(is.na(pseudotime), cellType_sub, paste0(cellType_sub, ":", branch_time)),
  cellType_sub = as.factor(cellType_sub)
)
sr@object@meta.data$cellType_sub %>% table

cc <- asjob_cellchat(sr, "cellType_sub")
cc <- step1(cc)
cc@plots$step1$p.aggre_count
```

Figure 23 (下方图) 为图 Overall communication count 概览。

(对应文件为 Figure+Table/Overall-communication-count.pdf)

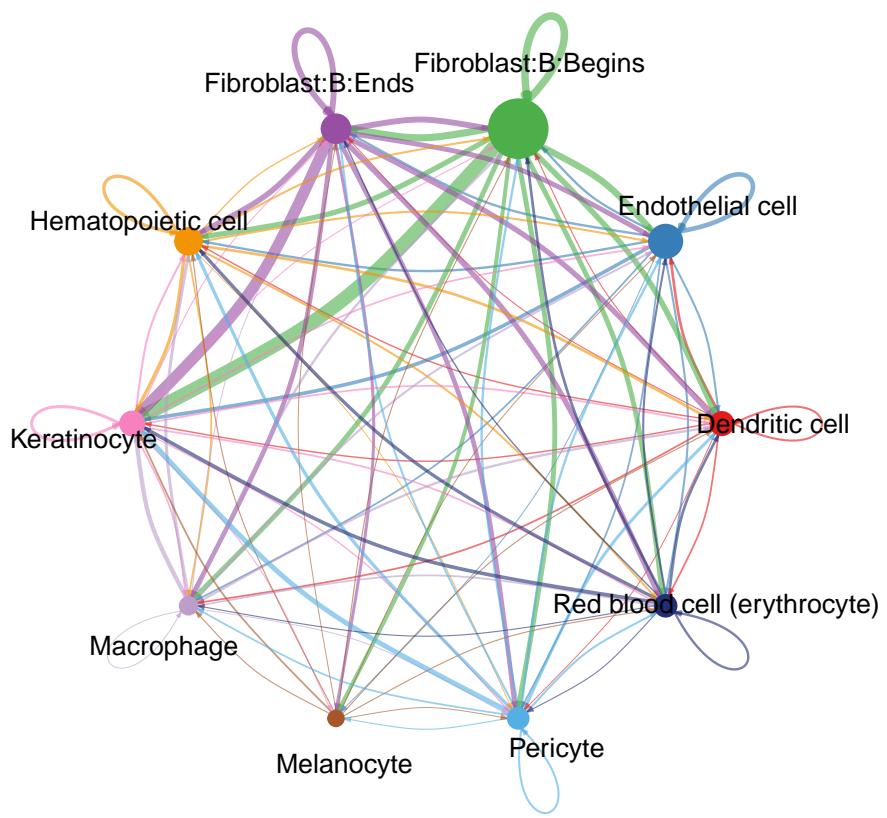


Figure 23: Overall communication count

Figure 24 (下方图) 为图 Overall communication weight 概览。

(对应文件为 [Figure+Table/Overall-communication-weight.pdf](#))

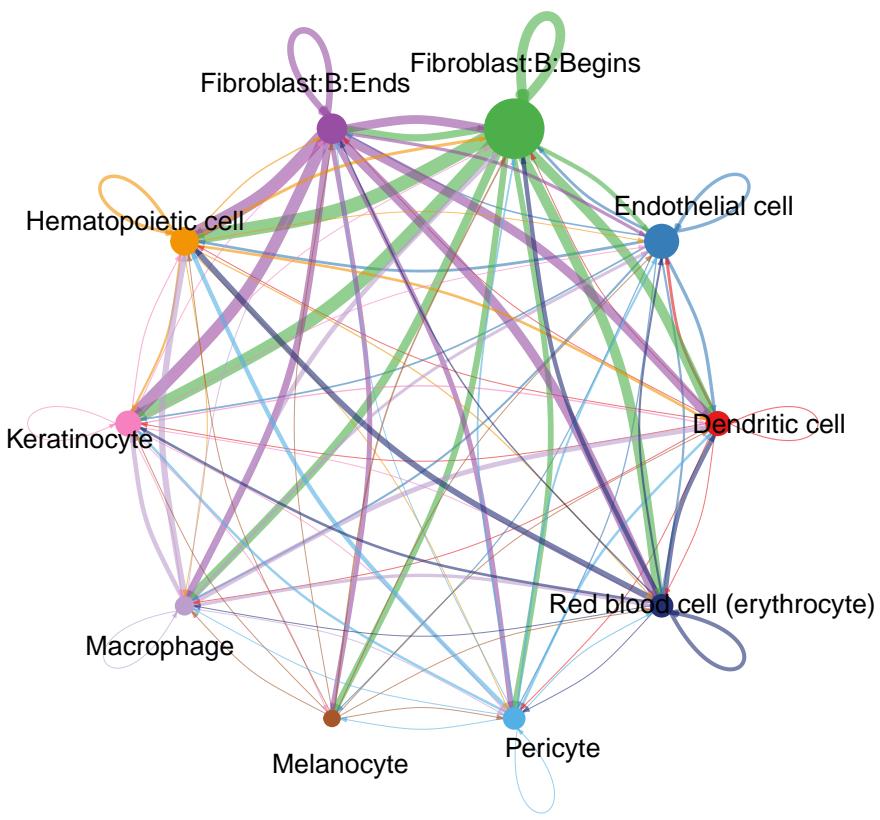


Figure 24: Overall communication weight

#### 6.4.5.2 Fibroblast 分支与免疫细胞

这里比较了 FB:ends 和 FB:begins 与两种免疫细胞 Microphage、dendritic cells 的通讯 (pathway) 的不同之处。

#### R input

```
fun_diff <- function(data, use) {  
  pair <- c("source", "target")  
  pair <- pair[ pair != use ]  
  lapply(split(data, data[[ use ]]),  
    function(x) {  
      fun <- function(pat) {  
        unique(dplyr::filter(x, grpl(!rlang::sym(pair), !pat))$pathway_name)  
      }  
      ends <- fun("Ends")  
      begins <- fun("Begin")  
      unique(c(setdiff(ends, begins), setdiff(begins, ends)))  
    })  
}  
  
chat.alltarget <- select_pathway(cc, "Begins|Ends", "^[^:]+$", "path")  
diff.tar <- fun_diff(chat.alltarget, "target")  
  
chat.allsource <- select_pathway(cc, "^[^:]+$", "Begins|Ends", "path")  
diff.sour <- fun_diff(chat.allsource, "source")  
  
diff.imm <- unique(unlist(lapply(list(diff.sour, diff.tar),  
  function(x) x[ grpl(names(x), "Dendri|Macro") ])))  
diff.imm  
# [1] "CD99"      "COLLAGEN"   "MIF"       "MK"
```

#### 6.4.5.3 差异通讯

#### R input

```
cc <- step2(cc, diff.imm)  
cc@plots$step2$cell_comm_heatmap$COLLAGEN  
cc@plots$step2$cell_comm_heatmap$ALL
```

Figure 25 (下方图) 为图 Diff path COLLAGEN 概览。

(对应文件为 Figure+Table/Diff-path-COLLAGEN.pdf)

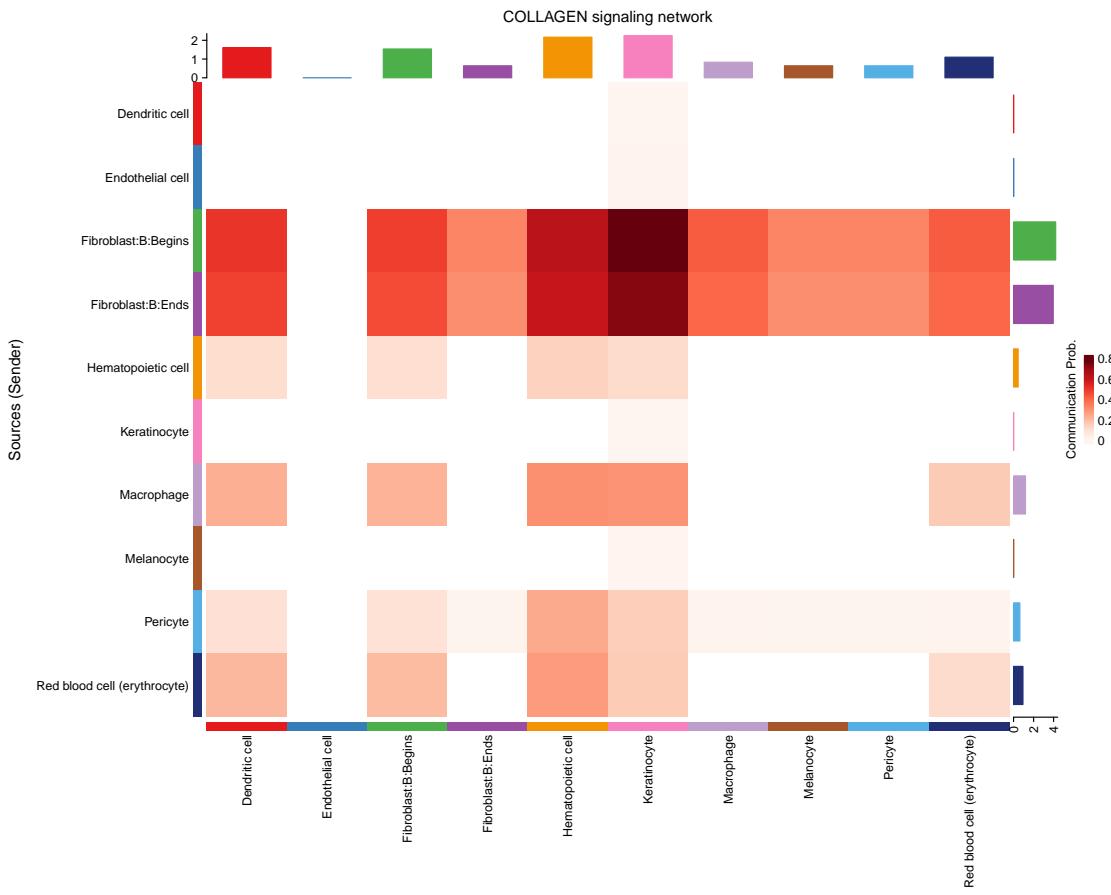


Figure 25: Diff path COLLAGEN

'Diff path others' 数据已全部提供。

(对应文件为 Figure+Table/Diff-path-others)

注：文件夹 Figure+Table/Diff-path-others 共包含 5 个文件。

1. 1\_ALL.pdf
2. 2\_COLLAGEN.pdf
3. 3\_MIF.pdf
4. 4\_MK.pdf
5. 5\_CD99.pdf

#### 6.4.5.4 Others

'Ligand receptor roles' 数据已全部提供。

(对应文件为 Figure+Table/ligand-receptor-roles)

注：文件夹 Figure+Table/ligand-receptor-roles 共包含 3 个文件。

1. 1\_outgoing.pdf
2. 2\_incoming.pdf
3. 3\_all.pdf

#### 6.4.5.5 蛋白互作 (PPI)

推测，Wnt 通路的表达变化可能影响到 FB:begins 和 FB:ends 与免疫细胞的通讯差异，因此这里试着构建 PPI 网络 (Functional, 功能网络)，首要查看姜黄素有调控作用的 Wnt 通路基因以及有调控作用的通讯的受体配体基因，两者之间是否存在可能的相互作用。

- Fig. 26, CD44 主要位于 COLLAGEN pathway, Tab. 6
- 联系 Fig. 25 可知，是 Macrophage 对 FB:begins 和 FB:ends 的 COLLAGEN 通讯有所不同。

R input

```
lp.imm <- dplyr::filter(cc@tables$step1$lp_net, pathway_name %in% diff.imm)
lp.imm <- dplyr::distinct(lp.imm[, -(1:2)], pathway_name, .keep_all = T)
genes.lp.imm <- c(lp.imm$ligand, lp.imm$receptor)
genes.lp.imm <- unlist(strsplit(genes.lp.imm, "_"))
genes.lp.imm

sdb.imm <- job_stringdb(c(genes.lp.imm, genes.allWnt))
sdb.imm <- step1(sdb.imm, 50, network_type = "full")

lstPPI.imm <- filter(sdb.imm,
  genes.wnt$geneName_list, genes.lp.imm,
  level.x = dplyr::select(Tops, hgnc_symbol, logFC),
  top = NULL, keep.ref = F, arrow = F, HLs = "CD44",
  label.shape = c(from = "Wnt", to = "Immune_LR")
)
lstPPI.imm$p.mcc
```

Figure 26 (下方图) 为图 PPI Wnt LR 概览。

(对应文件为 Figure+Table/PPI-Wnt-LR.pdf)

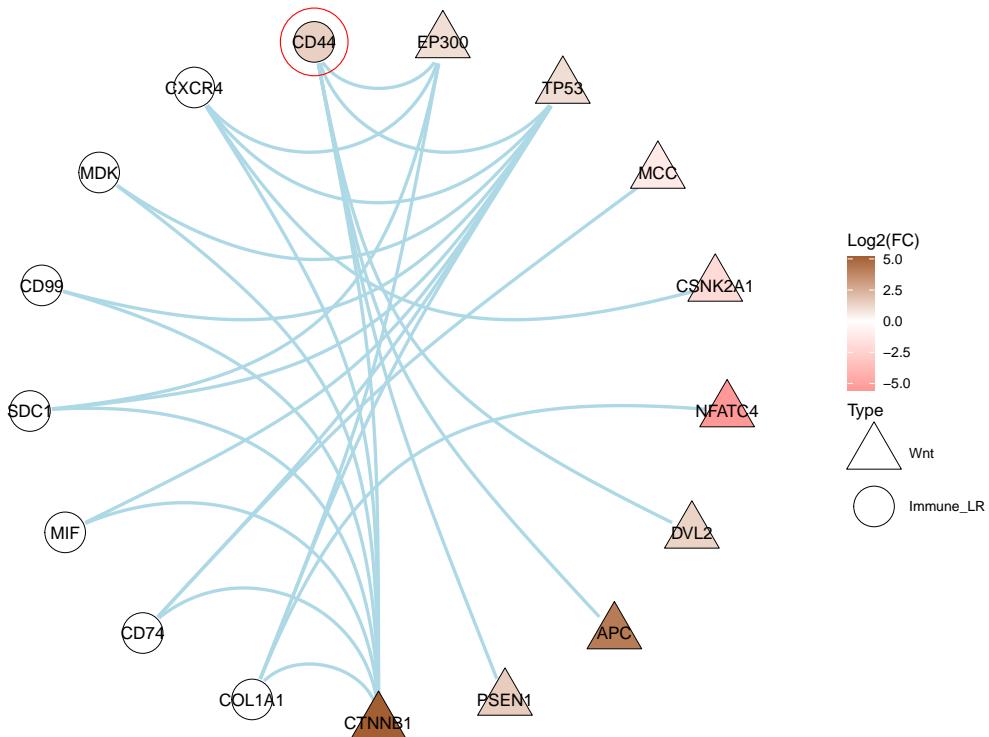


Figure 26: PPI Wnt LR

Table 6 (下方表格) 为表格 LR information 概览。

(对应文件为 Figure+Table/LR-information.csv)

注：表格共有 4 行 9 列，以下预览的表格可能省略部分数据；含有 4 个唯一 ‘ligand’。

1. evidence: 证据，相关文献中的描述。

Table 6: LR information

ligand	receptor	prob	pval	interact... er	interact... er	pathway... s	annotation... s	evidence
MIF	CD74_C...	0.0088...	0	MIF_CD...	MIF - ...	MIF	Secret... ory	PMID: ...
MDK	SDC1	0.0013...	0	MDK_SDC1	MDK - ...	MK	Secret... ory	PMID: ...
COL1A1	CD44	0.1542...	0	COL1A1...	COL1A1...	COLLAGEN	ECM-Re... ceptor	KEGG: ...
CD99	CD99	0.0712...	0	CD99_CD99	CD99 - ...	CD99	Cell-C... ell	KEGG: ...

## 6.5 HNRNPH1、Wnt 与姜黄素

### 6.5.1 分子对接结果

注：以下蛋白的 PDB 获取于 alphaFold。

- APCDD1 = “./material/APCDD1.pdb”,
- PLCB4 = “./material/PLCB4.pdb”,
- PRICKLE1 = “./material/PRICKLE1.pdb”

其余 PDB 文件获取于 PDB 数据库

R input

```
vn <- job_vina(c(Curcumin = 969516), c(genes.wnt$geneName_list, "HNRNPH1"))
# file.copy("~/Downloads/AF-Q8J025-F1-model_v4.pdb", "./material/APCDD1.pdb")
# file.copy("~/Downloads/AF-Q15147-F1-model_v4.pdb", "./material/PLCB4.pdb")
# file.copy("~/Downloads/AF-Q96MT3-F1-model_v4.pdb", "./material/PRICKLE1.pdb")

vn <- step1(vn, pdbs = c(CAMK2G = "2V70"))
vn <- step2(vn)
vn <- step3(vn, extra_pdb.files = c(
    APCDD1 = "./material/APCDD1.pdb",
    PLCB4 = "./material/PLCB4.pdb",
    PRICKLE1 = "./material/PRICKLE1.pdb")
)
# vn <- set_remote(vn)
vn <- step4(vn)
vn <- step5(vn, cutoff.af = 0)
wrap(vn@plots$step5$p.res_vina, 7, 5)
```

APCDD1 的对接取得了优异的亲和度能量。

Figure 27 (下方图) 为图 Overall combining Affinity 概览。

(对应文件为 Figure+Table/Overall-combining-Affinity.pdf)

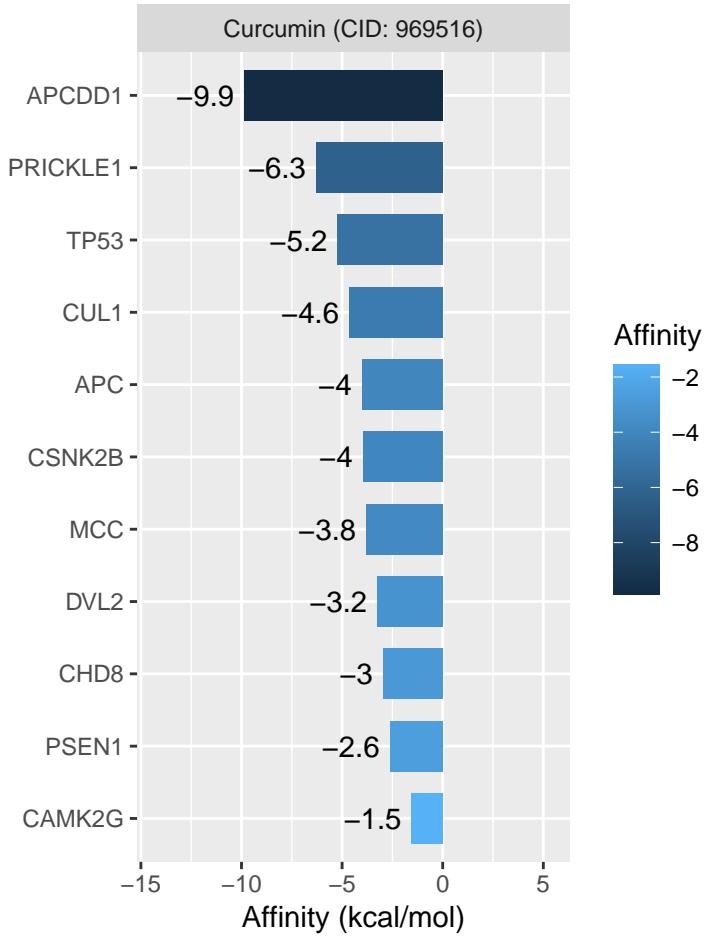


Figure 27: Overall combining Affinity

### 6.5.2 可视化

R input

```
vn <- step6(vn, top = 3)
vn@plots$step6$Top1_969516_into_APDD1
vn <- step7(vn)
vn@plots$step7$Top1_969516_into_APDD1
```

Figure 28 (下方图) 为图 Docking 969516 into APCDD1 概览。

(对应文件为 Figure+Table/Docking-969516-into-APCDD1.png)

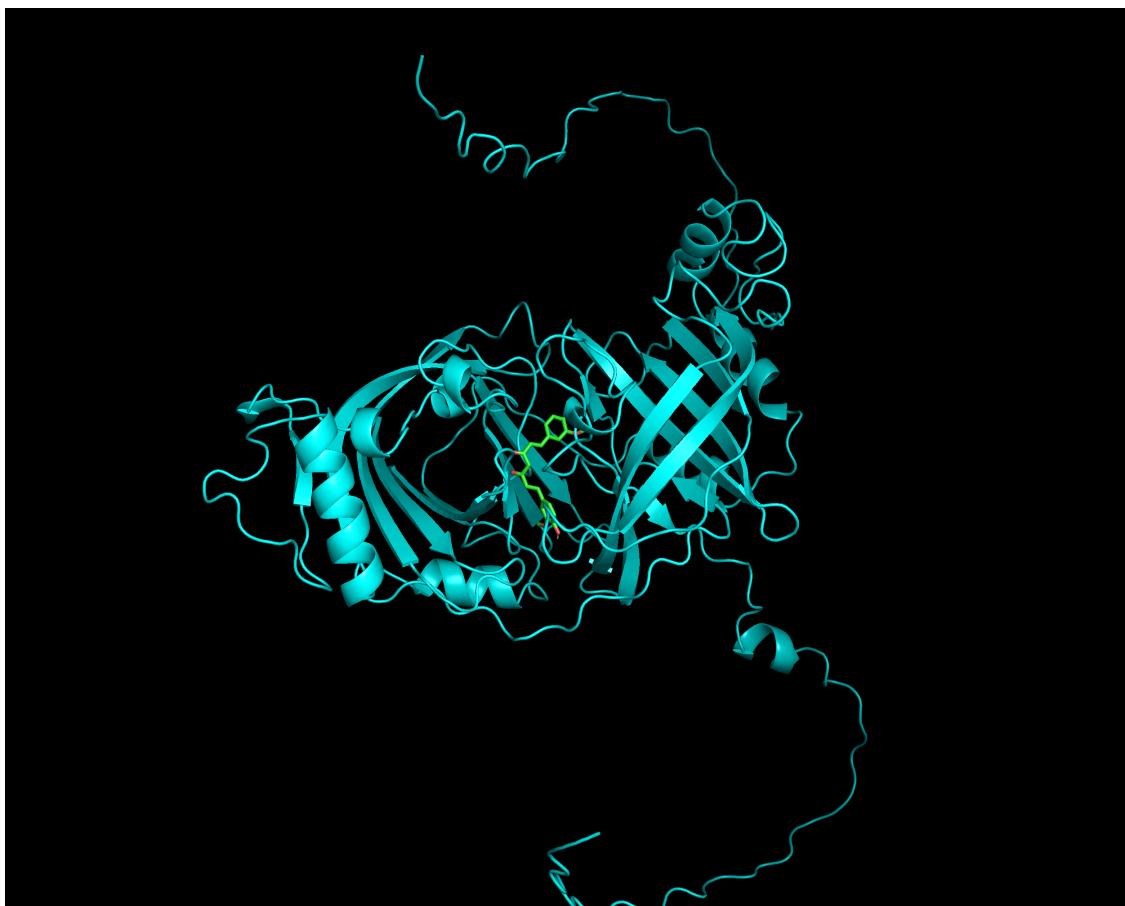


Figure 28: Docking 969516 into APCDD1

Figure 29 (下方图) 为图 Docking 969516 into APCDD1 detail 概览。

(对应文件为 [Figure+Table/Docking-969516-into-APCDD1-detail.png](#))

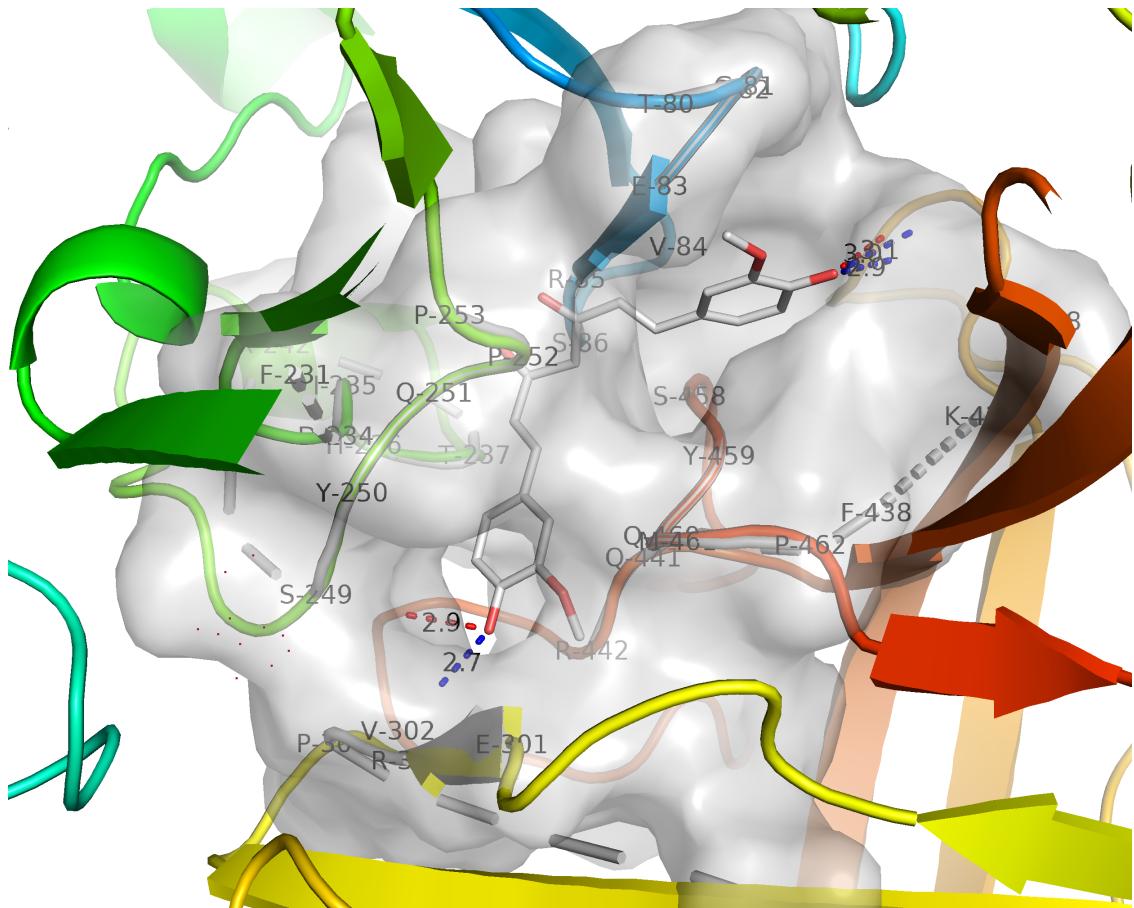


Figure 29: Docking 969516 into APCDD1 detail

## 6.6 Session Info

R input

```
sessionInfo()
```

```
## R version 4.4.0 (2024-04-24)
## Platform: x86_64-pc-linux-gnu
## Running under: Pop!_OS 22.04 LTS
##
## Matrix products: default
## BLAS:    /usr/lib/x86_64-linux-gnublas/libblas.so.3.10.0
## LAPACK:  /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.10.0
##
## locale:
## [1] LC_CTYPE=en_US.UTF-8          LC_NUMERIC=C           LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
## [5] LC_MONETARY=en_US.UTF-8       LC_MESSAGES=en_US.UTF-8      LC_PAPER=en_US.UTF-8       LC_NAME=C
## [9] LC_ADDRESS=C                  LC_TELEPHONE=C         LC_ME量UREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
```

```

## 
## time zone: Asia/Shanghai
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats4     grid      stats      graphics   grDevices  utils      datasets   methods    base
## 
## other attached packages:
## [1] Seurat_4.9.9.9067          SeuratObject_4.9.9.9091    sp_2.0-0           utils.tool_0.1.0
## [5] MCnebula2_0.0.9000         ggplot2_3.4.2              tidyHeatmap_1.10.1 monocle3_1.3.0
## [9] SummarizedExperiment_1.30.2 GenomicRanges_1.52.0       GenomeInfoDb_1.36.1 IRanges_2.34.0
## [13] S4Vectors_0.38.1          MatrixGenerics_1.12.3     matrixStats_1.0.0  biomaRt_2.56.0
## [17] Biobase_2.60.0            BiocGenerics_0.46.0       nvimcom_0.9-146
## 
## loaded via a namespace (and not attached):
## [1] igraph_1.5.1               hash_2.2.6.2             ica_1.0-3           plotly_4.10.2
## [5] Formula_1.2-5              zlibbioc_1.46.0          tidyselect_1.2.0    bit_4.0.5
## [9] doParallel_1.0.17           clue_0.3-64              lattice_0.22-6     rjson_0.2.21
## [13] blob_1.2.4                stringr_1.5.0            rngtools_1.5.2     S4Arrays_1.0.5
## [17] parallel_4.4.0              png_0.1-8               plotrix_3.8-2      cli_3.6.1
## [21] ggridify_0.1.2              registry_0.5-1          askpass_1.1         openssl_2.1.0
## [25] goftest_1.2-3              textshaping_0.3.6       purrr_1.0.2         officer_0.6.2
## [29] BiocNeighbors_1.18.0        ggnetwork_0.5.12        uwot_0.1.16        dendextend_1.17.0
## [33] shadowtext_0.1.2           curl_5.0.1              mime_0.12          evaluate_0.21
## [37] tidytree_0.4.5              leiden_0.4.3             gsubfn_0.7          ComplexHeatmap_2.0
## [41] stringi_1.7.12             backports_1.4.1          desc_1.4.2          XML_3.99-0.14
## [45] ggVennDiagram_1.2.2         httpuv_1.6.11            AnnotationDbi_1.62.2 magrittr_2.0.3
## [49] clusterProfiler_4.9.0.002   rappdirs_0.3.3          splines_4.4.0        CellChat_1.6.1
## [53] jpeg_0.1-10                ggraph_2.1.0             dplyr_1.1.2          sctransform_0.4.0
## [57] DBI_1.1.3                 terra_1.7-39             withr_2.5.0          systemfonts_1.0.0
## [61] rprojroot_2.0.3             enrichplot_1.20.0        lmtest_0.9-40        brio_1.1.3
## [65] tidygraph_1.2.3             BiocManager_1.30.22      htmlwidgets_1.6.2    fs_1.6.3
## [69] ggrepel_0.9.3              statnet.common_4.9.0      BPCells_0.1.0        reticulate_1.31
## [73] zoo_1.8-12                 XVector_0.40.0           knitr_1.43          network_1.18.1
## [77] foreach_1.5.2              fansi_1.0.4              patchwork_1.1.2     caTools_1.18.2
## [81] data.table_1.14.8            RSpectra_0.16-1          irlba_2.3.5.1       fastDummies_1.7.3
## [85] gridGraphics_0.5-1           ellipsis_0.3.2          lazyeval_0.2.2       yaml_2.3.7
## [89] survival_3.5-7              scattermore_1.2          crayon_1.5.2        RcppAnnoy_0.0.21
## [93] RColorBrewer_1.1-3          tidyR_1.3.0              progressr_0.14.0     tweenr_2.0.2
## [97] later_1.3.1                ggridges_0.5.4            codetools_0.2-20     base64enc_0.1-3
## [ reached getOption("max.print") -- omitted 149 entries ]

```

## Reference

1. Jin, S. *et al.* Inference and analysis of cell-cell communication using cellchat. *Nature Communications* **12**, (2021).
2. Szklarczyk, D. *et al.* The string database in 2021: Customizable proteinprotein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research* **49**, D605–D612 (2021).
3. Chin, C.-H. *et al.* CytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology* **8**, S11 (2014).
4. Wu, T. *et al.* ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, (2021).
5. Ritchie, M. E. *et al.* Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47 (2015).
6. Chen, Y., McCarthy, D., Ritchie, M., Robinson, M. & Smyth, G. EdgeR: Differential analysis of sequence read count data user's guide. 119.
7. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nature Methods* **14**, (2017).
8. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology* **32**, (2014).
9. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, (2021).
10. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, (2019).
11. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling* **61**, 3891–3898 (2021).
12. Zhang, Y., Forli, S., Omelchenko, A. & Sanner, M. F. AutoGridFR: Improvements on autodock affinity maps and associated software tools. *Journal of computational chemistry* **40**, 2882–2886 (2019).
13. Zhang, Y. & Sanner, M. F. AutoDock crankpep: Combining folding and docking to predict protein-peptide complexes. *Bioinformatics (Oxford, England)* **35**, 5121–5127 (2019).
14. Ravindranath, P. A. & Sanner, M. F. AutoSite: An automated approach for pseudo-ligands prediction-from ligand-binding sites identification to predicting key ligand atoms. *Bioinformatics (Oxford, England)* **32**, 3142–3149 (2016).
15. Ravindranath, P. A., Forli, S., Goodsell, D. S., Olson, A. J. & Sanner, M. F. AutoDockFR: Advances in protein-ligand docking with explicitly specified binding site flexibility. *PLoS computational biology* **11**, (2015).
16. Luo, W. & Brouwer, C. Pathview: An r/bioconductor package for pathway-based data integration and visualization. *Bioinformatics (Oxford, England)* **29**, 1830–1831 (2013).

17. Cao, Y., Wang, X. & Peng, G. SCSA: A cell type annotation tool for single-cell rna-seq data. *Frontiers in genetics* **11**, (2020).