

Analysis

Huang LiChuang of Wie-Biotech

Contents

1 题目	4
2 摘要	4
3 前言	4
4 材料和方法	4
5 分析结果	4
5.1 RCC 数据集分析鉴定	4
5.2 不同类型 CKD 数据共表达分析和重聚类	5
5.3 RCC 和 CKD 数据集成	6
5.4 差异分析和富集分析	7
5.5 以新的 IgA 和 RCC 数据集验证	8
6 结论	9
7 附：分析流程	9
7.1 肾癌 (GEO)	9
7.1.1 GSE171306	9
7.1.1.1 细胞聚类和注释	9
7.1.1.2 癌细胞识别	10
7.1.1.3 癌细胞拟时分析和共表达分析	12
7.2 血管性疾患	15
7.2.1 高血压性肾炎: GSE210898 (hypertensive nephropathy)	15
7.2.1.1 细胞聚类和注释	15
7.2.1.2 共表达分析	17
7.3 (原发性) 肾小球性疾患	19
7.3.1 IgA 肾病: GSE171314 (IgA Nephropathy)	19
7.3.1.1 细胞聚类和注释	19
7.3.1.2 共表达分析	21
7.3.2 膜性肾病: GSE241302 (idiopathic membranous nephropathy)	23
7.3.2.1 细胞聚类和注释	23

7.3.2.2	共表达分析	25
7.4	正常对照 (Healthy control) (GSE171314)	27
7.4.0.1	细胞聚类和注释	27
7.5	多数据集成: ccRCC 和 CKD scRNA-seq	29
7.5.1	使用 RISC 对不同来源的数据集成	29
7.5.2	拟时分析	30
7.5.3	IgA、Cancer、HC 差异分析	31
7.5.4	富集分析 (GSEA)	33
7.5.4.1	Cancer vs IgA	33
7.5.4.2	Cancer vs Control	37
7.6	验证	38
7.6.1	另一组 IgA 肾病: GSE127136	38
7.6.1.1	细胞聚类和注释	38
7.6.2	另一组 RCC: GSE202374	38
7.6.2.1	细胞聚类和注释	38
7.6.3	集成验证: Cancer、IgA	39
7.6.3.1	RISC 集成	39
7.6.3.2	拟时分析	41
7.6.3.3	富集分析	42
	Reference	46

List of Figures

1	Main figure 1	5
2	Main figure 2	6
3	Main figure 3	6
4	Main figure 4	7
5	Main figure 5	8
6	RCC data1 QC	9
7	RCC data1 SCSA mapping	10
8	RCC data1 copykat prediction	11
9	RCC data1 SCSA mapping with copykat prediction	12
10	RCC data1 pseudotime	13
11	RCC data1 co expression modules	14
12	RCC data1 cancer subtype	15
13	HN data QC	16
14	HN data SCSA mapping	16
15	HN data pseudotime	17
16	HN data co expression modules	18
17	HN data Proximal tubule cell subtype	19
18	IgA data1 QC	20

19	IgA data1 SCSA mapping	20
20	IgA data1 pseudotime	21
21	IgA data1 co expression modules	22
22	IgA data1 Proximal tubule cell subtype	23
23	IMN data QC	24
24	IMN data SCSA mapping	24
25	IMN data pseudotime	25
26	IMN data co expression modules	26
27	IMN data Proximal tubule cell subtype	27
28	HC data QC	28
29	HC data SCSA mapping	28
30	SETS1 select reference dataset for integration	29
31	SETS1 umap mapping	30
32	SETS1 trajectory	30
33	SETS1 pseudotime	31
34	SETS1 intersection of contrasts DEGs	32
35	SETS1 GSEA enrichment of KEGG CancerVsIgA	34
36	SETS1 GSEA enrichment of GO CancerVsIgA	35
37	SETS1 GSEA show OXPHOS pathway CancerVsIgA	36
38	SETS1 GSEA show OXPHOS pathway by pathview CancerVsIgA	36
39	SETS1 GSEA enrichment of KEGG CancerVsControl	37
40	SETS1 GSEA enrichment of GO CancerVsControl	38
41	RCC data2 SCSA mapping with copykat prediction	39
42	SETS2 select reference dataset for integration	40
43	SETS2 umap mapping	41
44	SETS2 pseudotime	42
45	SETS2 GSEA enrichment of KEGG CancerVsIgA	43
46	SETS2 GSEA enrichment of GO CancerVsIgA	44
47	SETS2 GSEA show OXPHOS pathway CancerVsIgA	45
48	SETS2 GSEA show OXPHOS pathway by pathview CancerVsIgA	45

List of Tables

1	SETS1 DEGs	32
---	----------------------	----

1 题目

2 摘要

3 前言

慢性肾病（CKD）与患癌症风险之间的联系尚未明确。尽管多项研究观察到需要透析或肾移植的 ESRD 患者患癌症的风险较高，但相对不严重的肾脏疾病是否与癌症相关仍知之甚少¹⁻³。已有研究论及 CKD 和肾癌之间的关联性和转化的风险²⁻⁴，轻度至中度 CKD 和移植受者的癌症风险增加³。

在本研究中，为了探究 CKD 与肾癌（RCC）之间的相关性和转化风险，重新分析了多组公共数据库的单细胞数据集。考虑到 CKD 的复杂性，这里并不采用单一的 CKD 病型研究，而是搜集了不同类型的 CKD (hypertensive nephropathy, HN; IgA; idiopathic membranous nephropathy, IMN)，并采用了强大的单细胞数据集成算法 RISC 消除不同来源（不同实验室、不同批次、不同类型患者等不相关因素）批次效应，试探索 CKD 与肾癌之间的相关性。并在此之后，以一批新的 CKD 和 RCC 单细胞数据集验证。

4 材料和方法

涉及的 6 个 GEO 数据集可参考 7 的次级标题。

涉及的方法：

- Seurat^{5,6}
- Monocle^{7,8}
- copyKAT⁹
- RISC¹⁰
- ClusterProfiler¹¹
- ...

5 分析结果

5.1 RCC 数据集分析鉴定

Figure 1 为图 Main figure 1 概览。

(对应文件为 [./Figure+Table/fig1.pdf](#))

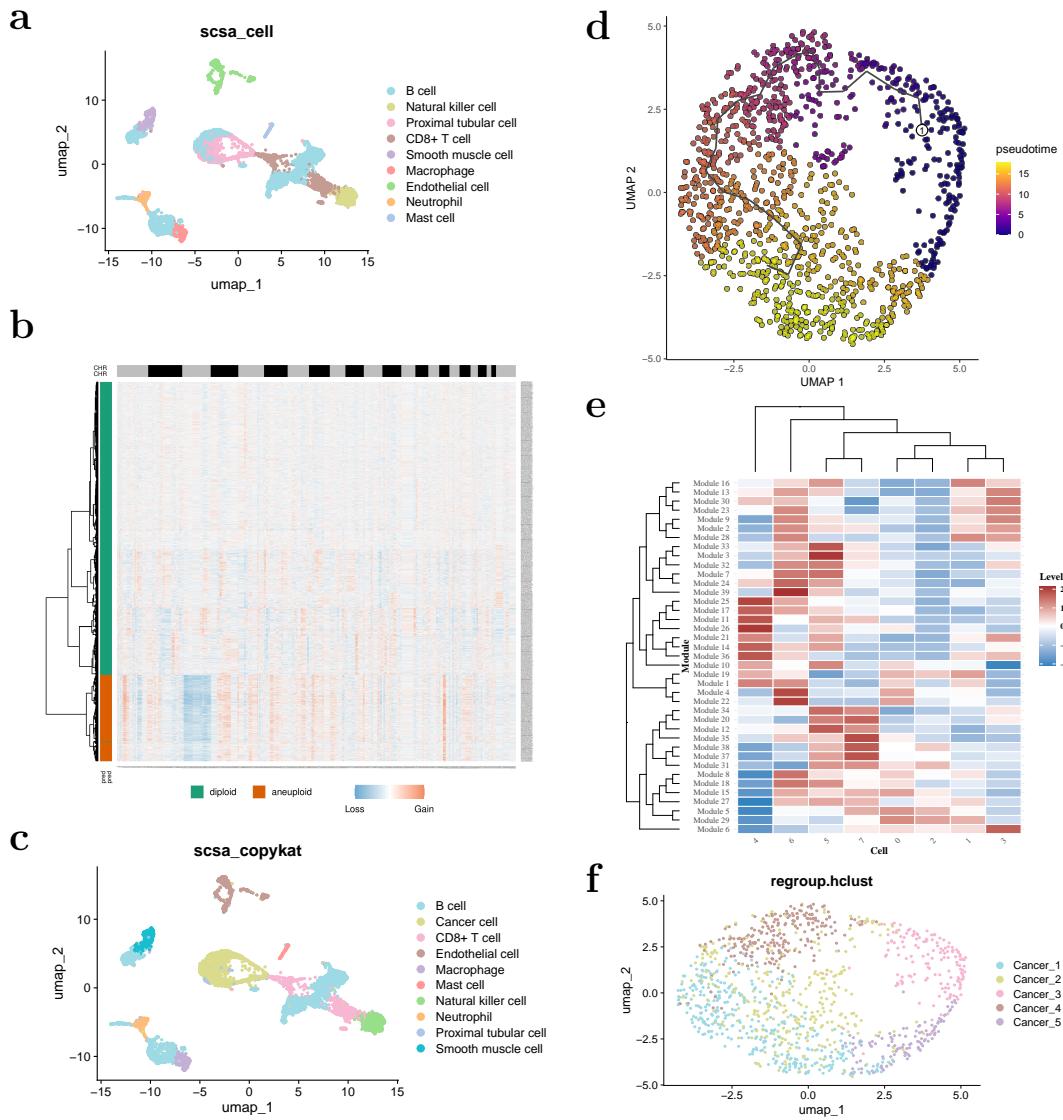


Figure 1: Main figure 1

5.2 不同类型 CKD 数据共表达分析和重聚类

Figure 2为图 Main figure 2 概览。

(对应文件为 ./Figure+Table/fig2.pdf)

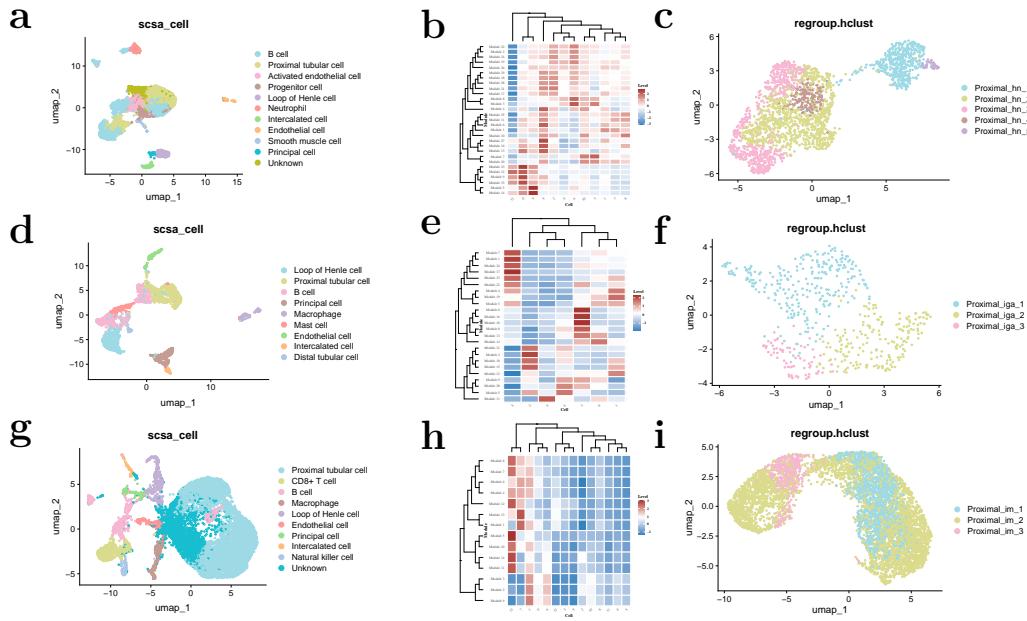


Figure 2: Main figure 2

5.3 RCC 和 CKD 数据集成

Figure 3为图 Main figure 3 概览。

(对应文件为 ./Figure+Table/fig3.pdf)

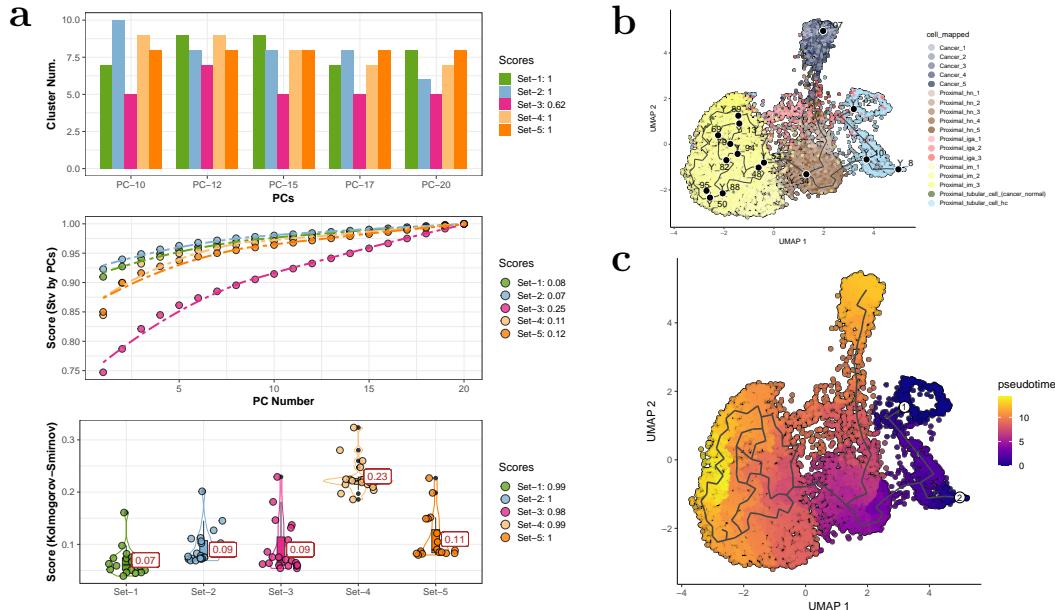


Figure 3: Main figure 3

Sets 1-5 对应 RCC, HM, IgA, IMN, HC 单细胞数据集。

5.4 差异分析和富集分析

Figure 4为图 Main figure 4 概览。

(对应文件为 ./Figure+Table/fig4.pdf)

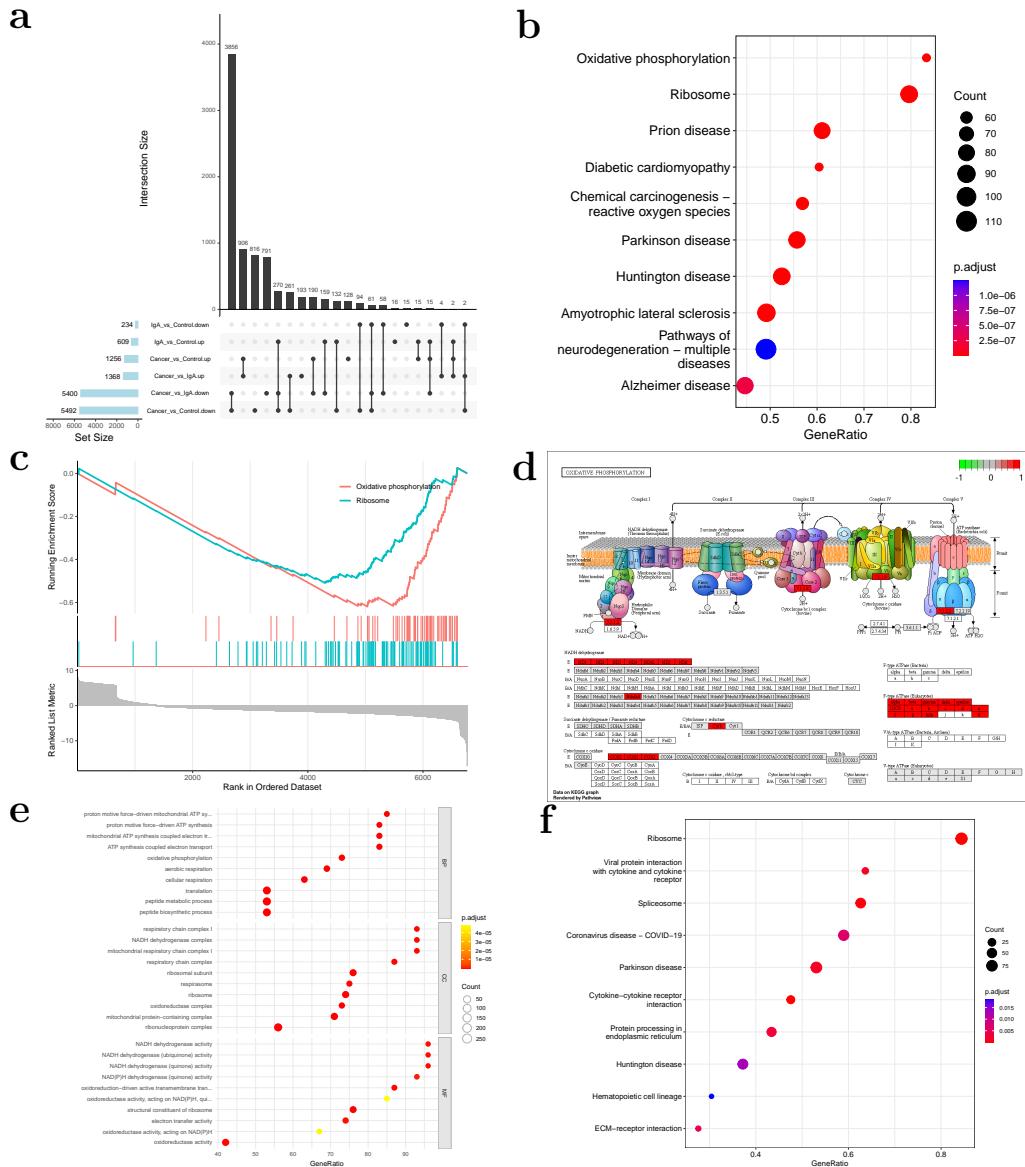


Figure 4: Main figure 4

5.5 以新的 IgA 和 RCC 数据集验证

Figure 5为图 Main figure 5 概览。

(对应文件为 ./Figure+Table/fig5.pdf)

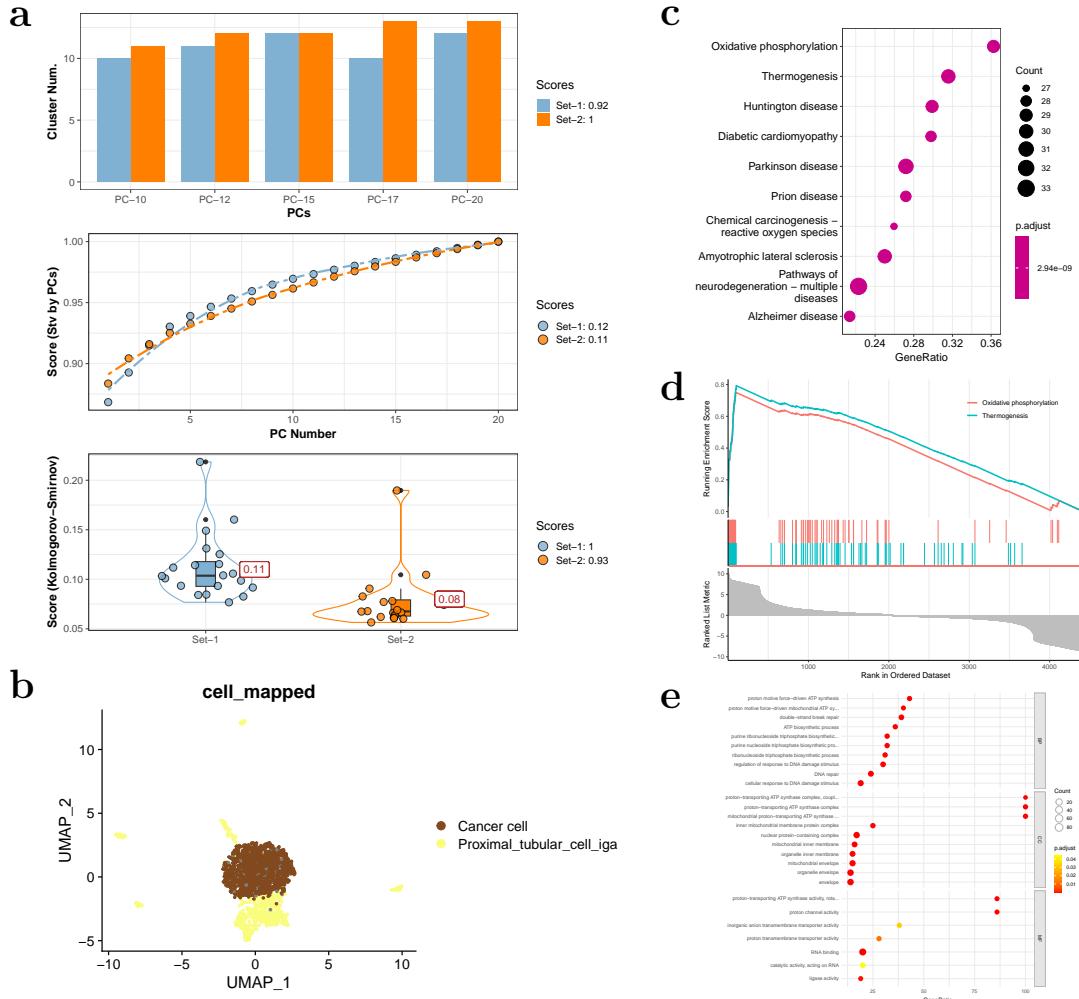


Figure 5: Main figure 5

6 结论

7 附：分析流程

7.1 肾癌 (GEO)

7.1.1 GSE171306

- Single-cell RNA sequencing (scRNA-seq) was performed on bilateral clear cell RCC (ccRCC). Primary kidney samples from 3 patients were used for single cell RNA sequencing by 10X Genomics

7.1.1.1 细胞聚类和注释 Figure 6为图 RCC data1 QC 概览。

(对应文件为 Figure+Table/RCC-data1-QC.pdf)

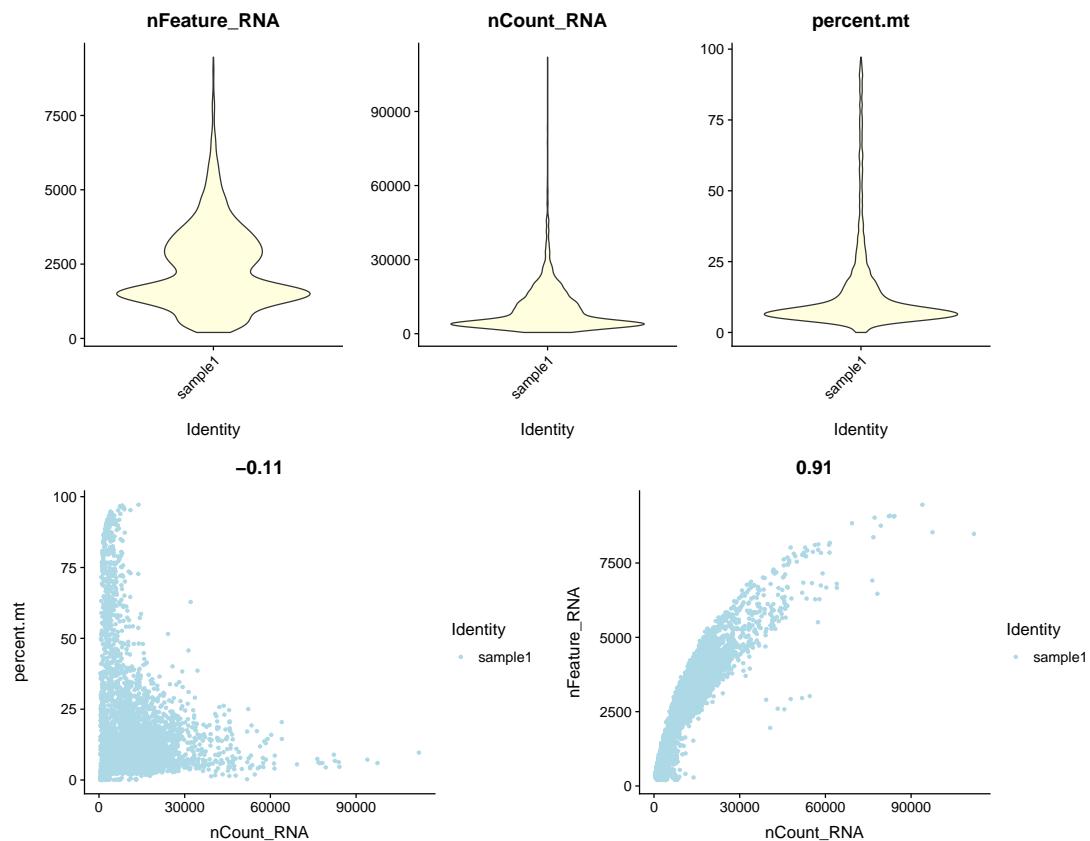


Figure 6: RCC data1 QC

使用 SCSA 对细胞类型注释（以下注释同）¹²。

Figure 7为图 RCC data1 SCSA mapping 概览。

(对应文件为 Figure+Table/RCC-data1-SCSA-mapping.pdf)

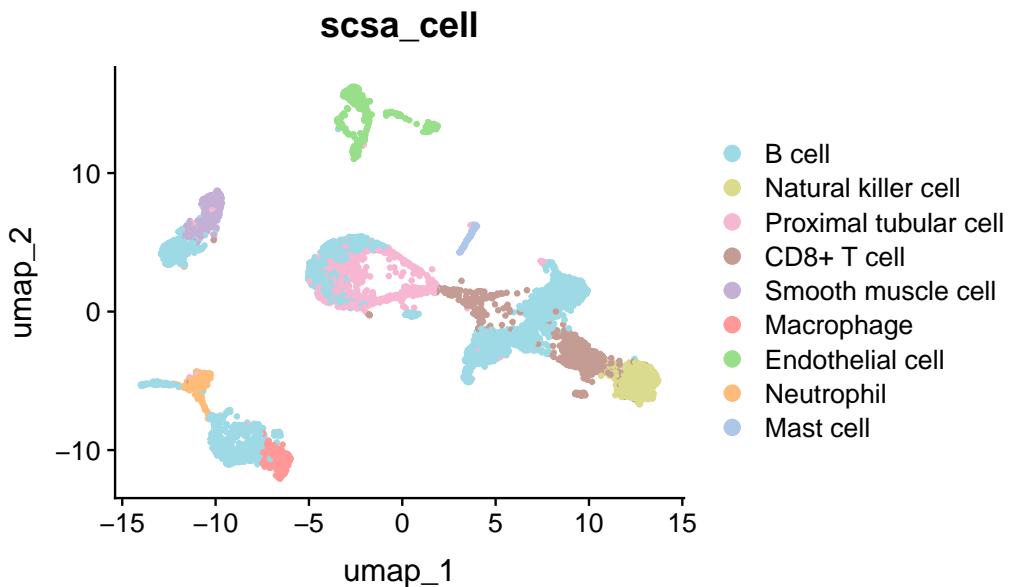


Figure 7: RCC data1 SCSA mapping

7.1.1.2 癌细胞识别 使用 copyKAT 预测癌细胞⁹。

Figure 8为图 RCC data1 copykat prediction 概览。

(对应文件为 Figure+Table/copykat_heatmap.png)

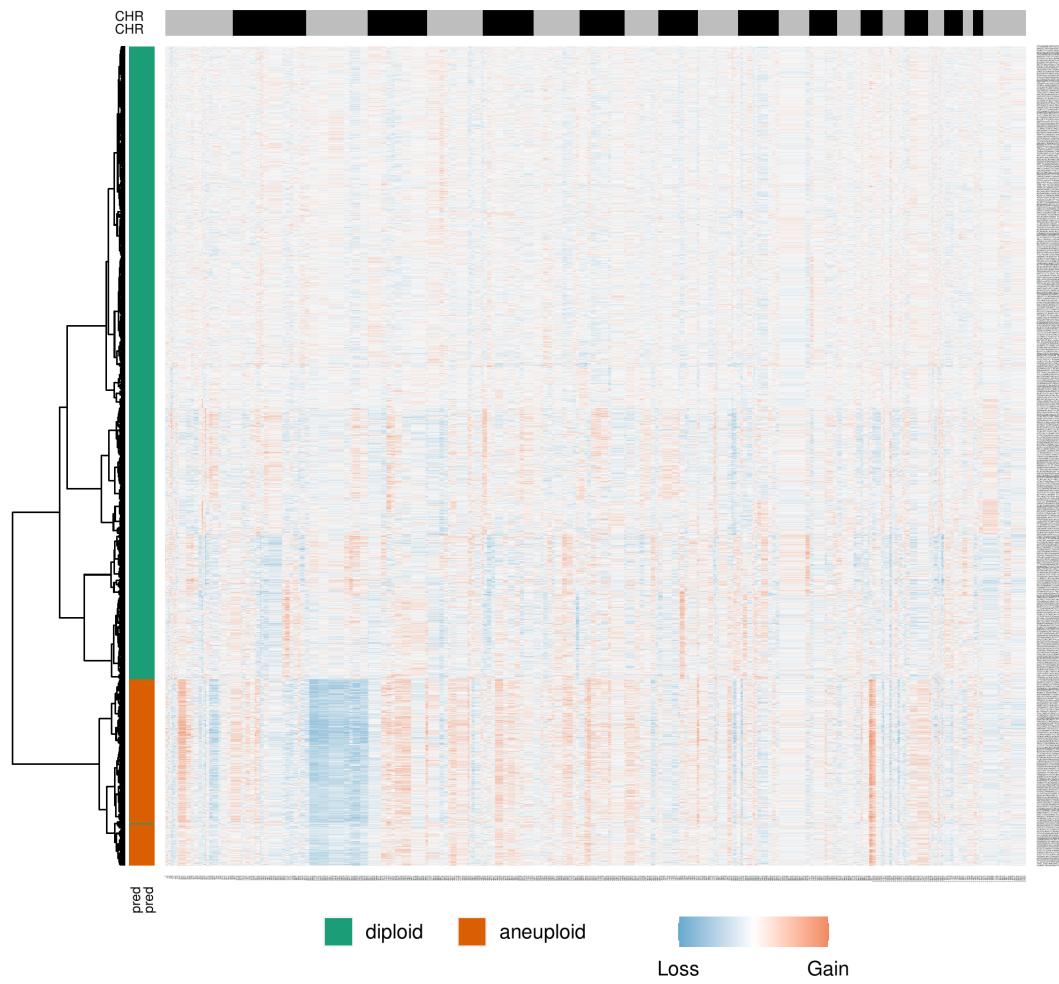


Figure 8: RCC data1 copykat prediction

Figure 9为图 RCC data1 SCSA mapping with copykat prediction 概览。

(对应文件为 [Figure+Table/RCC-data1-SCSA-mapping-with-copykat-prediction.pdf](#))

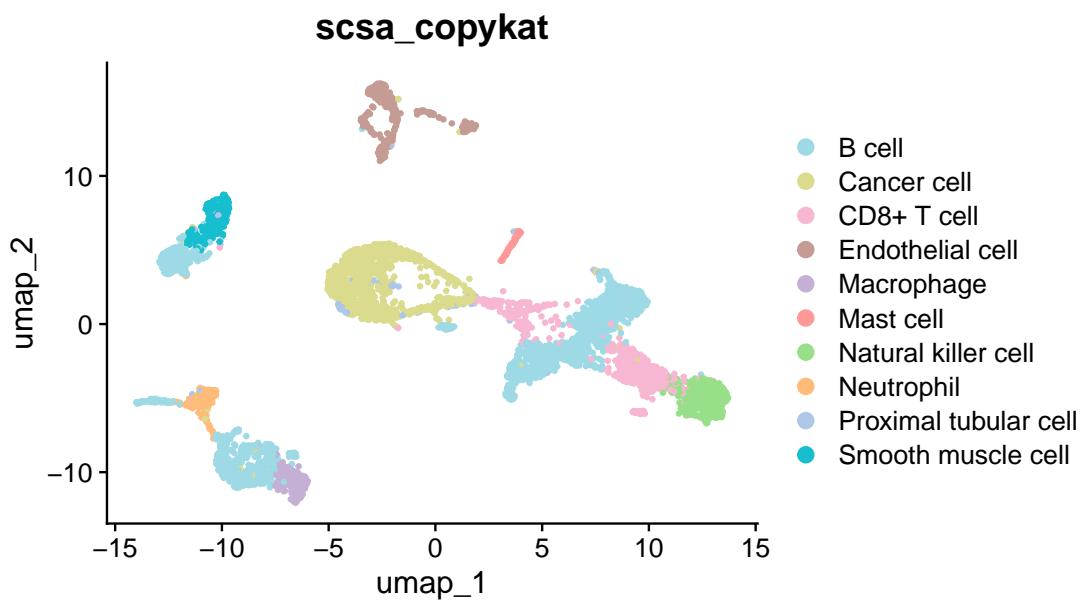


Figure 9: RCC data1 SCSA mapping with copykat prediction

7.1.1.3 癌细胞拟时分析和共表达分析 使用 monocle3 拟时分析^{7,8}。

Figure 10为图 RCC data1 pseudotime 概览。

(对应文件为 Figure+Table/RCC-data1-pseudotime.pdf)

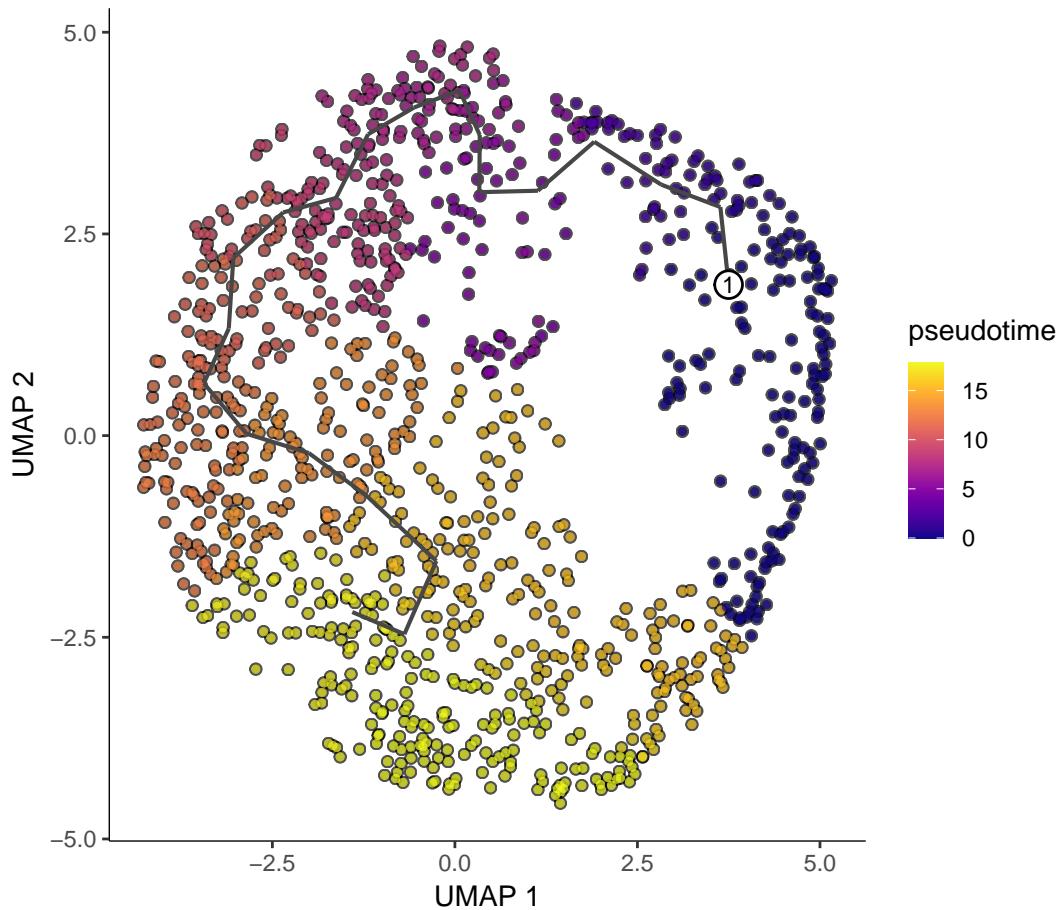


Figure 10: RCC data1 pseudotime

结合拟时分析和基因共表达模块，将肿瘤细胞类型切分为 5 个类型。

Figure 11为图 RCC data1 co expression modules 概览。

(对应文件为 [Figure+Table/RCC-data1-co-expression-modules.pdf](#))

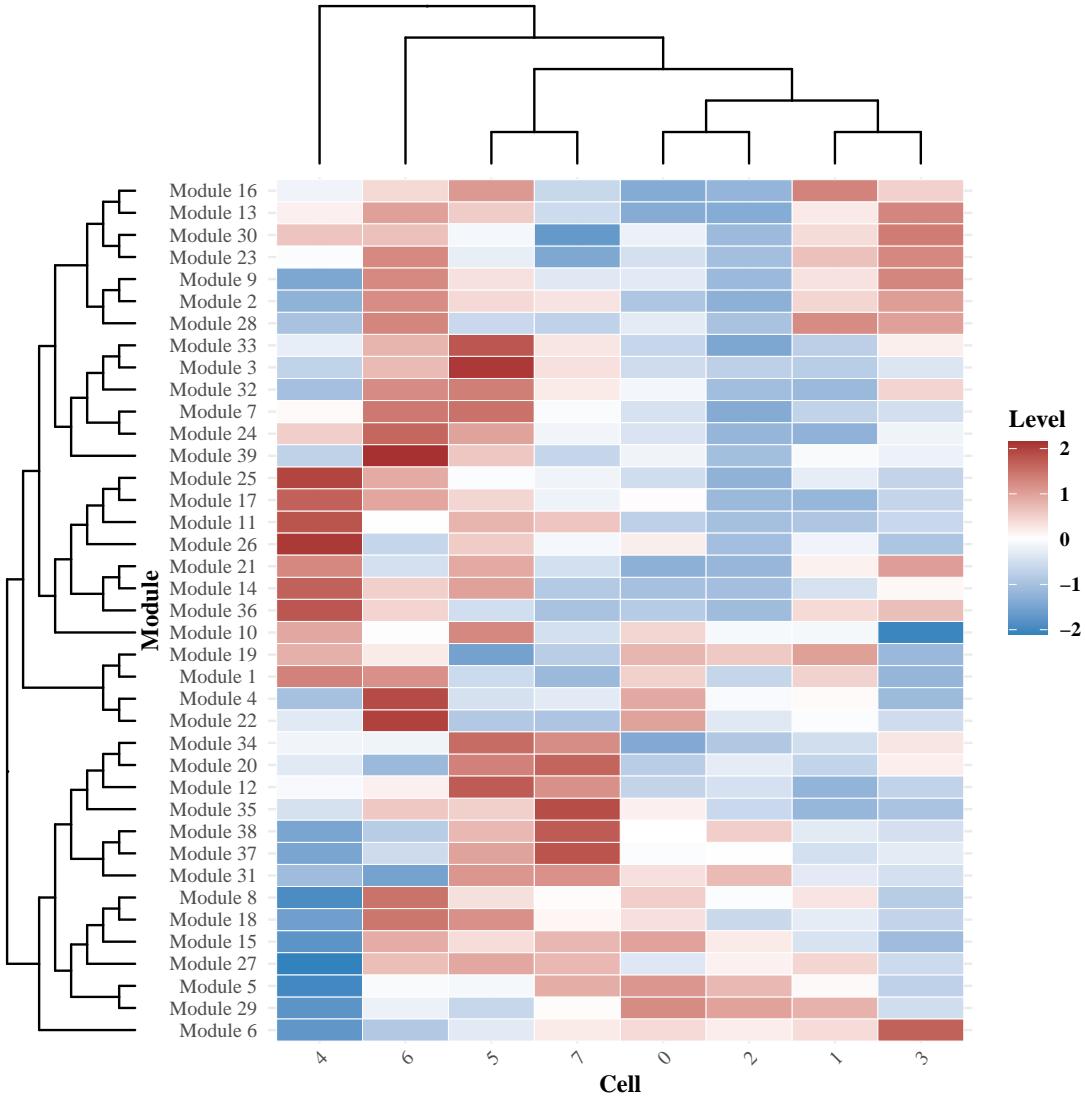


Figure 11: RCC data1 co expression modules

Figure 12为图 RCC data1 cancer subtype 概览。

(对应文件为 [Figure+Table/RCC-data1-cancer-subtype.pdf](#))

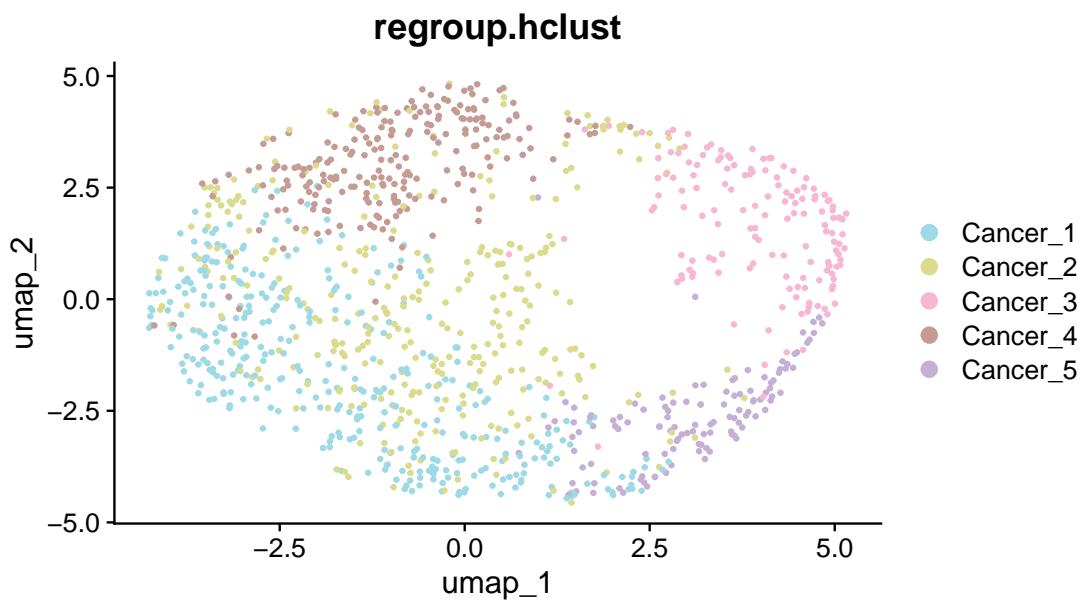


Figure 12: RCC data1 cancer subtype

7.2 血管性疾患

7.2.1 高血压性肾炎: GSE210898 (hypertensive nephropathy)

- Single-cell RNA transcriptomics of hypertensive nephropathy patients. We analyzed kidney samples from 3 patients with HTN using single-cell RNA sequencing, compared with previous data of controls

7.2.1.1 细胞聚类和注释 Figure 13为图 HN data QC 概览。

(对应文件为 Figure+Table/HN-data-QC.pdf)

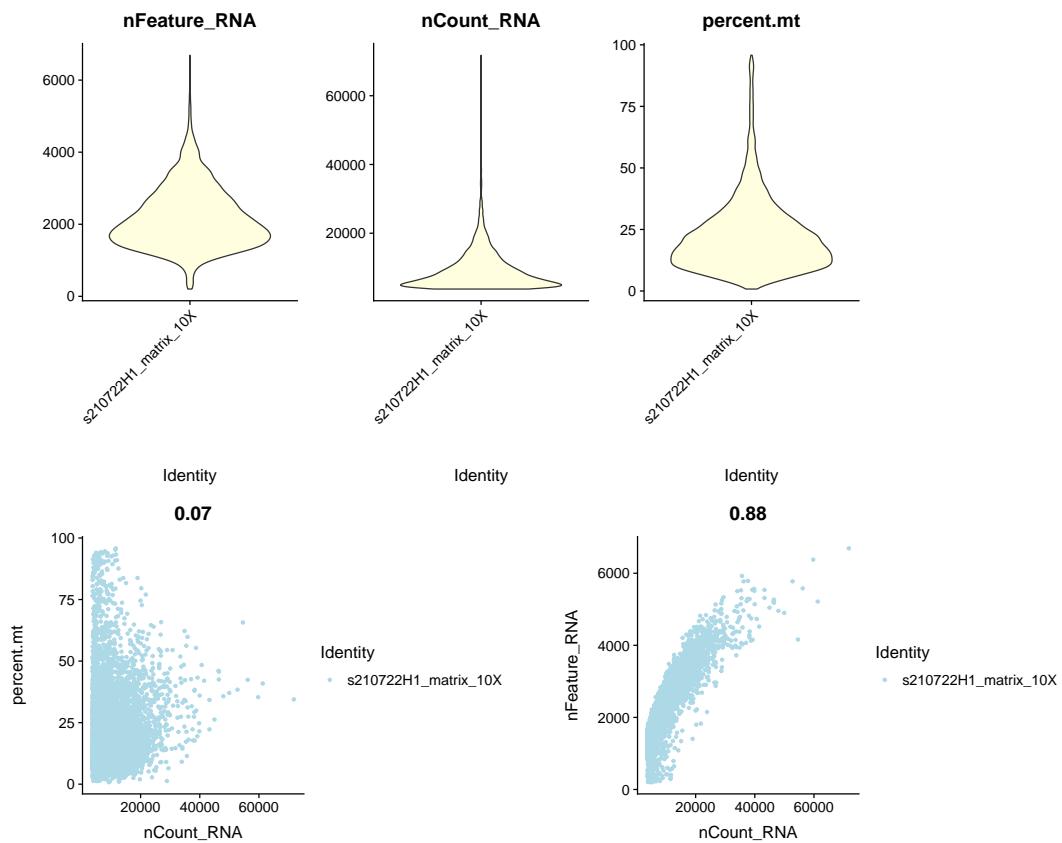


Figure 13: HN data QC

Figure 14为图 HN data SCSA mapping 概览。

(对应文件为 Figure+Table/HN-data-SCSA-mapping.pdf)

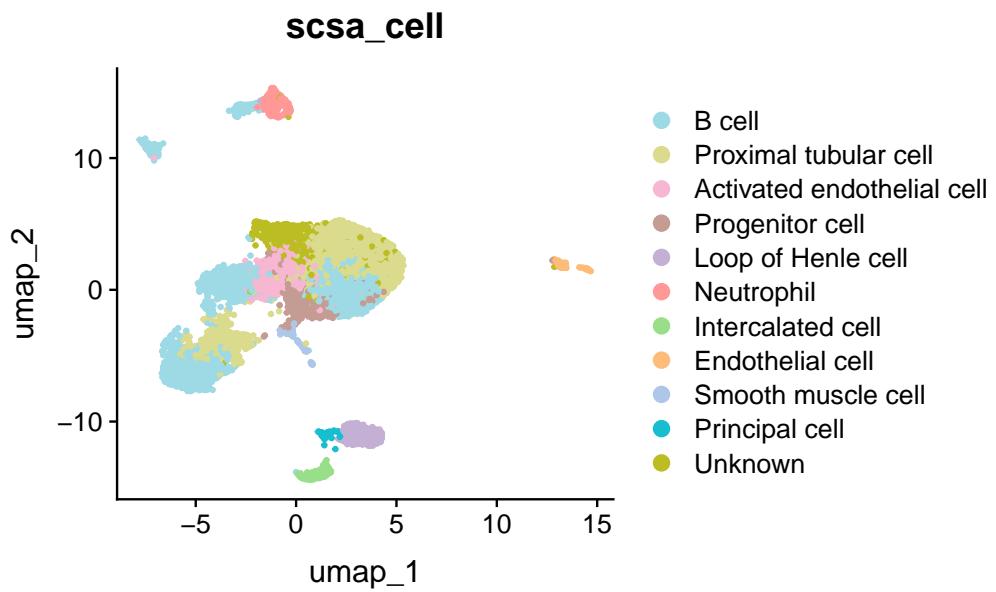


Figure 14: HN data SCSA mapping

7.2.1.2 共表达分析 Figure 15为图 HN data pseudotime 概览。

(对应文件为 Figure+Table/HN-data-pseudotime.pdf)

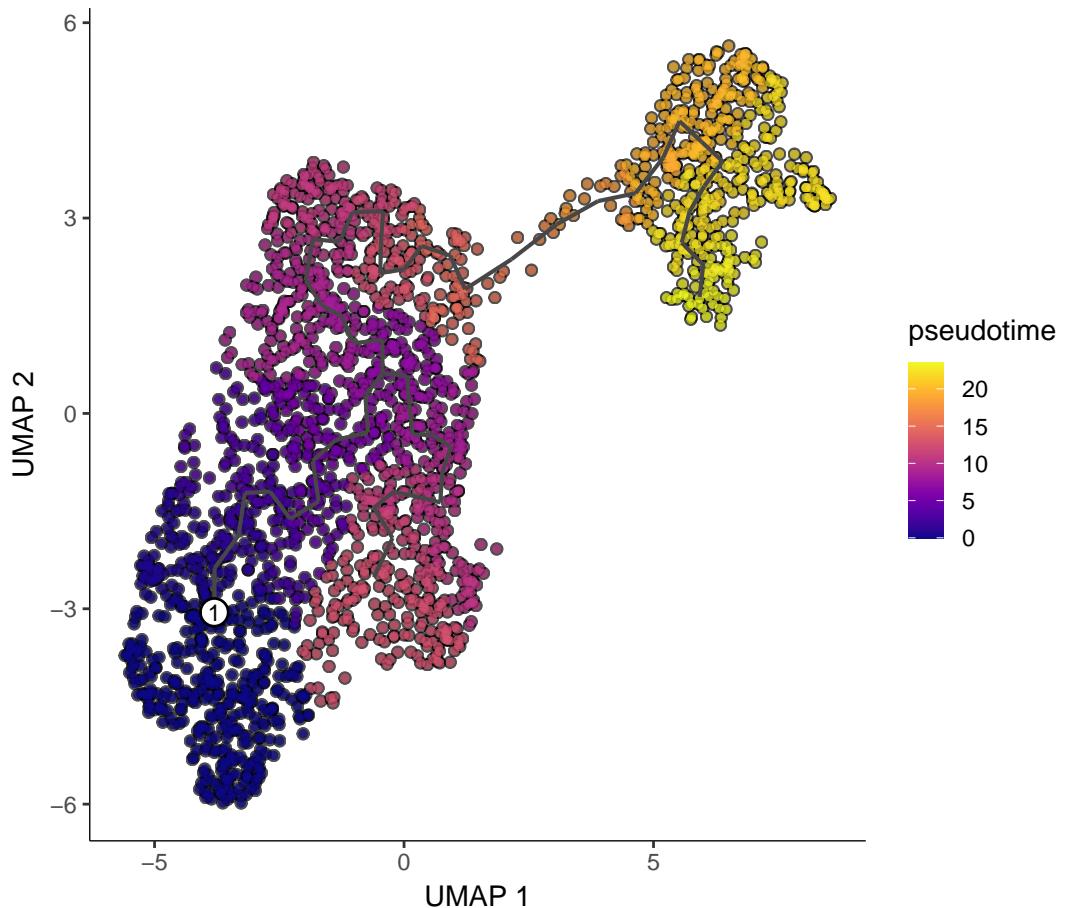


Figure 15: HN data pseudotime

Figure 16为图 HN data co expression modules 概览。

(对应文件为 Figure+Table/HN-data-co-expression-modules.pdf)

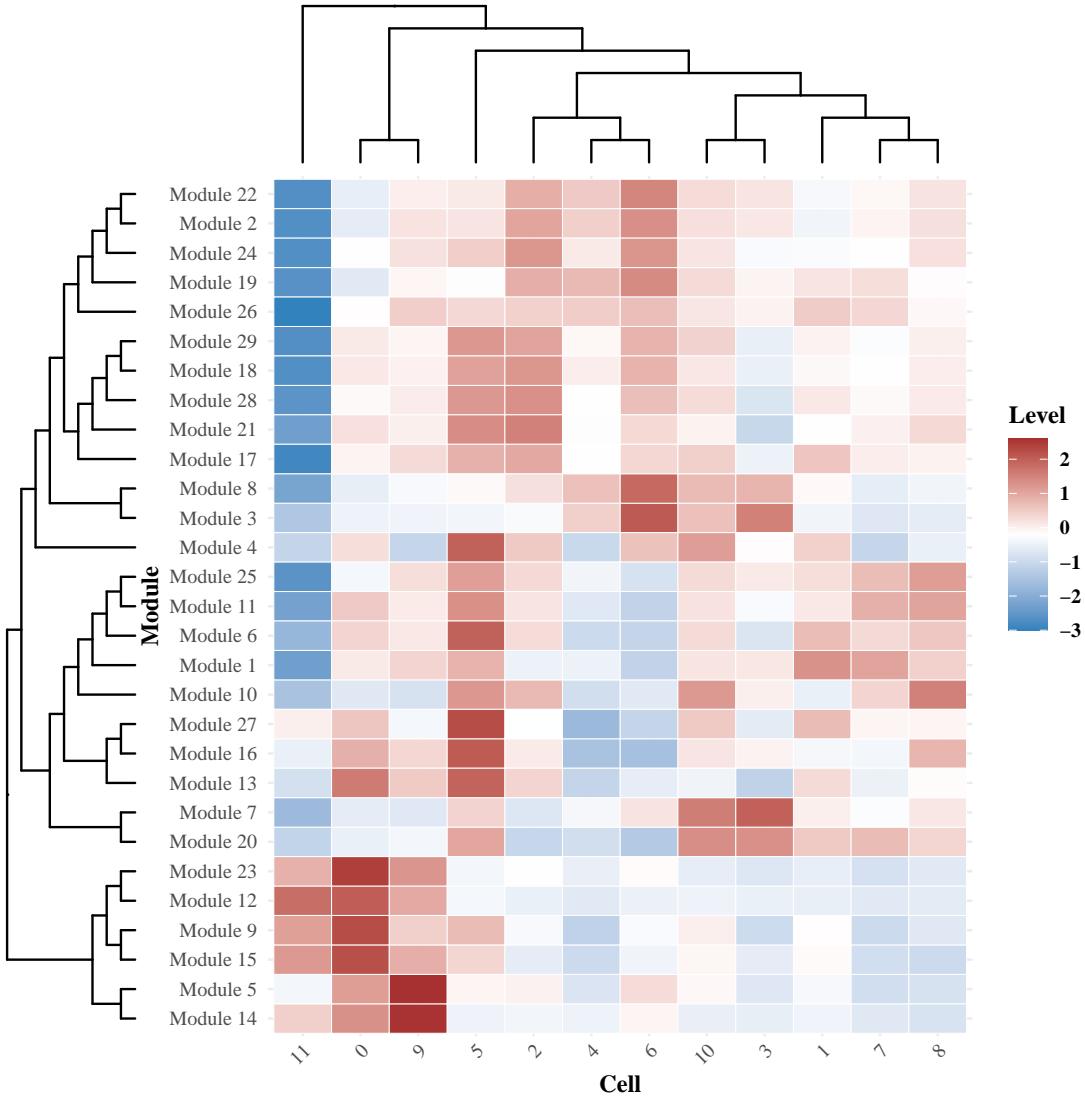


Figure 16: HN data co expression modules

Figure 17为图 HN data Proximal tubule cell subtype 概览。

(对应文件为 [Figure+Table/HN-data-Proximal-tubule-cell-subtype.pdf](#))

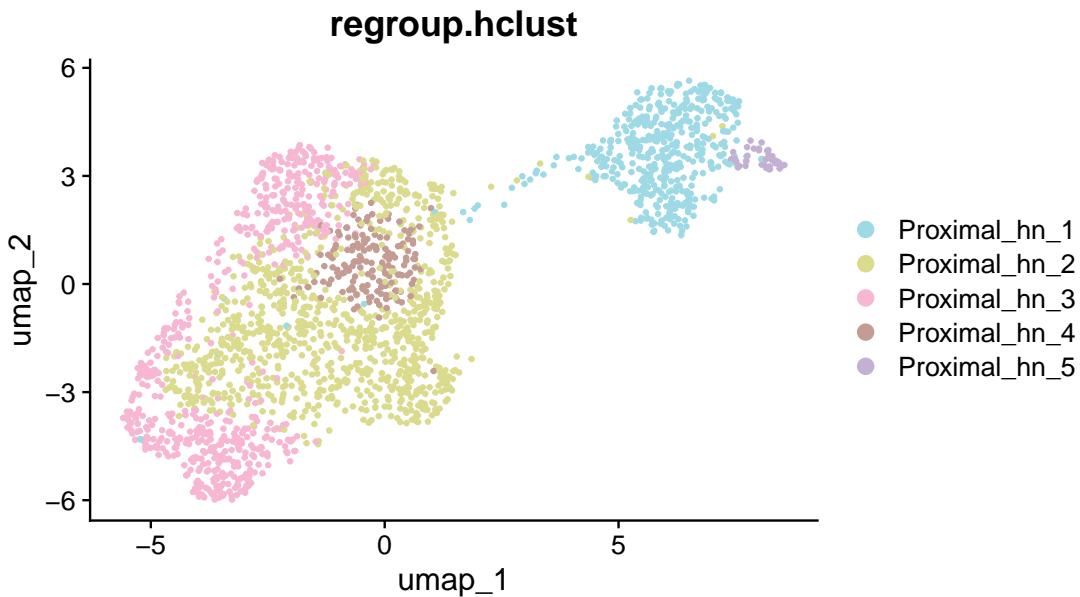


Figure 17: HN data Proximal tubule cell subtype

7.3 (原发性) 肾小球性疾患

7.3.1 IgA 肾病: GSE171314 (IgA Nephropathy)

- Single-cell RNA sequencing (scRNA-seq) was applied to kidney biopsies from 4 IgAN and 1 control subjects to define the transcriptomic landscape at the single-cell resolution.

7.3.1.1 细胞聚类和注释 Figure 18为图 IgA data1 QC 概览。

(对应文件为 Figure+Table/IgA-data1-QC.pdf)

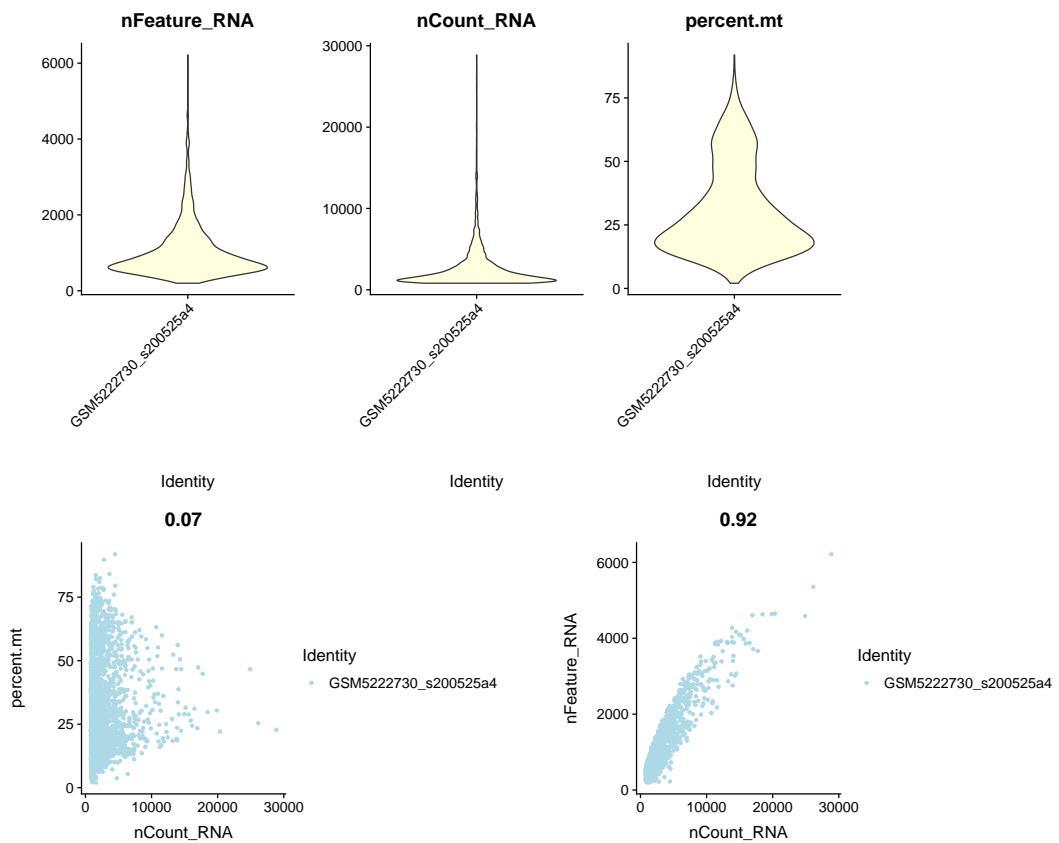


Figure 18: IgA data1 QC

Figure 19为图 IgA data1 SCSA mapping 概览。

(对应文件为 Figure+Table/IgA-data1-SCSA-mapping.pdf)

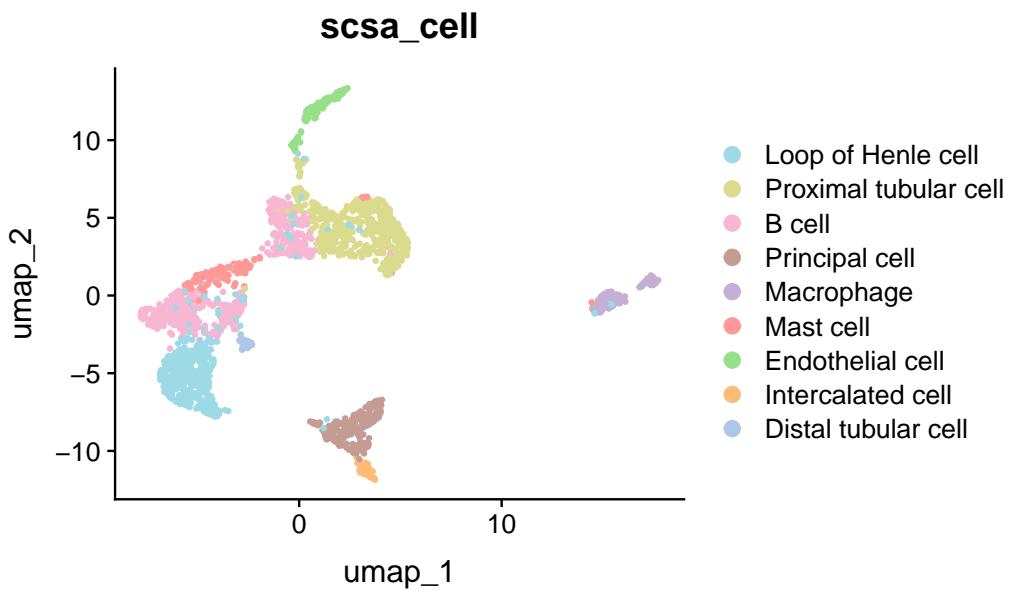


Figure 19: IgA data1 SCSA mapping

7.3.1.2 共表达分析 Figure 20为图 IgA data1 pseudotime 概览。

(对应文件为 Figure+Table/IgA-data1-pseudotime.pdf)

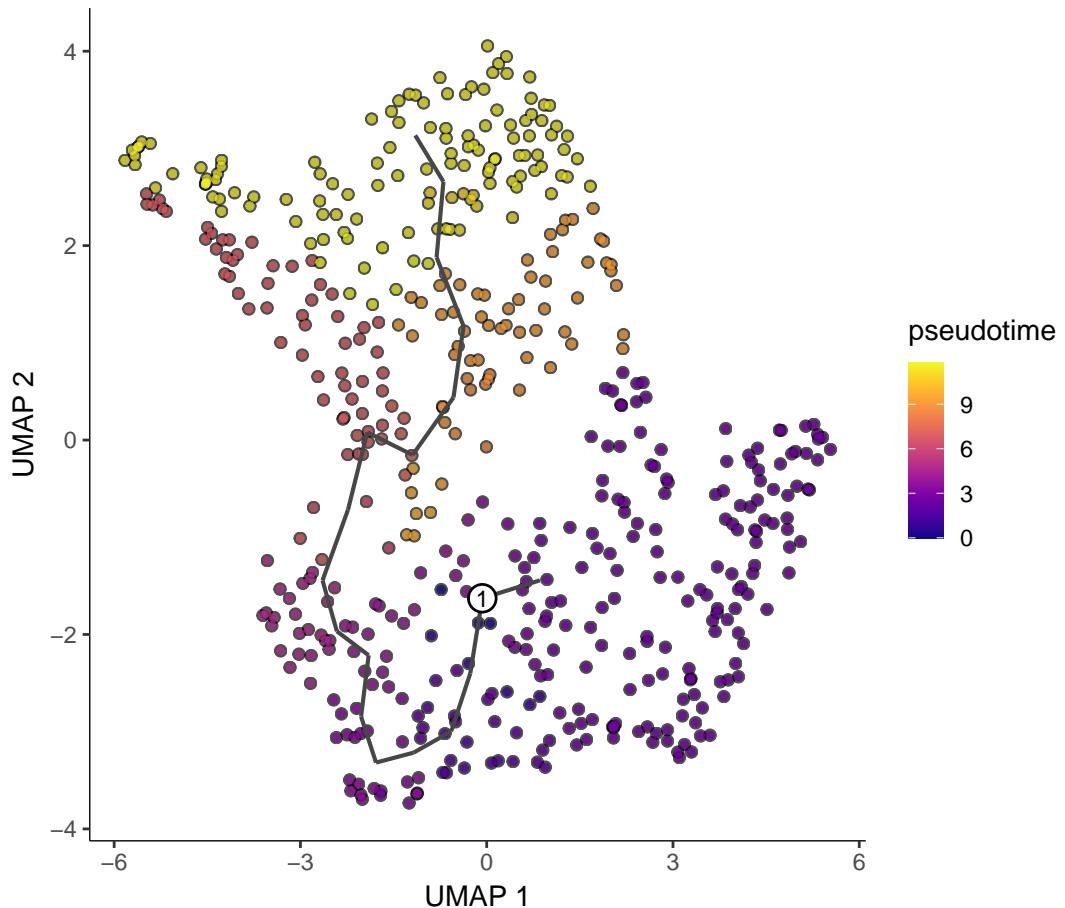


Figure 20: IgA data1 pseudotime

Figure 21为图 IgA data1 co expression modules 概览。

(对应文件为 Figure+Table/IgA-data1-co-expression-modules.pdf)

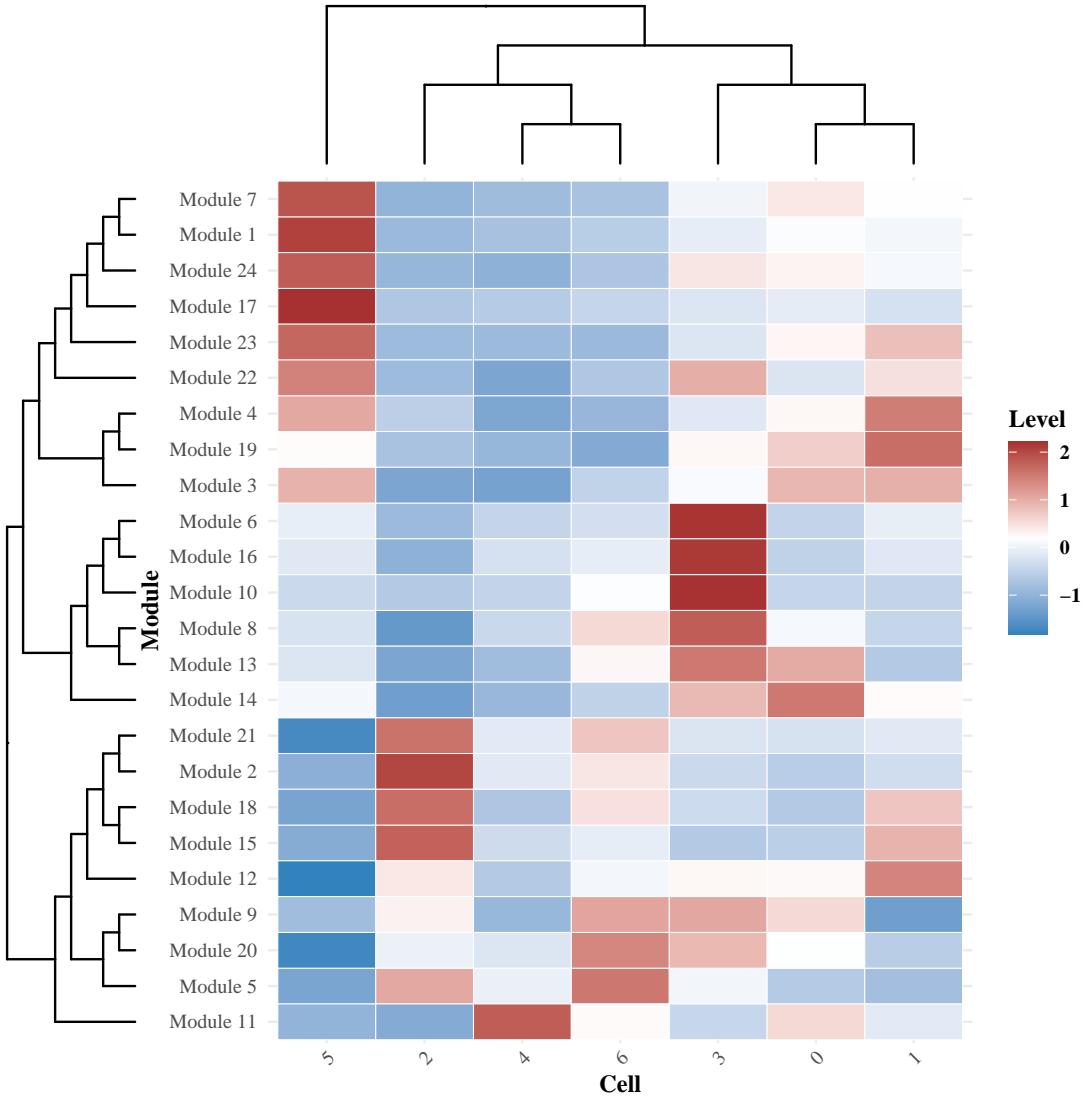


Figure 21: IgA data1 co expression modules

Figure 22为图 IgA data1 Proximal tubule cell subtype 概览。

(对应文件为 [Figure+Table/IgA-data1-Proximal-tubule-cell-subtype.pdf](#))

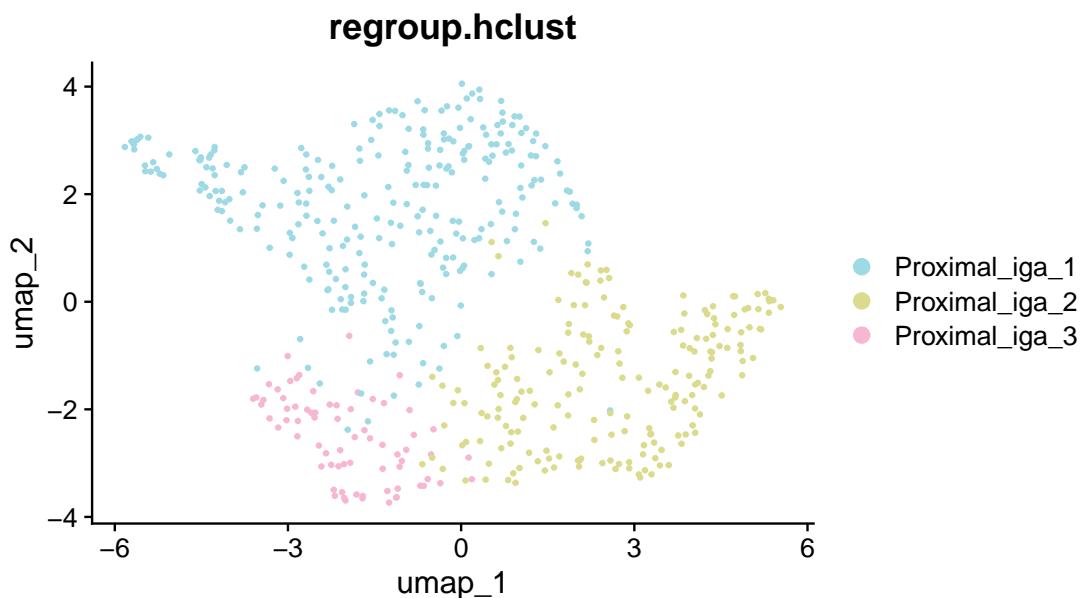


Figure 22: IgA data1 Proximal tubule cell subtype

7.3.2 膜性肾病: GSE241302 (idiopathic membranous nephropathy)

- In order to explore the molecular mechanism of IMN, we collected renal tissue samples from 3 IMN patients and 1 healthy controls and performed analysis by single-cell RNA sequencing.
- GSE241302, scRNA
- GSE216841, RNA-seq
- GSE175759, RNA-seq

7.3.2.1 细胞聚类和注释 Figure 23为图 IMN data QC 概览。

(对应文件为 Figure+Table/IMN-data-QC.pdf)

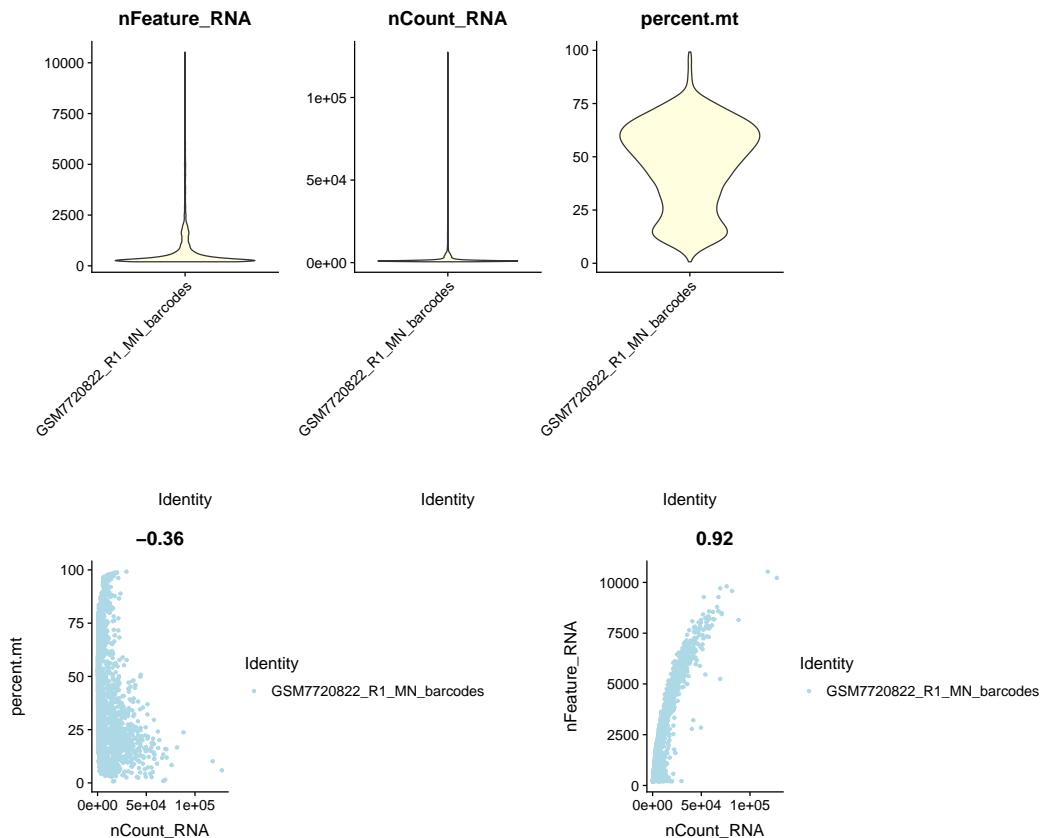


Figure 23: IMN data QC

Figure 24为图 IMN data SCSA mapping 概览。

(对应文件为 Figure+Table/IMN-data-SCSA-mapping.pdf)

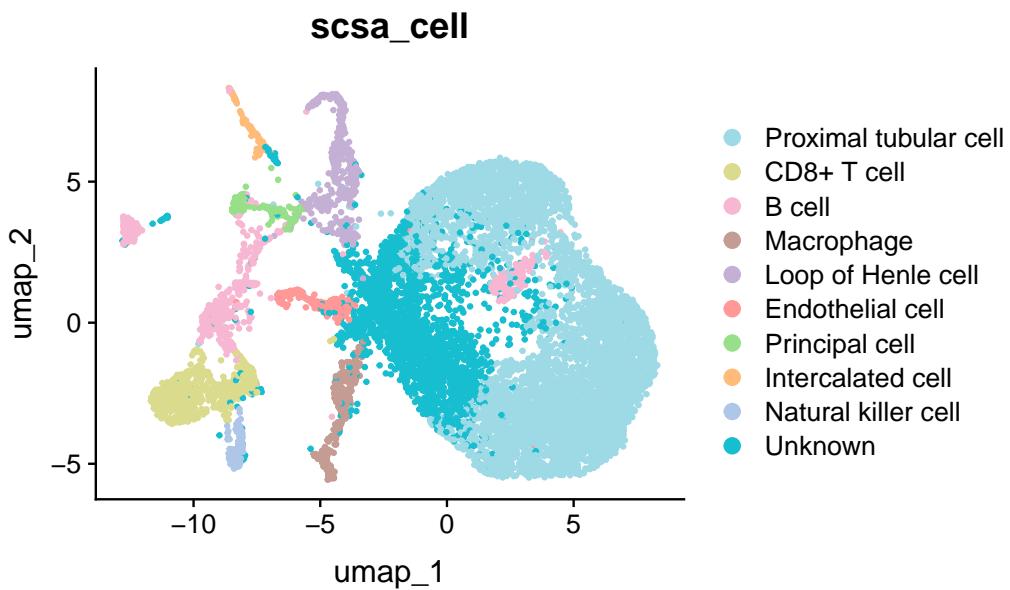


Figure 24: IMN data SCSA mapping

7.3.2.2 共表达分析 Figure 25为图 IMN data pseudotime 概览。

(对应文件为 Figure+Table/IMN-data-pseudotime.pdf)

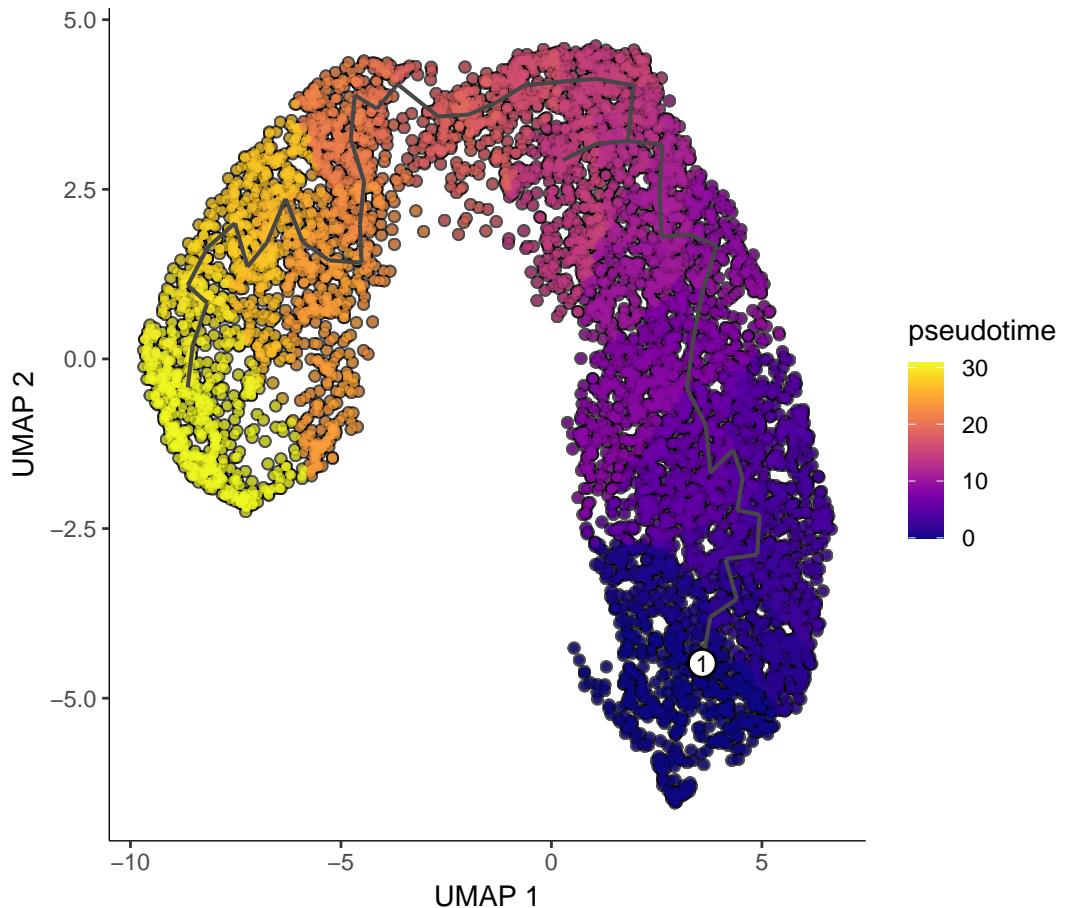


Figure 25: IMN data pseudotime

Figure 26为图 IMN data co expression modules 概览。

(对应文件为 Figure+Table/IMN-data-co-expression-modules.pdf)

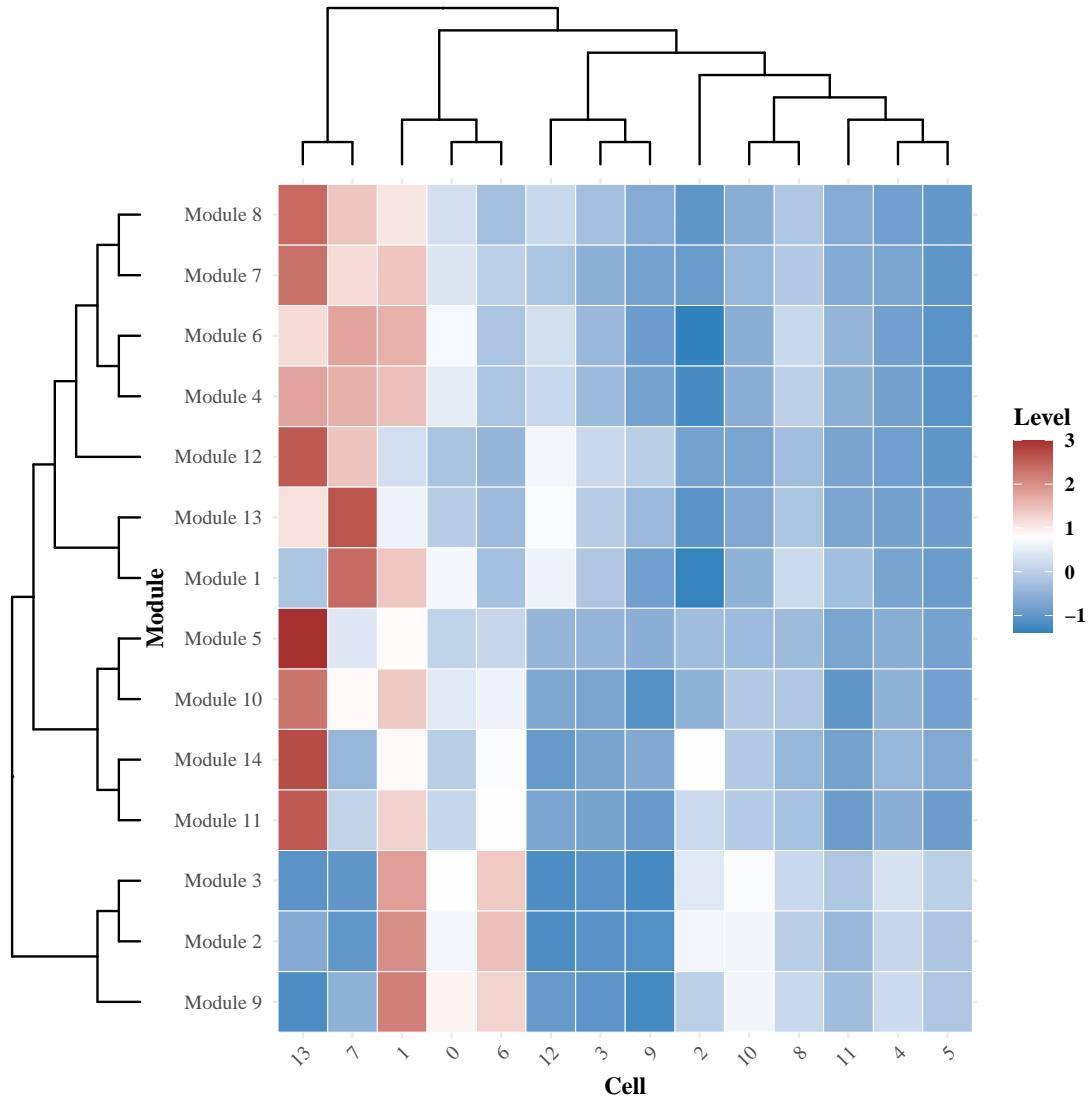


Figure 26: IMN data co expression modules

Figure 27为图 IMN data Proximal tubule cell subtype 概览。

(对应文件为 [Figure+Table/IMN-data-Proximal-tubule-cell-subtype.pdf](#))

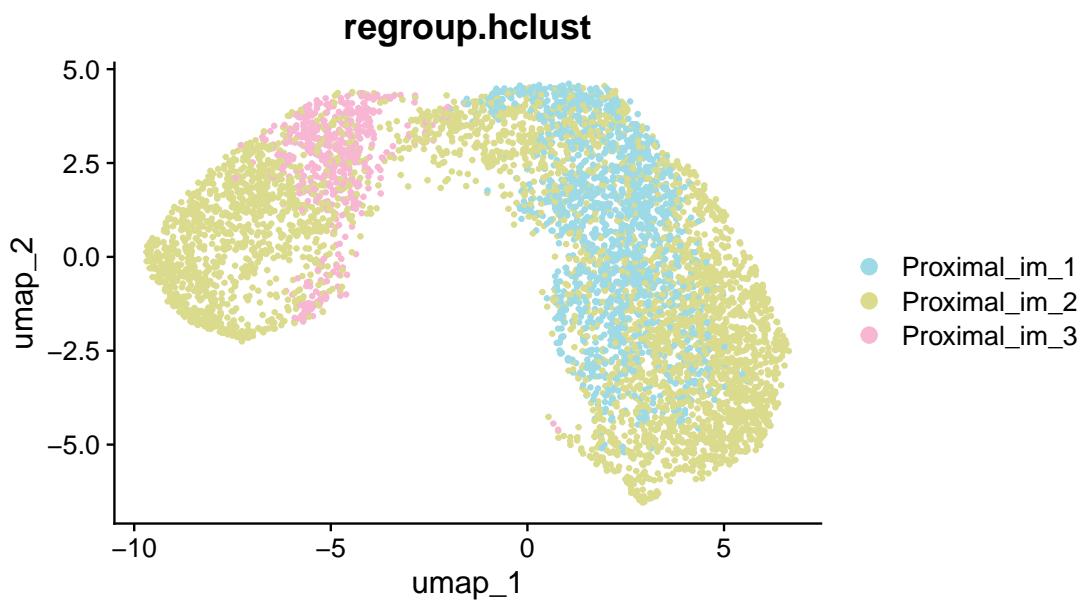


Figure 27: IMN data Proximal tubule cell subtype

7.4 正常对照 (Healthy control) (GSE171314)

7.4.0.1 细胞聚类和注释 Figure 28为图 HC data QC 概览。

(对应文件为 Figure+Table/HC-data-QC.pdf)

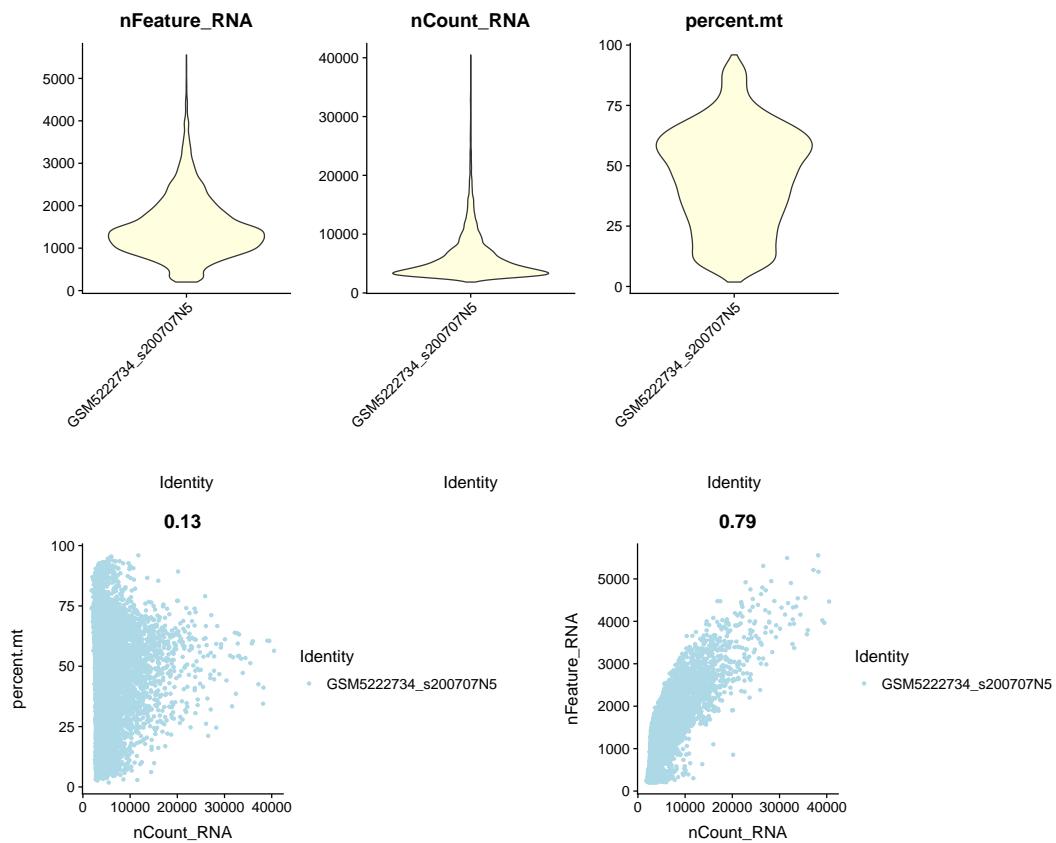


Figure 28: HC data QC

Figure 29为图 HC data SCSA mapping 概览。

(对应文件为 Figure+Table/HC-data-SCSA-mapping.pdf)

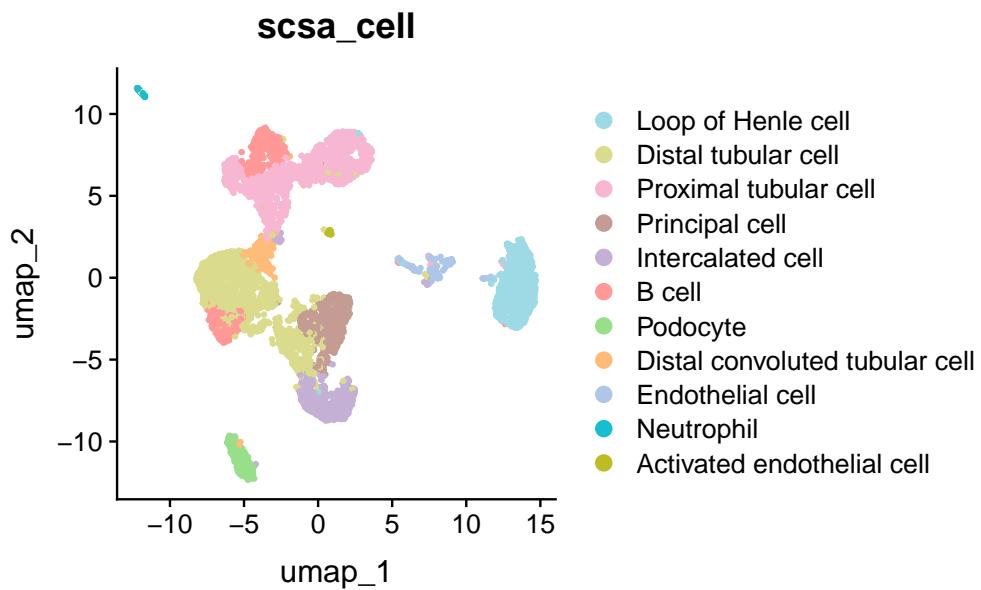


Figure 29: HC data SCSA mapping

7.5 多数据集成: ccRCC 和 CKD scRNA-seq

7.5.1 使用 RISC 对不同来源的数据集成

RISC¹⁰

Figure 30为图 SETS1 select reference dataset for integration 概览。

(对应文件为 Figure+Table/SETS1-select-reference-dataset-for-integration.pdf)

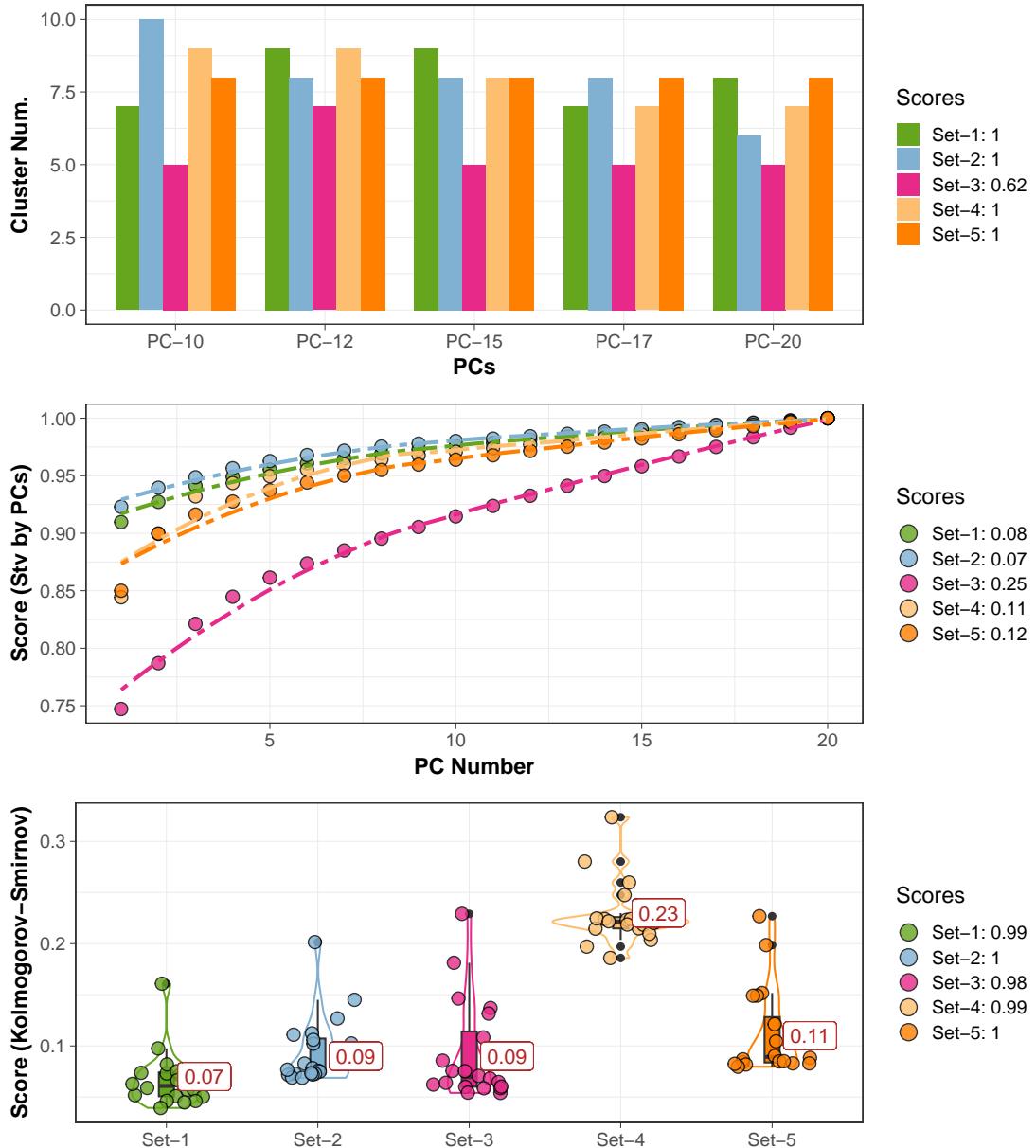


Figure 30: SETS1 select reference dataset for integration

Figure 31为图 SETS1 umap mapping 概览。

(对应文件为 Figure+Table/SETS1-umap-mapping.pdf)

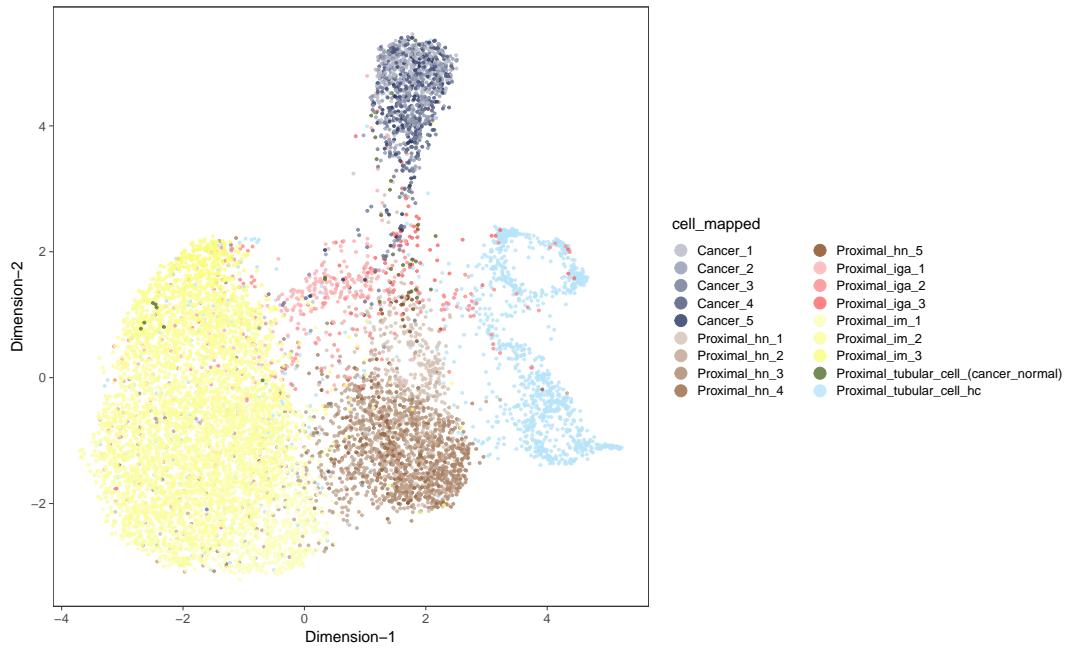


Figure 31: SETS1 umap mapping

7.5.2 拟时分析

在不同的肾病聚类中，IgA 与 RCC 最为接近。IgA 肾病可能发展成 RCC。

Figure 32为图 SETS1 trajectory 概览。

(对应文件为 Figure+Table/SETS1-trajectory.pdf)

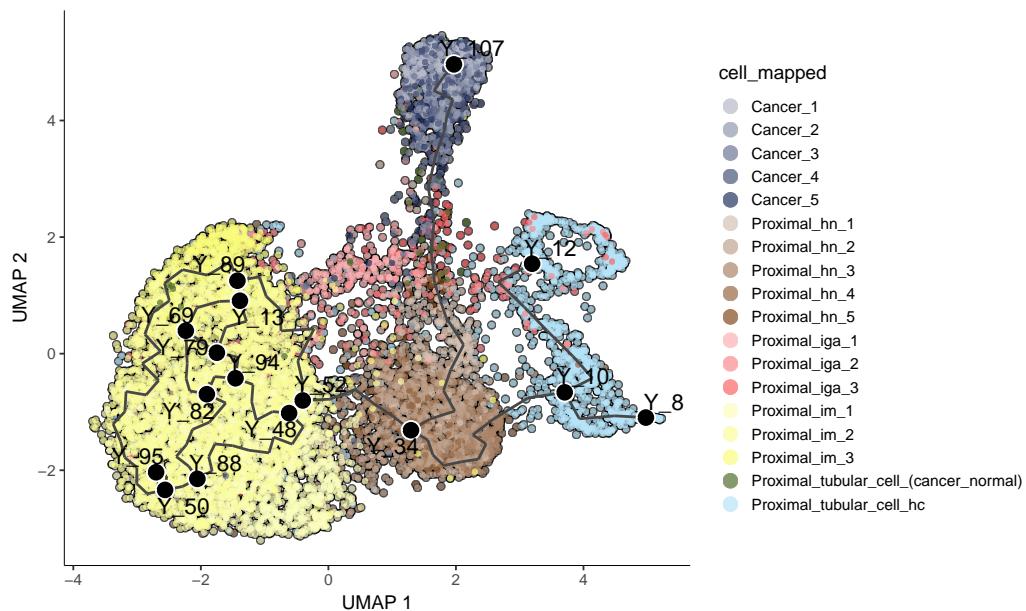


Figure 32: SETS1 trajectory

Figure 33为图 SETS1 pseudotime 概览。

(对应文件为 Figure+Table/SETS1-pseudotime.pdf)

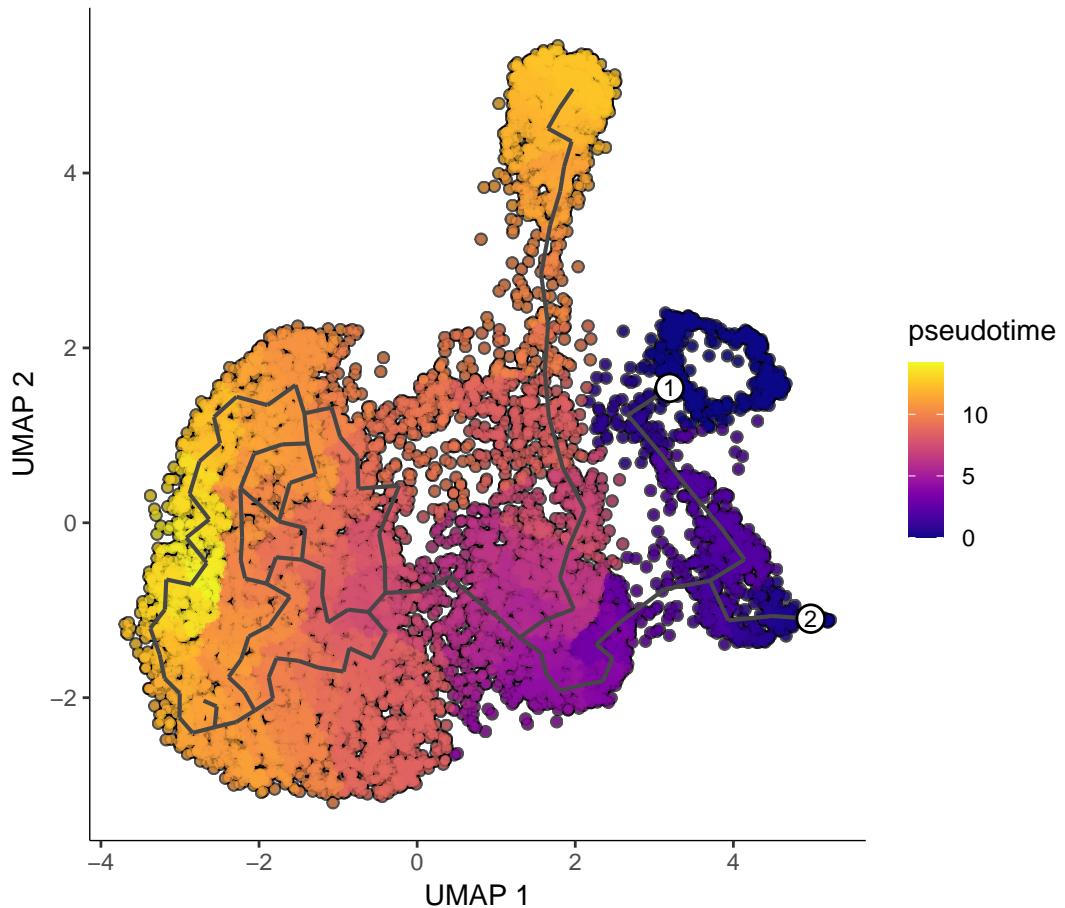


Figure 33: SETS1 pseudotime

7.5.3 IgA、Cancer、HC 差异分析

Figure 34为图 SETS1 intersection of contrasts DEGs 概览。

(对应文件为 Figure+Table/SETS1-intersection-of-contrasts-DEGs.pdf)

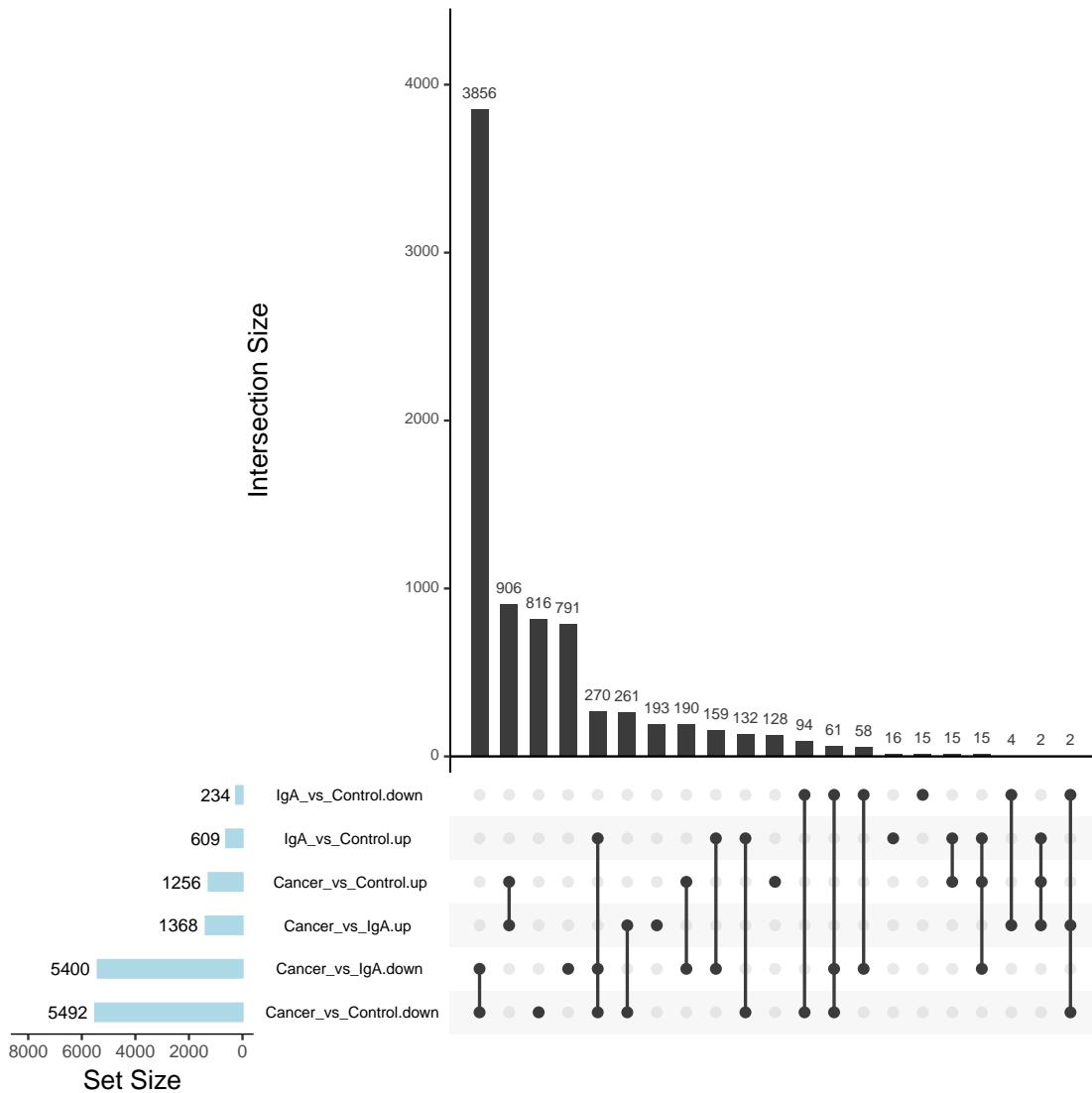


Figure 34: SETS1 intersection of contrasts DEGs

Table 1为表格 SETS1 DEGs 概览。

(对应文件为 [Figure+Table/SETS1-DEGs.csv](#))

注：表格共有 14359 行 7 列，以下预览的表格可能省略部分数据；表格含有 3 个唯一‘contrast’。

Table 1: SETS1 DEGs

contrast	p_val	avg_l...	pct.1	pct.2	p_val...	gene
Cance...	0	-11.8...	0	1	0	MT-RNR1
Cance...	0	-8.47...	0	0.809	0	FABP1
Cance...	0	-11.9...	0	0.806	0	MTATP6P1
Cance...	0	-12.0...	0	0.988	0	ALDOB

contrast	p_val	avg_l...	pct.1	pct.2	p_val...	gene
Cance...	0	-3.24...	0.97	0.056	0	ANXA2
Cance...	0	-9.49...	0	0.839	0	SLC5A12
Cance...	0	-4.24...	1	0.075	0	VIM
Cance...	0	-1.19...	0.928	0.015	0	PFKP
Cance...	0	-0.97...	0.91	0.014	0	PERP
Cance...	0	-2.67...	0.933	0.03	0	LGALS1
Cance...	0	-1.90...	0.887	0.029	0	BST2
Cance...	0	-1.09...	0.848	0.017	0	CYTOR
Cance...	0	-2.57...	0.981	0.023	0	TIMP1
Cance...	0	0.906...	0.784	0.007	0	MT3
Cance...	0	-1.71...	0.859	0.022	0	AHNAK
...

7.5.4 富集分析 (GSEA)

7.5.4.1 Cancer vs IgA Figure 35为图 SETS1 GSEA enrichment of KEGG CancerVsIgA 概览。

(对应文件为 Figure+Table/SETS1-GSEA-enrichment-of-KEGG-CancerVsIgA.pdf)

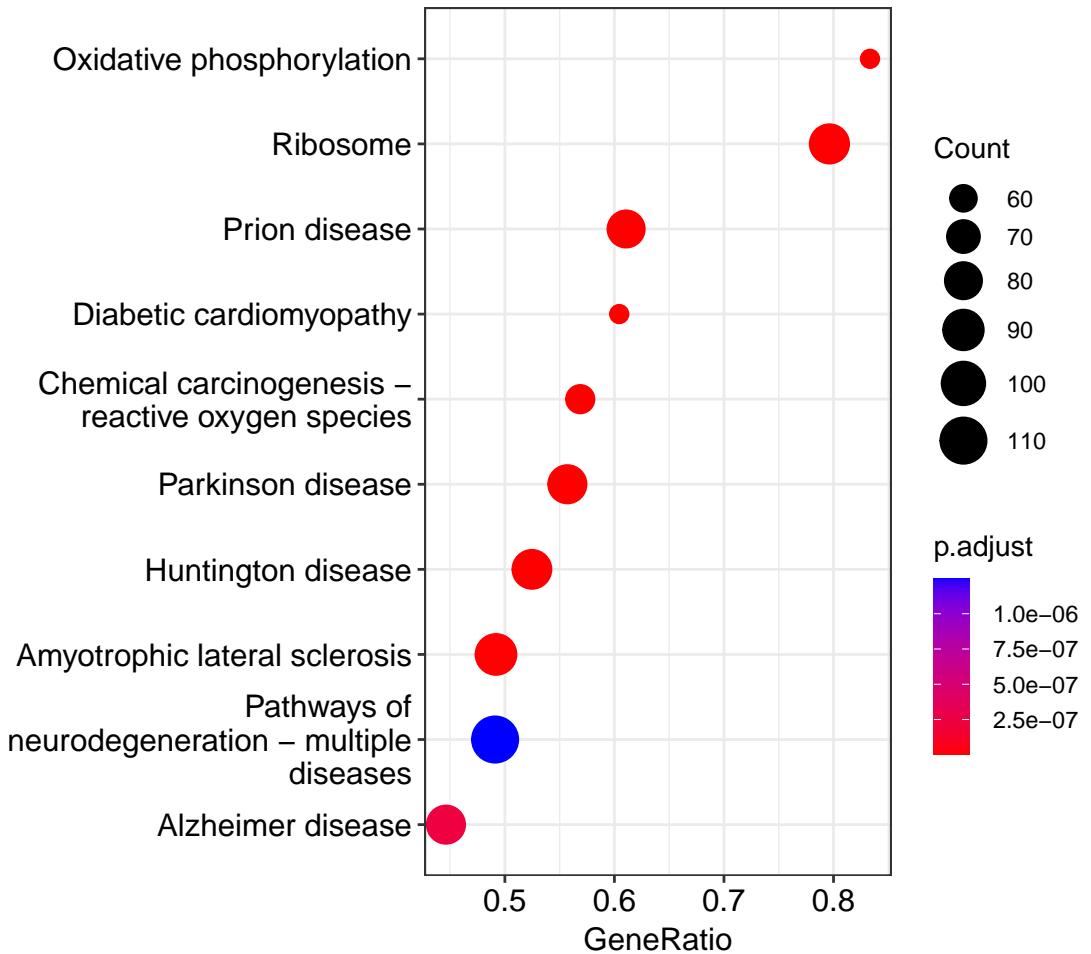


Figure 35: SETS1 GSEA enrichment of KEGG CancerVsIgA

Figure 36为图 SETS1 GSEA enrichment of GO CancerVsIgA 概览。

(对应文件为 Figure+Table/SETS1-GSEA-enrichment-of-GO-CancerVsIgA.pdf)

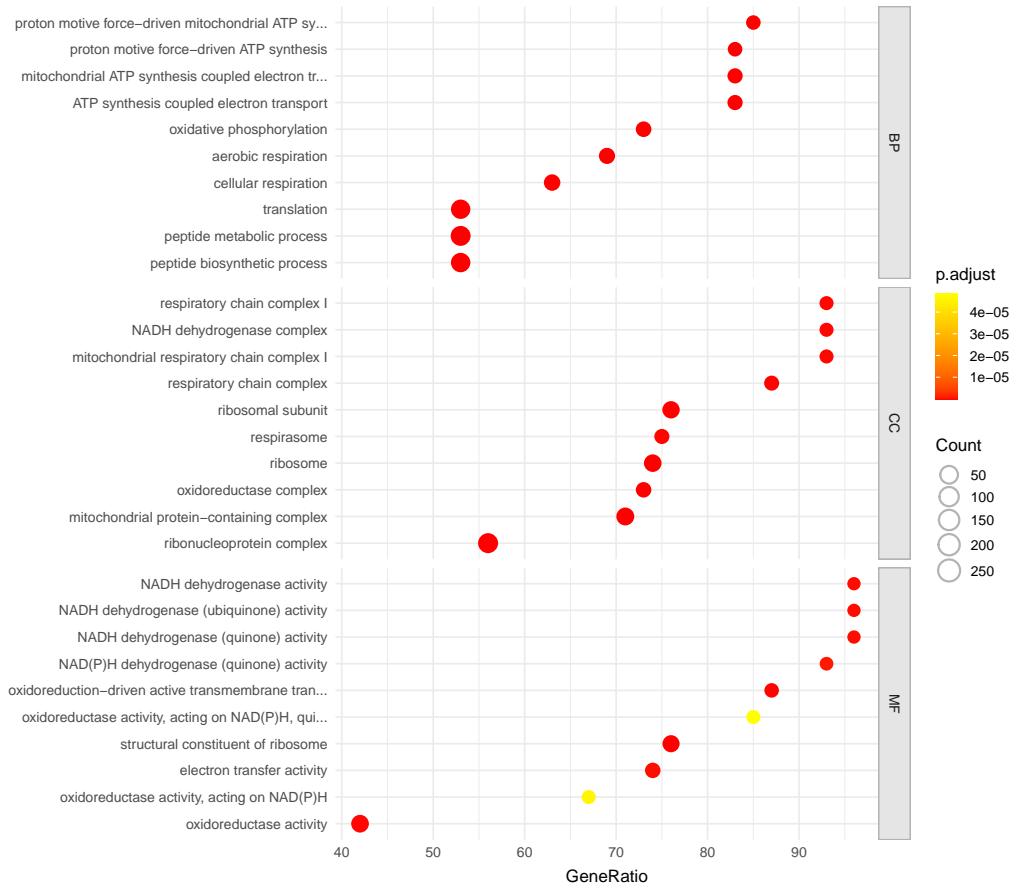


Figure 36: SETS1 GSEA enrichment of GO CancerVsIgA

Fig. ?? 富集于 OXPHOS 通路的基因表达量整体下降。

Figure 37为图 SETS1 GSEA show OXPHOS pathway CancerVsIgA 概览。

(对应文件为 Figure+Table/SETS1-GSEA-show-OXPHOS-pathway-CancerVsIgA.pdf)

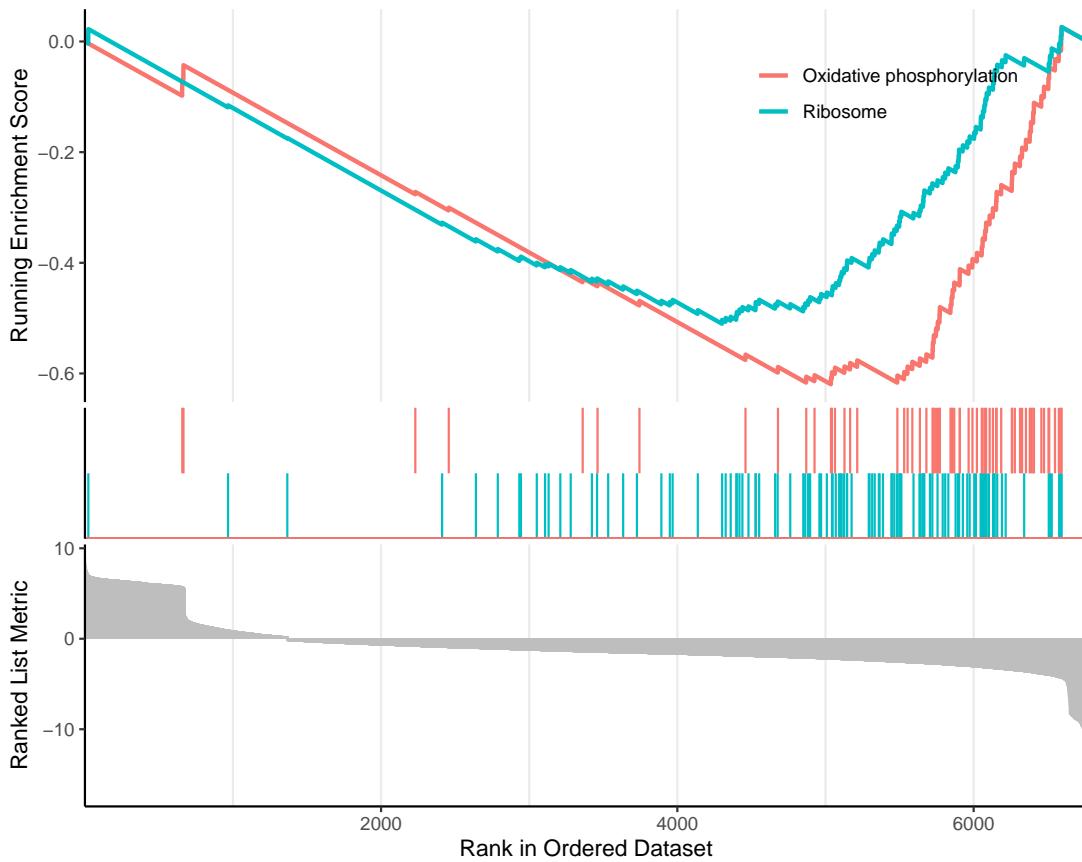


Figure 37: SETS1 GSEA show OXPHOS pathway CancerVsIgA

Figure 38为图 SETS1 GSEA show OXPHOS pathway by pathview CancerVsIgA 概览。

(对应文件为 Figure+Table/hsa00190.pathview.png)

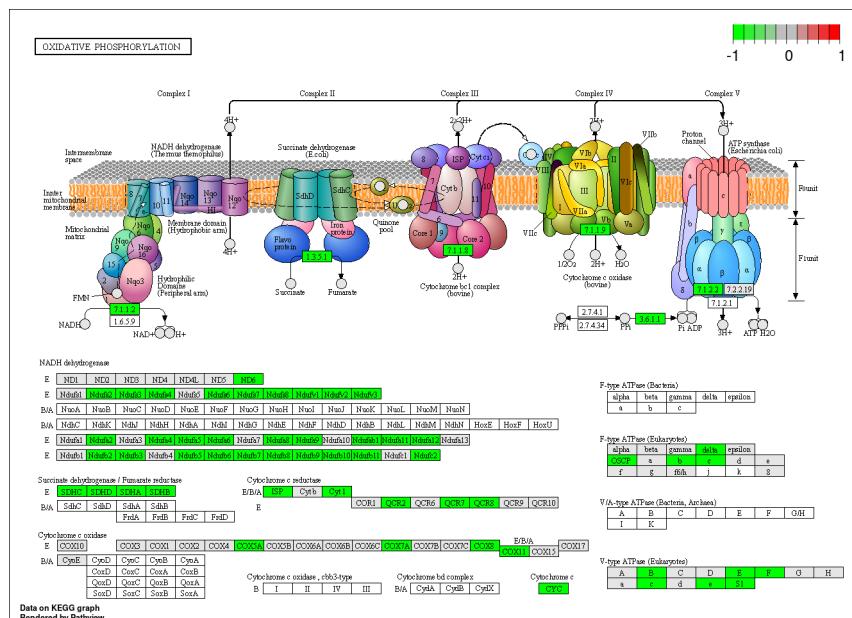


Figure 38: SETS1 GSEA show OXPHOS pathway by pathview CancerVsIgA

7.5.4.2 Cancer vs Control Fig. 39 与 Fig. 35 相比，并不富集于 OXPHOS 通路。

Figure 39为图 SETS1 GSEA enrichment of KEGG CancerVsControl 概览。

(对应文件为 Figure+Table/SETS1-GSEA-enrichment-of-KEGG-CancerVsControl.pdf)

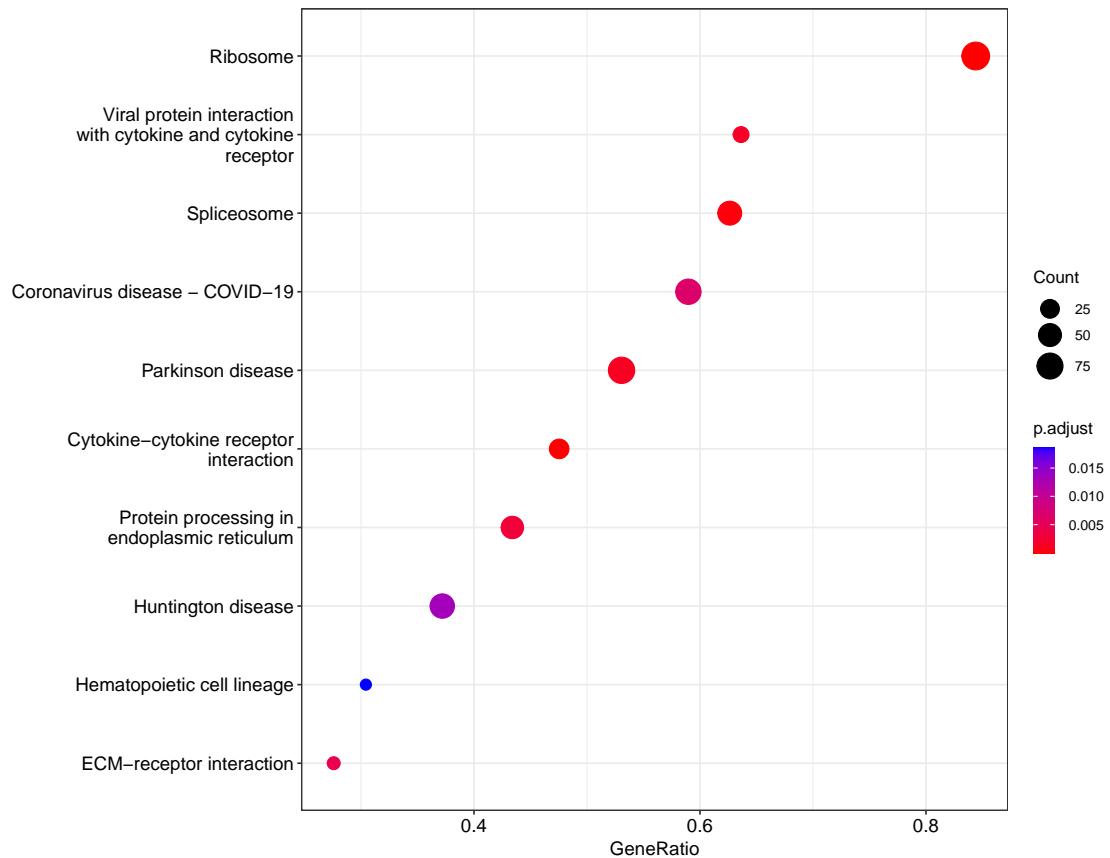


Figure 39: SETS1 GSEA enrichment of KEGG CancerVsControl

Figure 40为图 SETS1 GSEA enrichment of GO CancerVsControl 概览。

(对应文件为 Figure+Table/SETS1-GSEA-enrichment-of-GO-CancerVsControl.pdf)

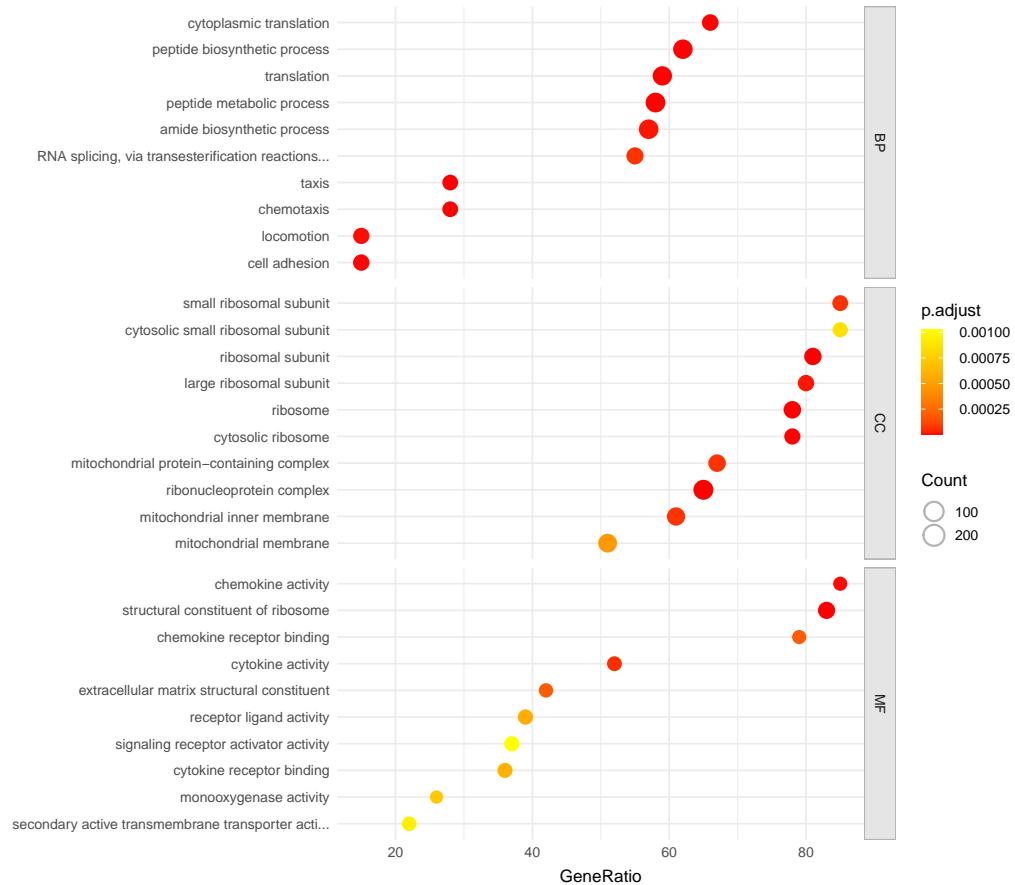


Figure 40: SETS1 GSEA enrichment of GO CancerVsControl

7.6 验证

7.6.1 另一组 IgA 肾病: GSE127136

7.6.1.1 细胞聚类和注释

7.6.2 另一组 RCC: GSE202374

7.6.2.1 细胞聚类和注释 同 7.1, 使用 SCSA 注释后, 以 copyKAT 预测癌细胞, 随后将肿瘤细胞映射到 UMAP 聚类图中。

Figure 41为图 RCC data2 SCSA mapping with copykat prediction 概览。

(对应文件为 Figure+Table/RCC-data2-SCSA-mapping-with-copykat-prediction.pdf)

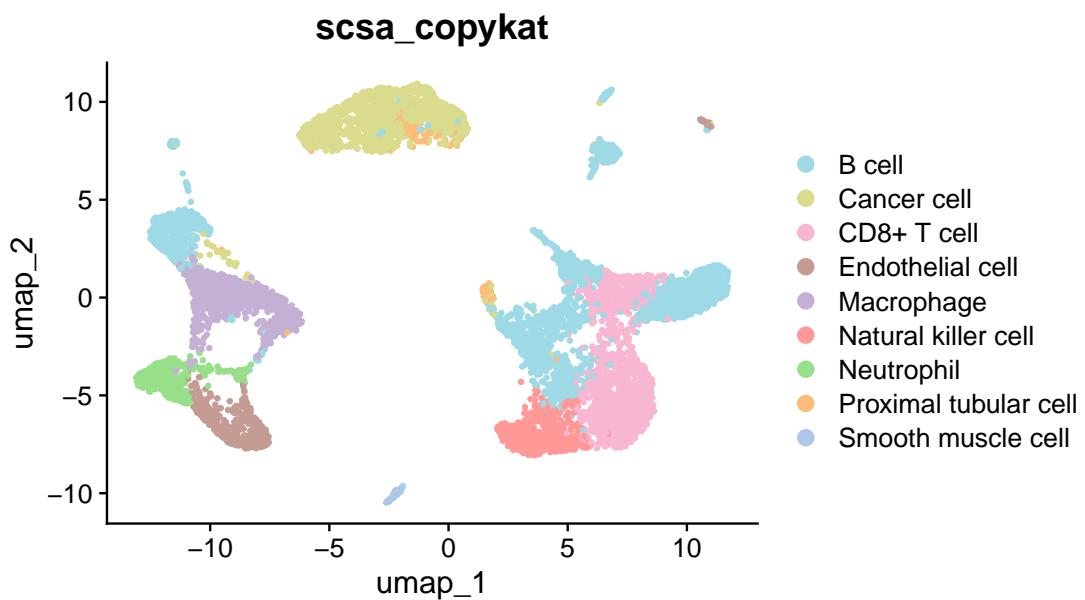


Figure 41: RCC data2 SCSA mapping with copykat prediction

7.6.3 集成验证: Cancer、IgA

7.6.3.1 RISC 集成 Figure 42为图 SETS2 select reference dataset for integration 概览。

(对应文件为 Figure+Table/SETS2-select-reference-dataset-for-integration.pdf)

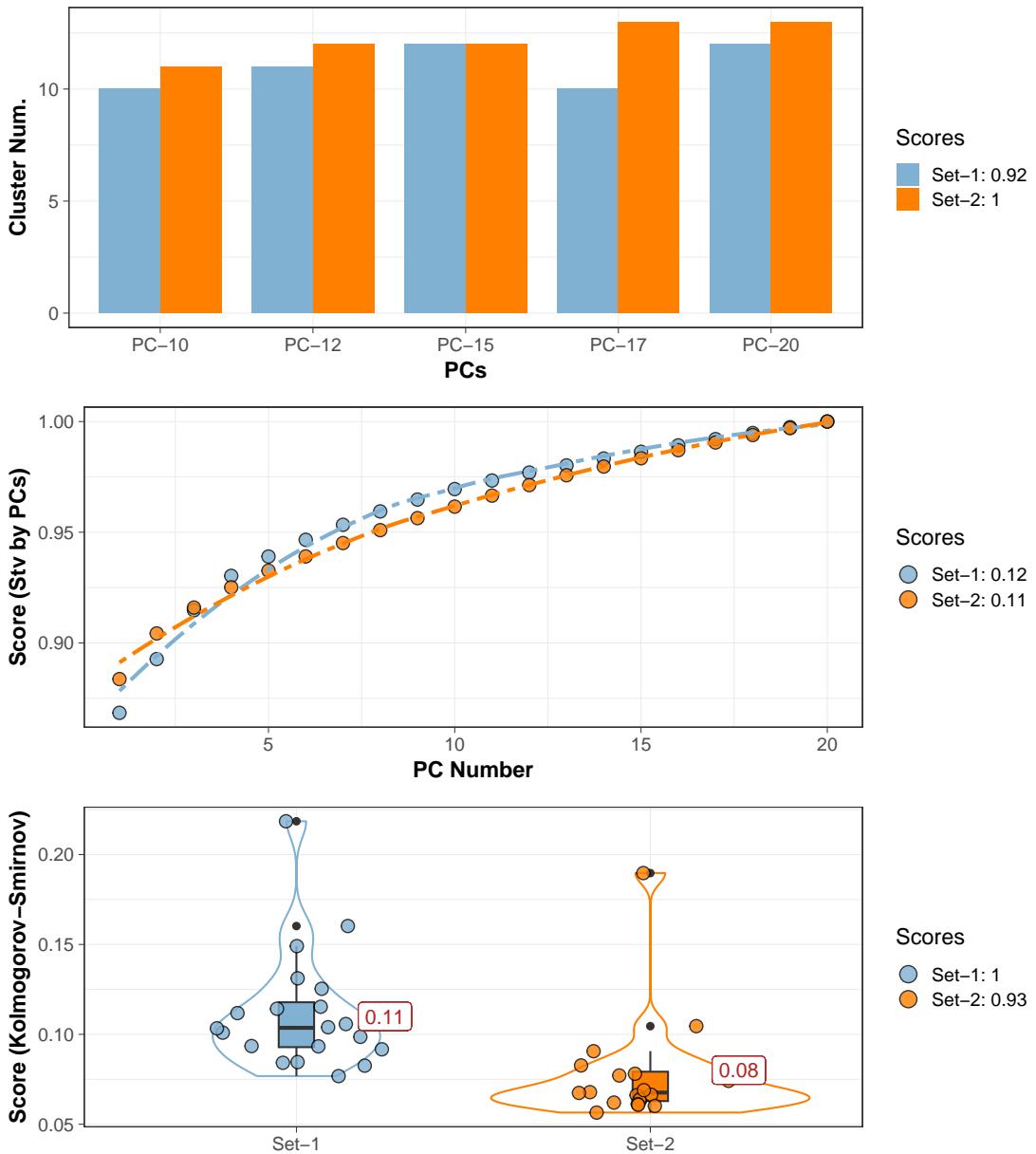


Figure 42: SETS2 select reference dataset for integration

Figure 43为图 SETS2 umap mapping 概览。

(对应文件为 Figure+Table/SETS2-umap-mapping.pdf)

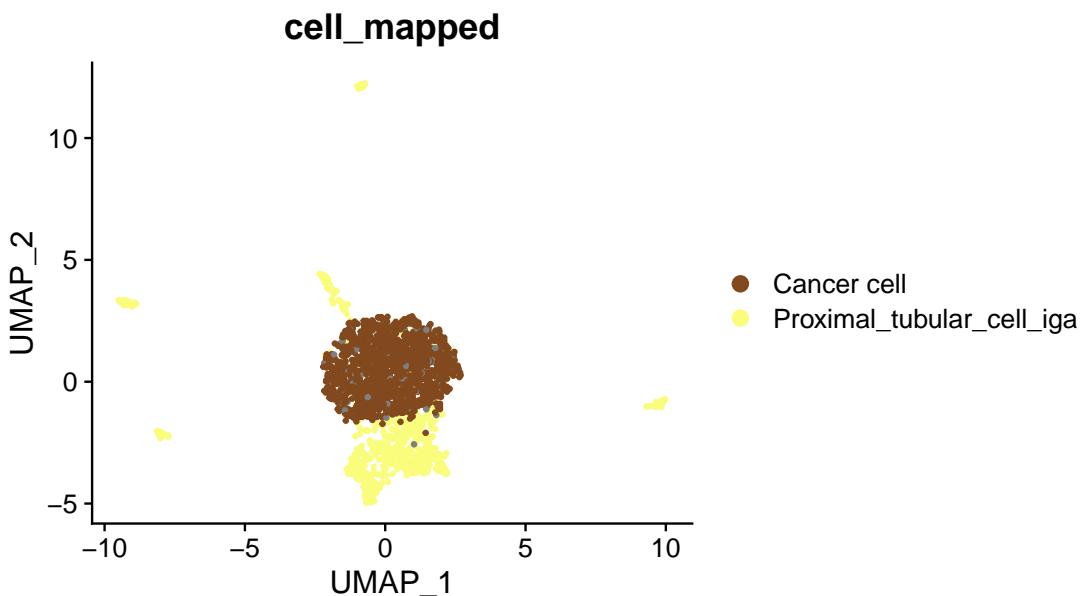


Figure 43: SETS2 umap mapping

7.6.3.2 拟时分析 依据拟时轨迹 Fig. 33, 同 Fig. 32相似, IgA 与 RCC 聚类邻近, 可能在疾病发展过程中转化为癌症。

Figure 44为图 SETS2 pseudotime 概览。

(对应文件为 Figure+Table/SETS2-pseudotime.pdf)

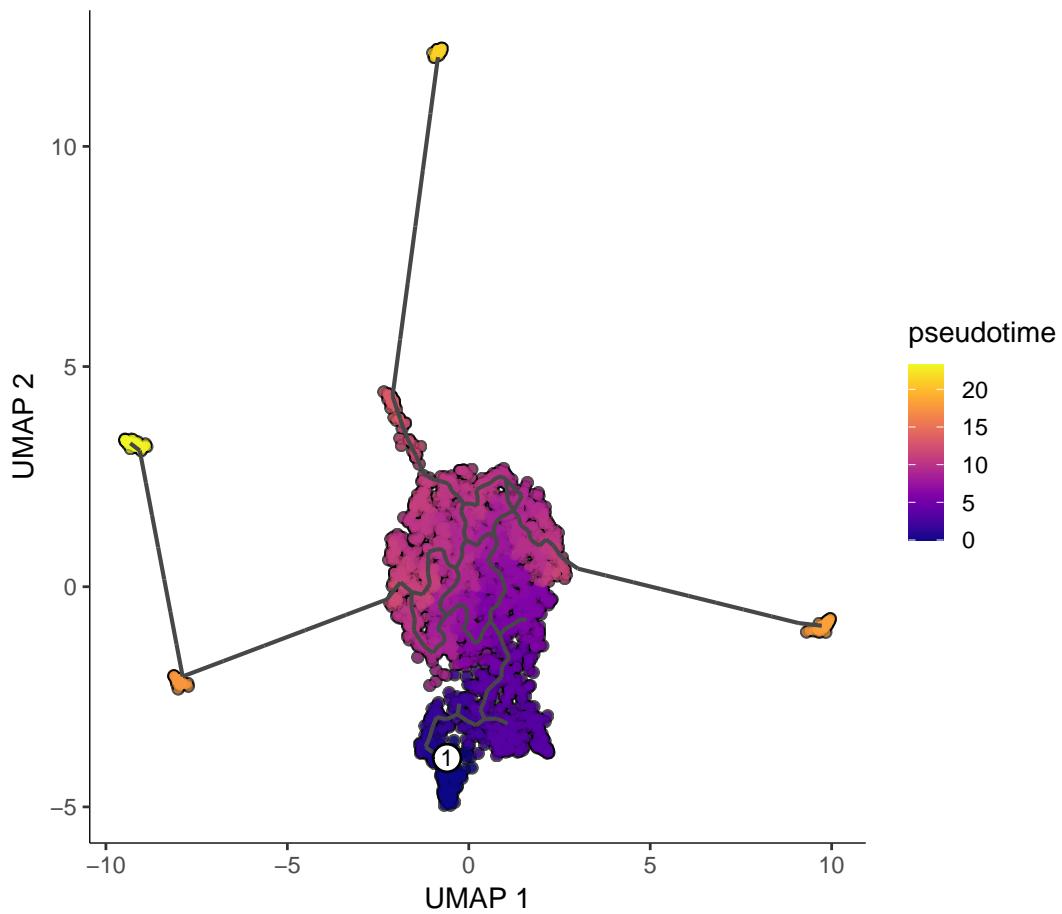


Figure 44: SETS2 pseudotime

7.6.3.3 富集分析 Figure 45为图 SETS2 GSEA enrichment of KEGG CancerVsIgA 概览。
(对应文件为 Figure+Table/SETS2-GSEA-enrichment-of-KEGG-CancerVsIgA.pdf)

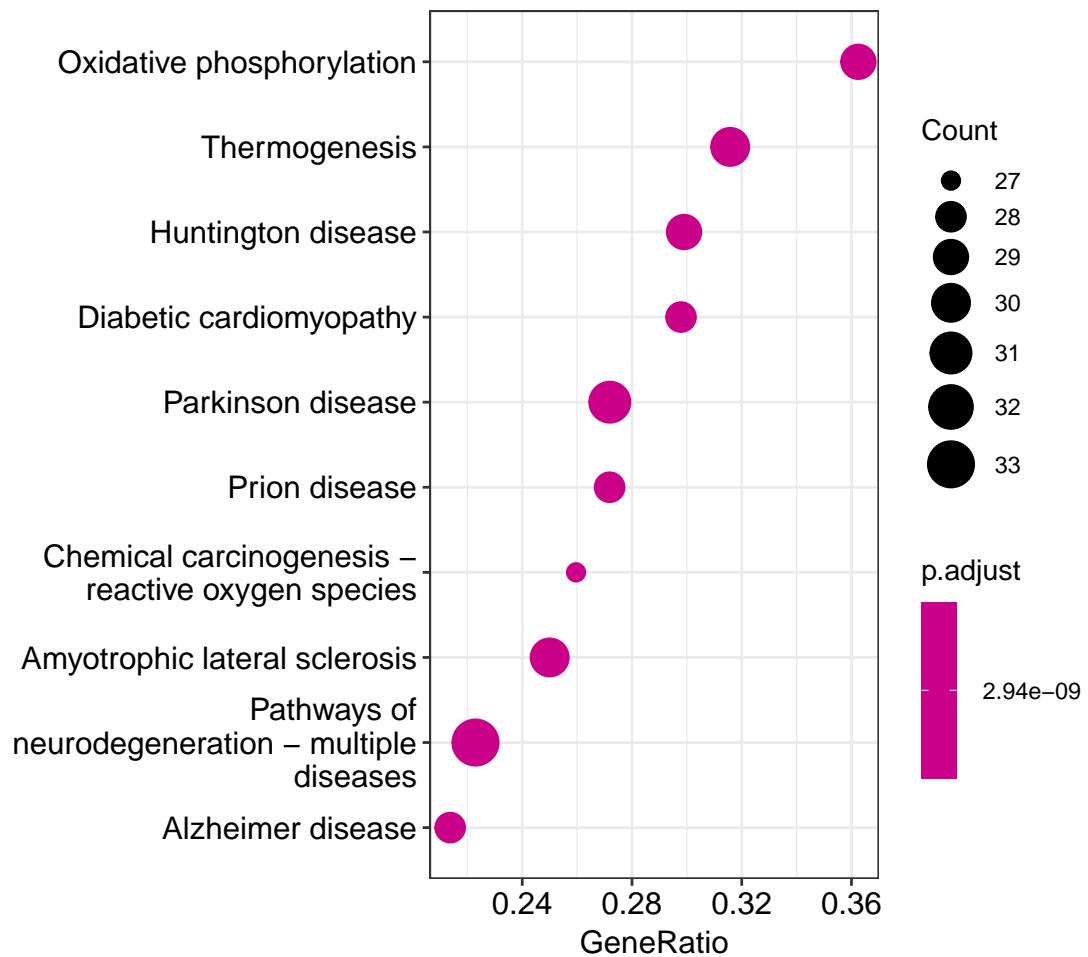


Figure 45: SETS2 GSEA enrichment of KEGG CancerVsIgA

Figure 46为图 SETS2 GSEA enrichment of GO CancerVsIgA 概览。

(对应文件为 Figure+Table/SETS2-GSEA-enrichment-of-GO-CancerVsIgA.pdf)

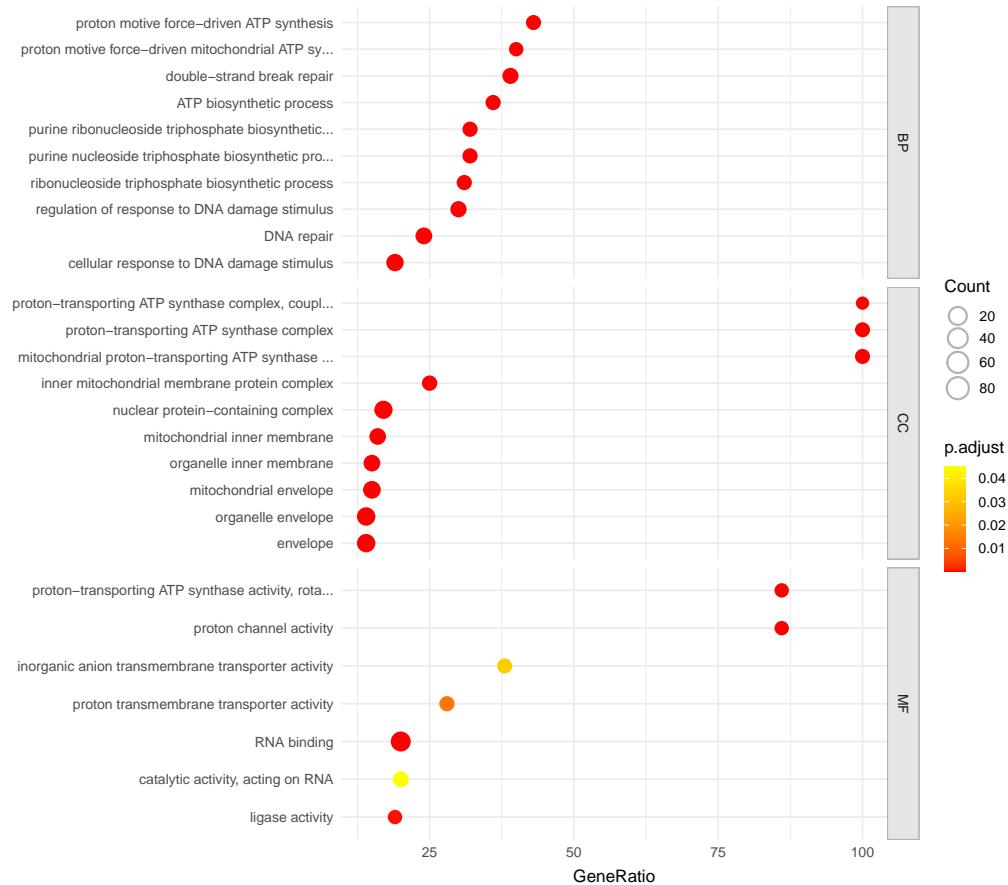


Figure 46: SETS2 GSEA enrichment of GO CancerVsIgA

Fig. ?? 富集于 OXPHOS 通路的基因表达量整体上升（与 Fig. ?? 相反）。系肿瘤细胞的异质性所致¹³，RCC 肿瘤可分为 OXPHOS 依赖型和非依赖型代谢增强。

Figure 47为图 SETS2 GSEA show OXPHOS pathway CancerVsIgA 概览。

(对应文件为 Figure+Table/SETS2-GSEA-show-OXPHOS-pathway-CancerVsIgA.pdf)

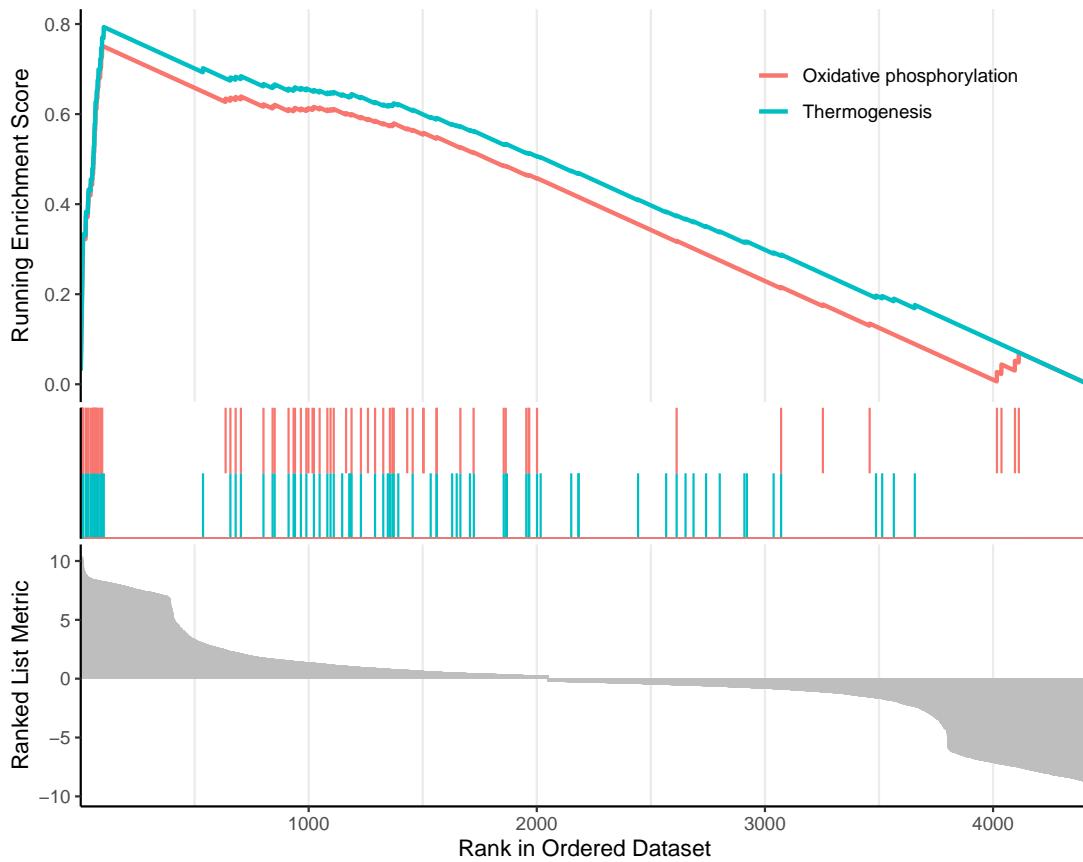


Figure 47: SETS2 GSEA show OXPHOS pathway CancerVsIgA

Figure 48为图 SETS2 GSEA show OXPHOS pathway by pathview CancerVsIgA 概览。

(对应文件为 Figure+Table/hsa00190.pathview.png)

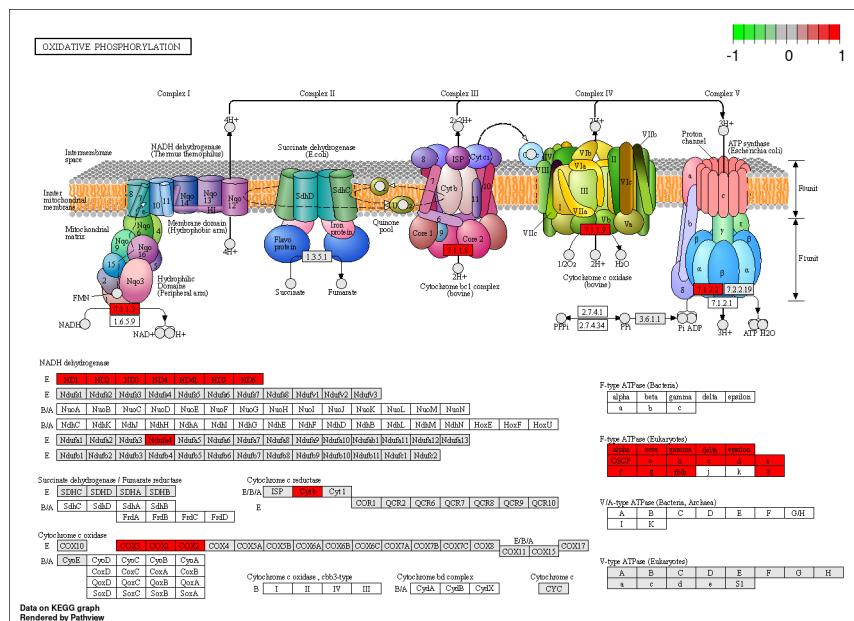


Figure 48: SETS2 GSEA show OXPHOS pathway by pathview CancerVsIgA

Reference

1. Rosner, M. H., Jhaveri, K. D., McMahon, B. A. & Perazella, M. A. Onconephrology: The intersections between the kidney and cancer. *CA: a cancer journal for clinicians* **71**, 47–77 (2021).
2. Lowrance, W. T., Ordoñez, J., Udaltssova, N., Russo, P. & Go, A. S. CKD and the risk of incident cancer. *Journal of the American Society of Nephrology* **25**, (2014).
3. Kitchlu, A. *et al.* Cancer risk and mortality in patients with kidney disease: A population-based cohort study. *American journal of kidney diseases : the official journal of the National Kidney Foundation* **80**, 436–448.e1 (2022).
4. Saly, D. L., Eswarappa, M. S., Street, S. E. & Deshpande, P. Renal cell cancer and chronic kidney disease. *Advances in chronic kidney disease* **28**, 460–468.e1 (2021).
5. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, (2021).
6. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, (2019).
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. Gao, R. *et al.* Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nature Biotechnology* **39**, 599–608 (2021).
10. Liu, Y., Wang, T., Zhou, B. & Zheng, D. Robust integration of multiple single-cell rna sequencing datasets using a single reference space. *Nature biotechnology* **39**, 877–884 (2021).
11. Wu, T. *et al.* ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, (2021).
12. Cao, Y., Wang, X. & Peng, G. SCSA: A cell type annotation tool for single-cell rna-seq data. *Frontiers in genetics* **11**, (2020).
13. Sancho, P., Barneda, D. & Heeschen, C. Hallmarks of cancer stem cell metabolism. *British journal of cancer* **114**, 1305–1312 (2016).