

Analysis

Huang LiChuang of Wie-Biotech

Contents

1 题目	3
2 摘要	4
3 前言	4
4 材料和方法	4
4.1 材料	4
4.2 方法	5
5 分析结果	5
5.1 RCC 数据集分析鉴定	5
5.2 不同类型 CKD 数据共表达分析和重聚类	6
5.3 RCC 和 CKD 数据集成	7
5.4 差异分析和富集分析	8
5.5 以新的 IgA 和 RCC 数据集验证	9
6 结论	11
7 附：分析流程	11
7.1 肾癌 (GEO)	11
7.1.1 GSE171306	11
7.1.1.1 细胞聚类和注释	11
7.1.1.2 癌细胞识别	12
7.1.1.3 癌细胞拟时分析和共表达分析	14
7.2 血管性疾患	18
7.2.1 高血压性肾炎: GSE210898 (hypertensive nephropathy)	18
7.2.1.1 细胞聚类和注释	18
7.2.1.2 共表达分析	20
7.3 (原发性) 肾小球性疾患	22
7.3.1 IgA 肾病: GSE171314 (IgA Nephropathy)	22
7.3.1.1 细胞聚类和注释	22
7.3.1.2 共表达分析	24

7.3.2 膜性肾病: GSE241302 (idiopathic membranous nephropathy)	26
7.3.2.1 细胞聚类和注释	26
7.3.2.2 共表达分析	28
7.4 正常对照 (Healthy control) (GSE171314)	30
7.4.0.1 细胞聚类和注释	30
7.5 多数据集成: ccRCC 和 CKD scRNA-seq	32
7.5.1 使用 RISC 对不同来源的数据集成	32
7.5.2 拟时分析	33
7.5.3 IgA、Cancer、HC 差异分析	34
7.5.4 富集分析 (GSEA)	36
7.5.4.1 Cancer vs IgA	36
7.5.4.2 Cancer vs Control	40
7.6 验证	41
7.6.1 另一组 IgA 肾病: GSE127136	41
7.6.1.1 细胞聚类和注释	41
7.6.2 另一组 RCC: GSE202374	41
7.6.2.1 细胞聚类和注释	41
7.6.3 集成验证: Cancer、IgA	42
7.6.3.1 RISC 集成	42
7.6.3.2 拟时分析	44
7.6.3.3 富集分析	45
Reference	49

List of Figures

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

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

List of Tables

1	SETS1 DEGs	35
---	----------------------	----

1 题目

集成多组 CKD 单细胞数据集发现 IgA 相比于其它 CKD 的与 RCC 之间转化的更高风险

2 摘要

为了探究慢性肾病（CKD）与癌症之间的联系，本研究以 HN、IMN、IgA 类型的 CKD 为例，分析其和 RCC 之间的转化可能和差异关系。为了克服不同来源数据的分析困难（多疾病关联分析往往面临的难题），本研究选用 RISC 集成不同来源的疾病数据。随后，以拟时分析的方式来探究不同疾病之间的近似关系和转化可能。创新性的探究了多种 CKD 疾病和 RCC 癌症之间关系。随后发现 IgA 可能是向 RCC 转化风险更高的一个 CKD 疾病（结论总结见 6）。

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

4.1 材料

All used GEO expression data and their design:

- **GSE210898:** We analyzed kidney samples from 3 patients with HTN using single-cell RNA sequencing, compared with previous data of controls. We then identified distinct cell clusters through unsupervised cluster...
- **GSE171314:** 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.
- **GSE127136:** In this study, we collected single cells from 13 IgAN patients' renal biopsies and normal renal cells from 6 kidney cancer patients' paracancerous tissues. As glomerulus are difficult to digest, we...
- **GSE241302:** 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.
- **GSE202374:** Single cell comparisons of radiotherapy treated resected RCC tumors and nephrectomy only controls.
- **GSE171306:** Primary kidney samples from 3 patients were used for single cell RNA sequencing by 10X Genomics

4.2 方法

Mainly used method:

- GEO <https://www.ncbi.nlm.nih.gov/geo/> used for expression dataset acquisition .
- Seurat used for scRNA-seq processing.^{5,6}
- Monocle3 used for cell pseudotime analysis.^{7,8}
- Package copyKAT used for aneuploid cell or cancer cell prediction.⁹
- RISC used for scRNA-seq data integration.¹⁰
- ClusterProfiler used for gene enrichment analysis.¹¹
- Other R packages used for statistic analysis or data visualization.

5 分析结果

5.1 RCC 数据集分析鉴定

RCC 数据集 (RCC1) 以 Seurat 预处理、质控、聚类后，以 SCSA 注释细胞类型 (Fig. 1a)¹²，随后以 copyKAT 预测肿瘤细胞⁹ (Fig. 1b)。约 90% 的肿瘤细胞为非整倍体¹³，而人类为二倍体生物，因此可以通过 copyKAT 的非整倍体预测判断肿瘤细胞⁹。对于 Fig. 1a，由于 SCSA 无法注释肿瘤细胞，而 RCC 是一种源于上皮细胞的癌症，非 B 细胞癌，因此，Fig. 1a 中与 Proximal tubular cells (PTCs) 相邻，但在随后被 copyKAT 预测为肿瘤细胞的 B 细胞应是 PTCs 细胞。所以 Fig. 1c 中，肿瘤细胞主要为 PTCs。将癌细胞数据单独分离，以 Seurat 重新聚类，随后以 Monocle3 拟时分析，得到拟时轨迹图 Fig. 1d。将图示的位置作为拟时起点是因为，我们发现，如果将 Fig. 1b 的癌细胞切分成不同数量的聚类簇 Fig. 10，则其中与正常细胞更加近似的细胞会主要分布于 Fig. 1d 所示的拟时起点。随后，以这些癌细胞的差异基因 (Graph test) 共表达分析。Fig. 1e，根据细胞聚类树，RCC 细胞主要可以分为 5 种亚型 (Fig. 1f)，与拟时分析的结果相对应 (Fig. 1d)。通过将 RCC 细胞分为亚型，以备 RISC 集成，考察其在多组数据集成后，是否还具备原先数据集的特征。

Figure 1为图 Main figure 1 概览。

(对应文件为 ./Figure+Table/fig1.pdf)

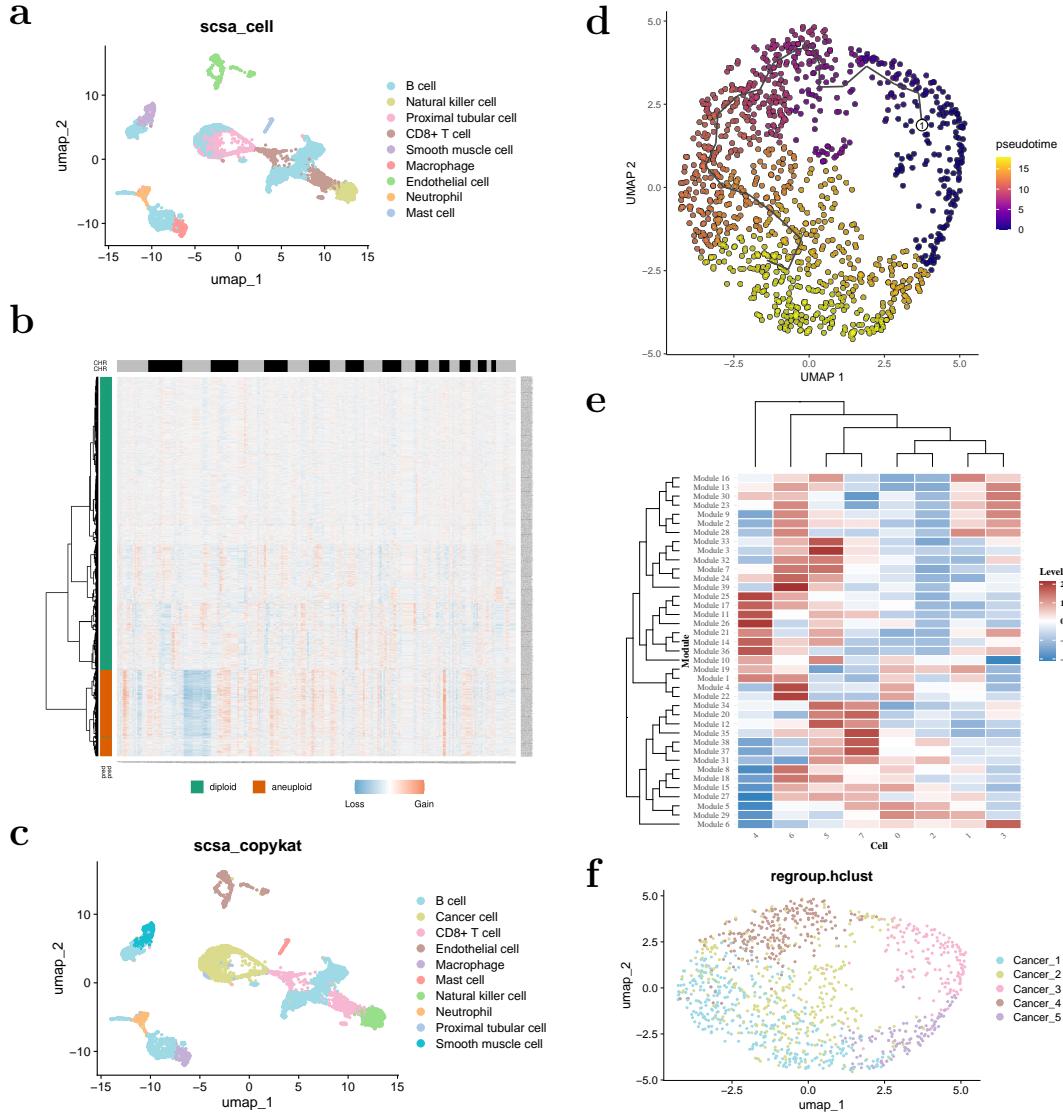


Figure 1: Main figure 1

5.2 不同类型 CKD 数据共表达分析和重聚类

对应 CKD 细胞的数据集，这里执行了 5.1 中近似的处理。即，以 Seurat 将数据集预处理后，以 SCSA 注释，随后提取 PTCs 重新聚类分析，继之以拟时分析，以 Graph test 获取差异表达基因，以此做聚类共表达分析。总共有 3 组 CKD 数据集，对应疾病 HN、IgA、IMN，分别对应 Fig. 2 的：a, b, c; d, e, f; g, h, i。对这些 CKD 疾病的亚型区分聚类如 Fig. 2 c、f、i 所示，可以观察到，HN 和 IgA 型 CKD 具有拟时间梯度的亚型变化，而 IMN 型 CKD 疾病则并不具备这样的亚型分布特征（IM-2 分散在不同的位置）。对于 CKD 的数据集，这里并不做拟时起点的判别，而只是作为 RISC 多数据集成之前的数据检查和质控。事实上，在随后的分析中，可以在 UMAP 图 (Fig. 3b) 中观察到 CKD 或 RCC 这些数据集的亚型特征。

Figure 2为图 Main figure 2 概览。

(对应文件为 ./Figure+Table/fig2.pdf)

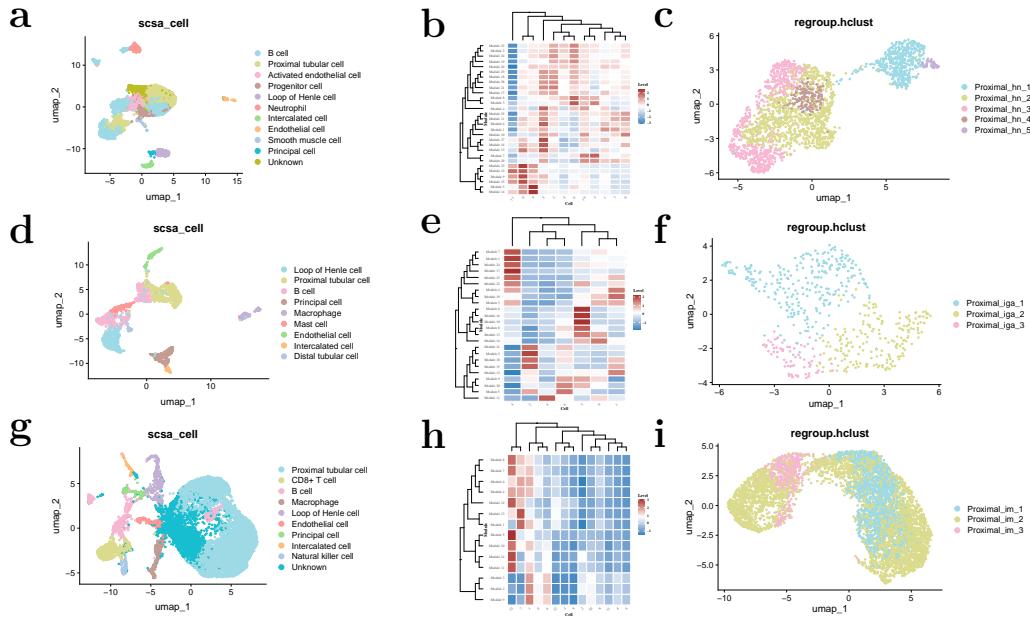


Figure 2: Main figure 2

5.3 RCC 和 CKD 数据集成

以 RISC¹⁰ 将 RCC 和 CKD 的数据集集成。除了上述的 RCC 和 CKD 数据组外，额外加入了一组健康对照 (Healthy Control, HC) 的数据以供对照。Fig. 3a 为 RISC 计算的各个数据集的得分参考 (Sets 1-5 对应 RCC, HM, IgA, IMN, HC 单细胞数据集。)，据此，选择 HC 数据集作为参考数据，将一共 5 组的单细胞数据集集成。UMAP 聚类如 Fig. 3b 所示。同时，Fig. 3b、c 也展示了拟时分析轨迹图。这里，我们在 RISC 中实现数据集成和 UMAP 聚类后，将对应的结果转移到 Monocle3 种进行拟时分析，而没有重新以 Monocle3 的算法重新聚类并拟时分析，体现了分析的延续性（考虑到 RISC 的聚类可能更适应于自身的数据集成的结果）。可以发现，RCC 细胞 (Cancer 1-5)，IgA 细胞 (Proximal iga 1-3)，HN 细胞 (Proximal hn 1-5)，三者的聚类结果保留了各自单独数据集的特征（颜色的渐变梯度，源自于单独数据集的亚型区分）。RISC 算法并没有影响各自数据集的特征。进一步观察可以发现，IgA 型细胞与 RCC 细胞的聚类更加相邻，而 HN 型和 IM 型细胞位于更远端，这可能预示着：相比于 HN 和 IM，IgA 更有可能发展为 RCC。

Figure 3为图 Main figure 3 概览。

(对应文件为 ./Figure+Table/fig3.pdf)

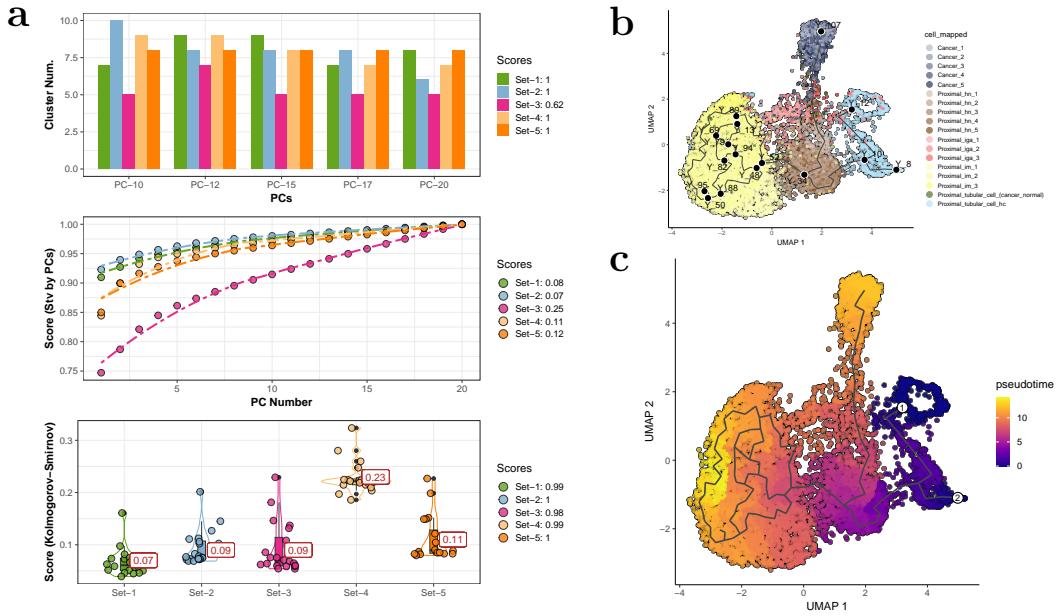


Figure 3: Main figure 3

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

在 5.3 中，聚焦于 IgA 和 RCC 发现了两者之间可能存在的转化风险。此处，将 IgA 细胞、RCC 细胞还有 HC 细胞进行了差异分析。将三者的差异分析结果分集得到 Fig. 4a。取 Cancer vs IgA 的合集（包括 up 组和 down 组），进行 GSEA 的 KEGG 富集分析，发现富集结果首要为 Oxidative phosphorylation (OXPHOS) 和 Ribosome 通路。可以观察到，对比 IgA 肾病，RCC 的两条通路的基因表达主要展现出下调趋势 Fig. 4c、d。结合已有研究可知，某些亚型的癌细胞需要氧化磷酸化 (OXPHOS) 才能生存，增加的 OXPHOS 依赖性通常是癌症干细胞和细胞对化疗和靶向治疗产生耐药性的标志¹⁴。GO 富集分析的结果同样引证了 KEGG 分析的结果 (Fig. 4e)。然而，当以 Cancer vs Control 的差异基因富集分析时 (Fig. 4f)，并没有发现富集到 OXPHOS 这一通路。这一差异通路似乎仅存在于 RCC 和 IgA 之间。

Figure 4为图 Main figure 4 概览。

(对应文件为 ./Figure+Table/fig4.pdf)

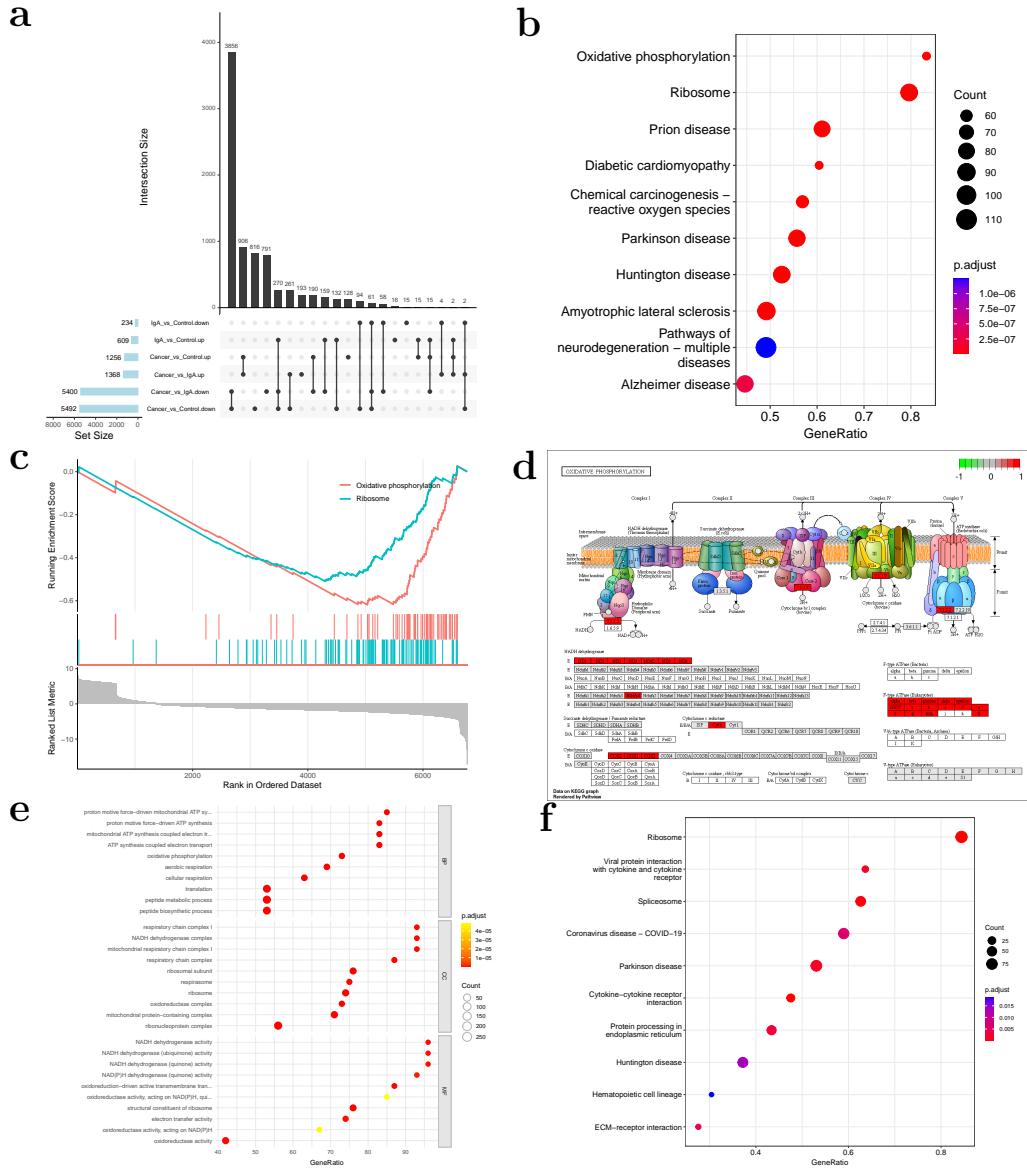


Figure 4: Main figure 4

5.5 以新的 IgA 和 RCC 数据集验证

为了验证 5.4 的分析并非数据特异的，接下来的分析以一组全新的 RCC 和 IgA 数据集（来自不同上述的 GSE 数据）重新考察（两个数据集的处理和细胞注释参考 7.6）。Fig. 5a，选择 IgA 数据集作为参考集（Sets 1-2 对应 RCC 和 IgA 数据集），将两个数据集集成。随后的 UMAP 聚类同样发现，RCC 细胞和 IgA 细胞表现出相邻近似的聚类关系（Fig. 5b），与 5.3 的结果一致。随后的差异分析和富集分析表明（Fig. 5c），

OXPHOS 通路的确是两者的差异通路。然而，GSEA 的富集结果却发现 (Fig. 5d)，新的这一批数据集表现出 RCC 对应 IgA 细胞的 OXPHOS 通路的整体上调趋势。结合 5.3，合理的推测是：由于肿瘤的异质性，不同亚型的肿瘤细胞分为 OXPHOS 型或非 OXPHOS 依赖型¹⁵；对于非 OXPHOS 依赖型，使用效率较低的糖酵解来产生 ATP 和结构单元 (Warburg effect) 是其主要方式¹⁵。进一步结合富集分析结果，在 Fig. 4b 中可以发现富集的 Ribosome 通路呈总体下调趋势，而对应的 RCC 细胞可能是 OXPHOS 非依赖型，有氧糖酵解是其主要产热方式，而其发生的场所主要为细胞质。而 Fig. 5c 和 d 的结果相反，没有观察到富集的 Ribosome 通路的整体下调，却可以发现产热 (Thermogenesis) 可能增强了 (基因表达整体上调)，说明该 RCC 细胞由于为 OXPHOS 依赖型，并不依赖于糖酵解产生热量，因此 Ribosome 通路没有表现出下调趋势。然而，上述推测还有待进一步验证。

Figure 5 为图 Main figure 5 概览。

(对应文件为 ./Figure+Table/fig5.pdf)

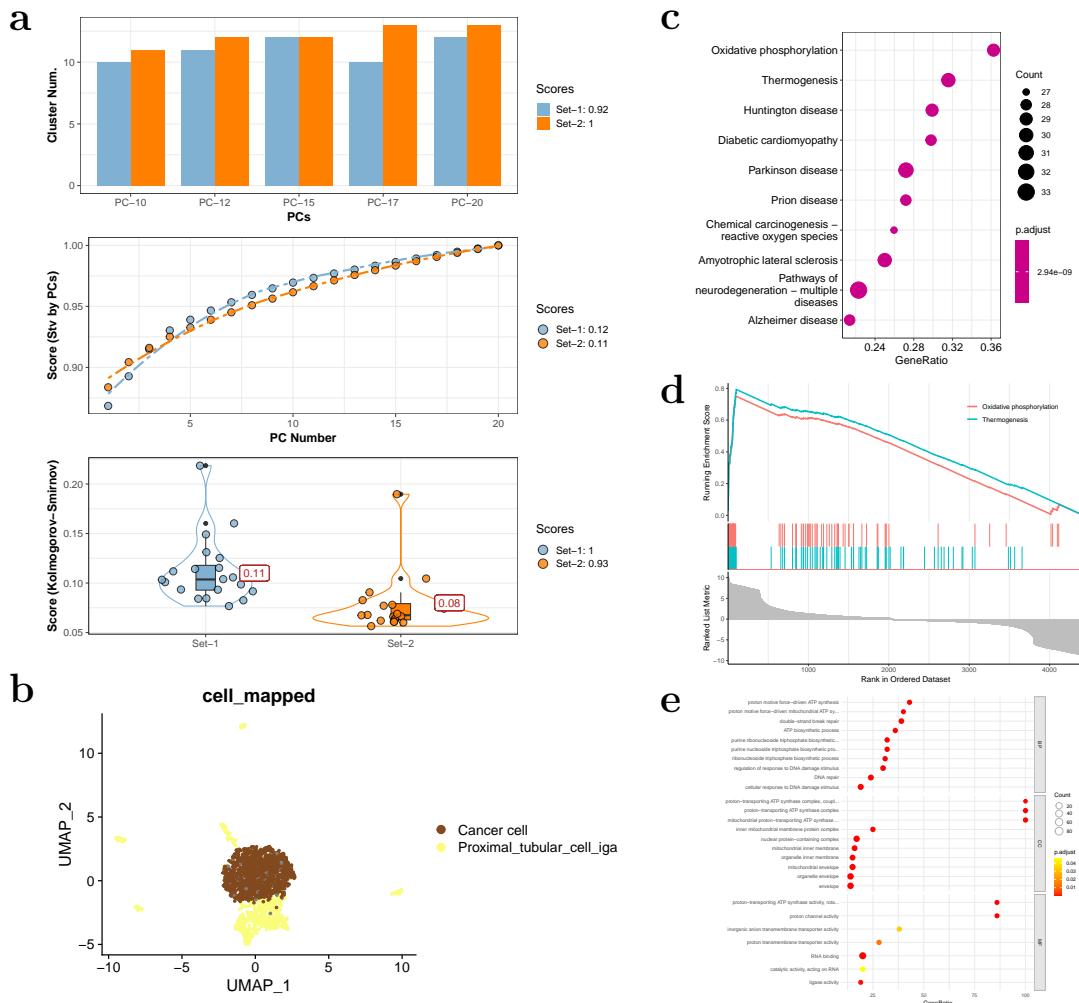


Figure 5: Main figure 5

6 结论

为了探究 CKD 和 RCC 之间可能存在的转化关系，本研究利用了 GEO 数据库的多组不同类型的 CKD 数据与 RCC 数据集成分析，主要发现：

- 基于拟时分析，相比于 HN 和 IMN 型 CKD，IgA 有更高的可能往 RCC 方向发展（PTCs）。
- RCC 和 IgA（PTCs）的首要差异通路为 OXPHOS，而这一差异通路没有在 RCC 和 HC 中发现。
- 以新的数据集验证后，OXPHOS 同样为 RCC 和 IgA 的差异通路，但 RCC 表现出了 OXPHOS 通路表达的异质性（见 5.5）。

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)

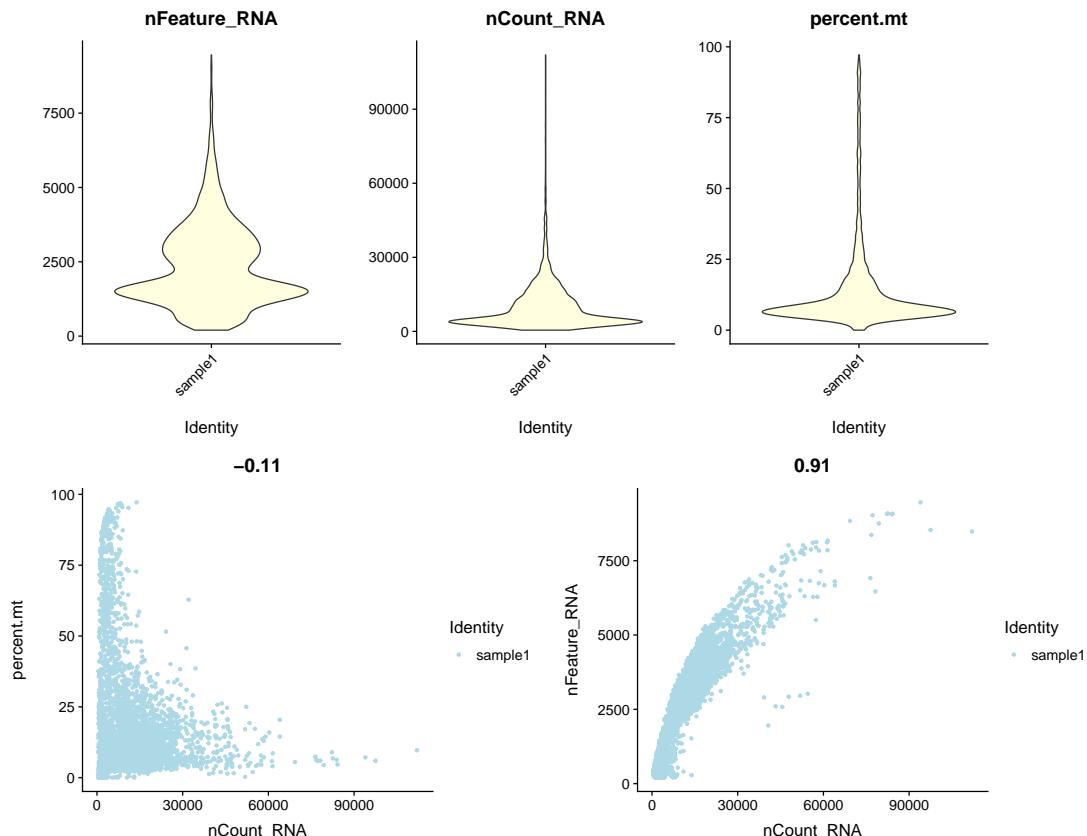


Figure 6: RCC data1 QC

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）

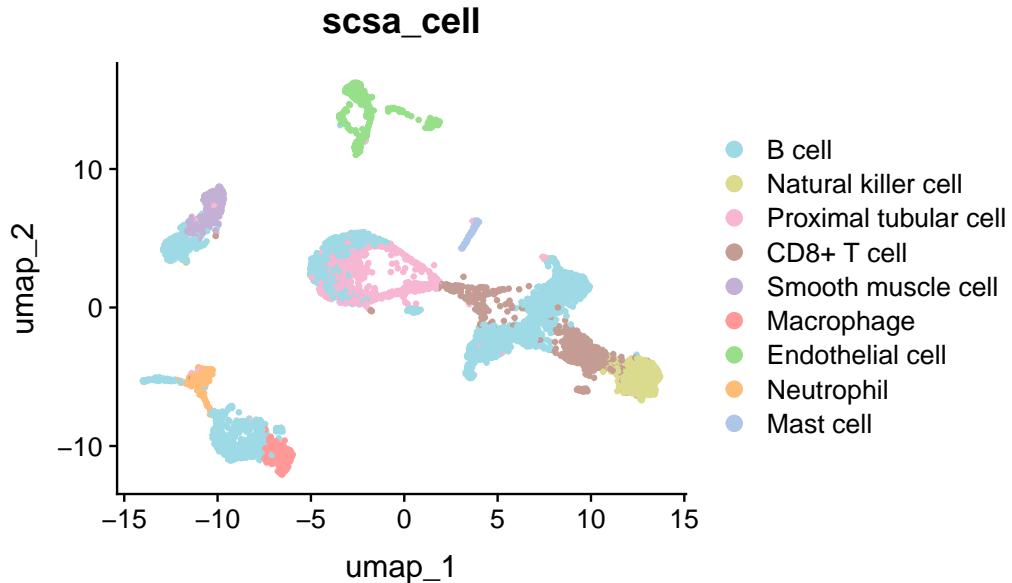


Figure 7: RCC data1 SCSA mapping

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）

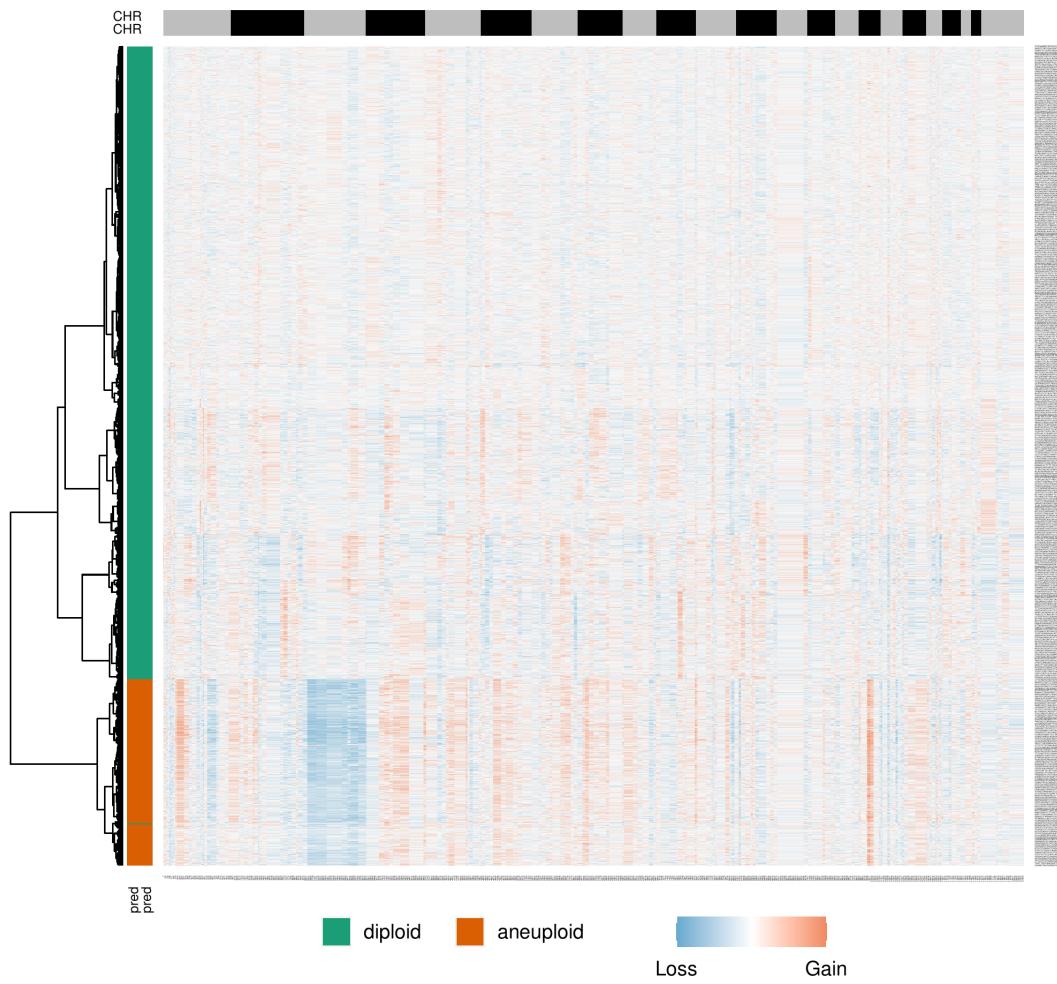


Figure 8: RCC data1 copykat prediction

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](#))

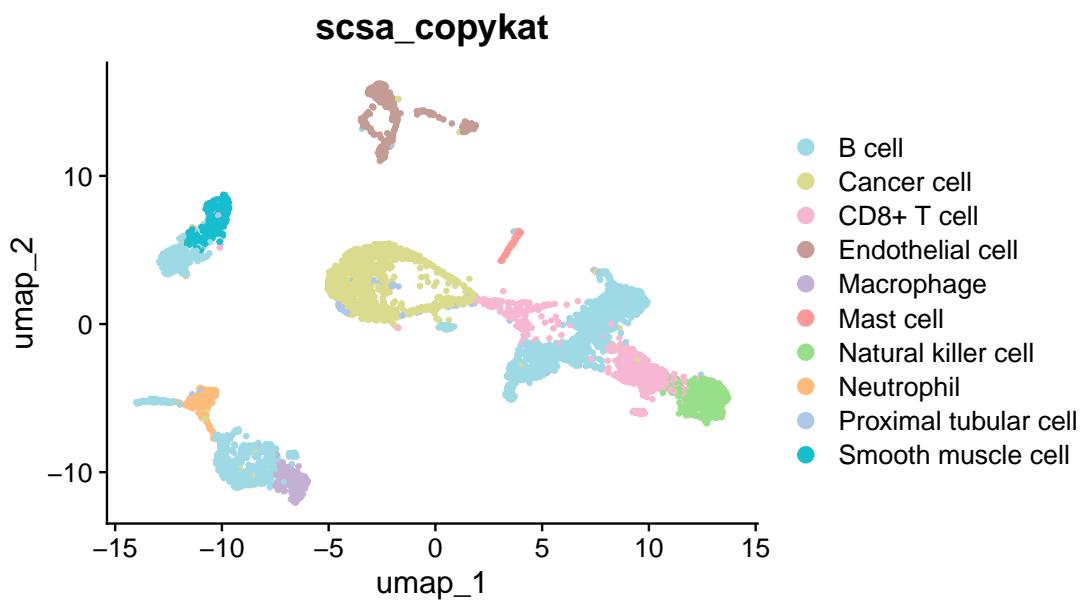


Figure 9: RCC data1 SCSA mapping with copykat prediction

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 Select start point for pseudotime 概览。

(对应文件为 Figure+Table/RCC-Select-start-point-for-pseudotime.pdf)

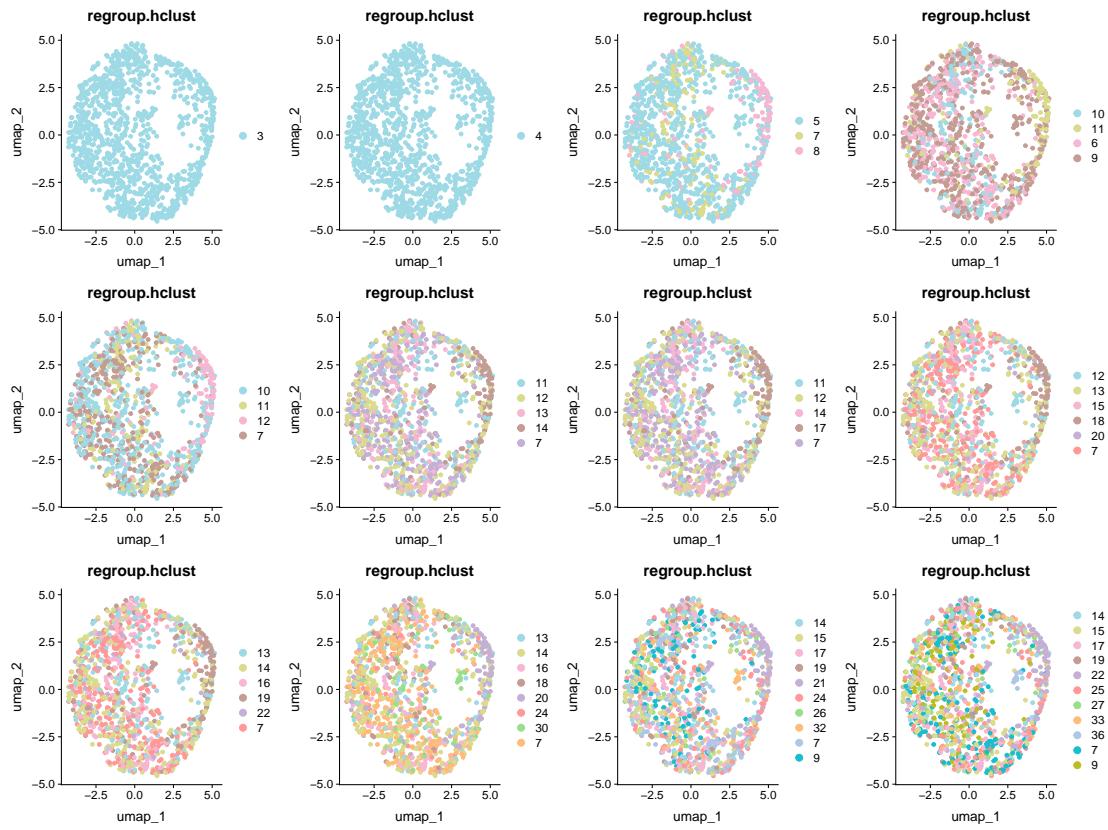


Figure 10: RCC Select start point for pseudotime

Figure 11为图 RCC data1 pseudotime 概览。

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

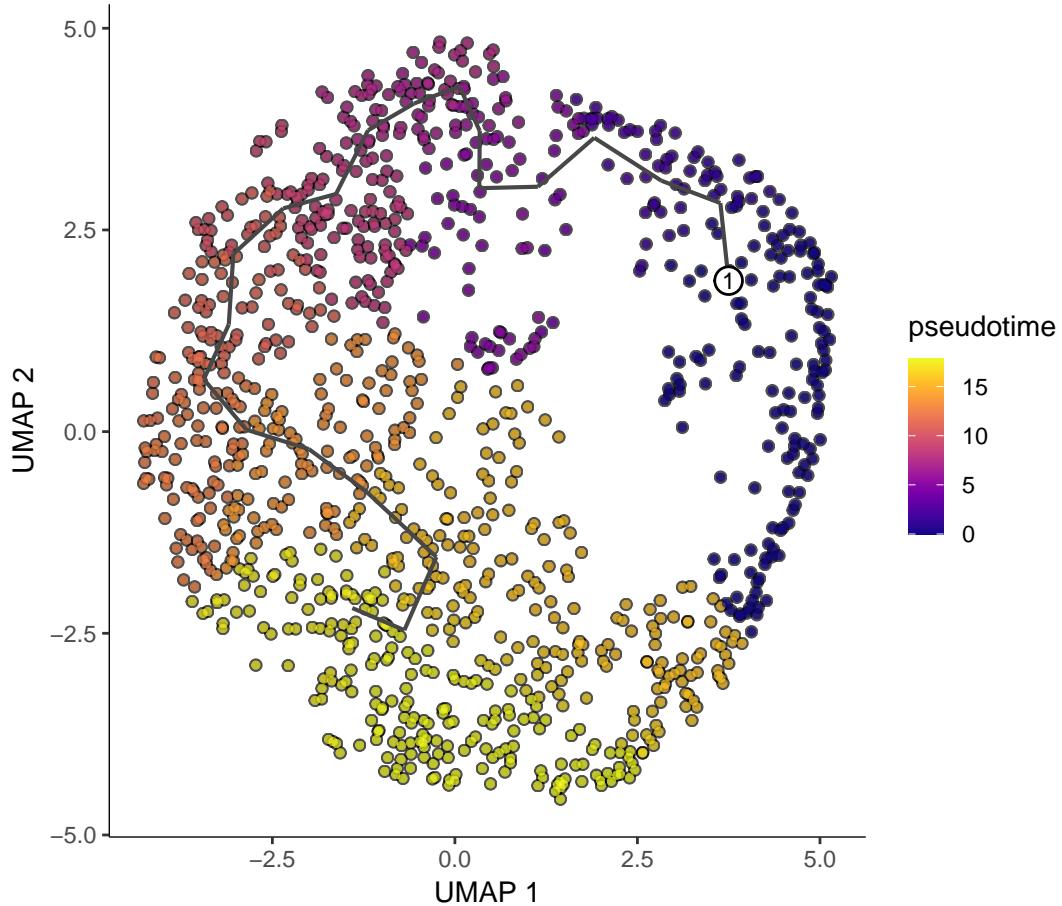


Figure 11: RCC data1 pseudotime

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 12为图 RCC data1 co expression modules 概览。

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

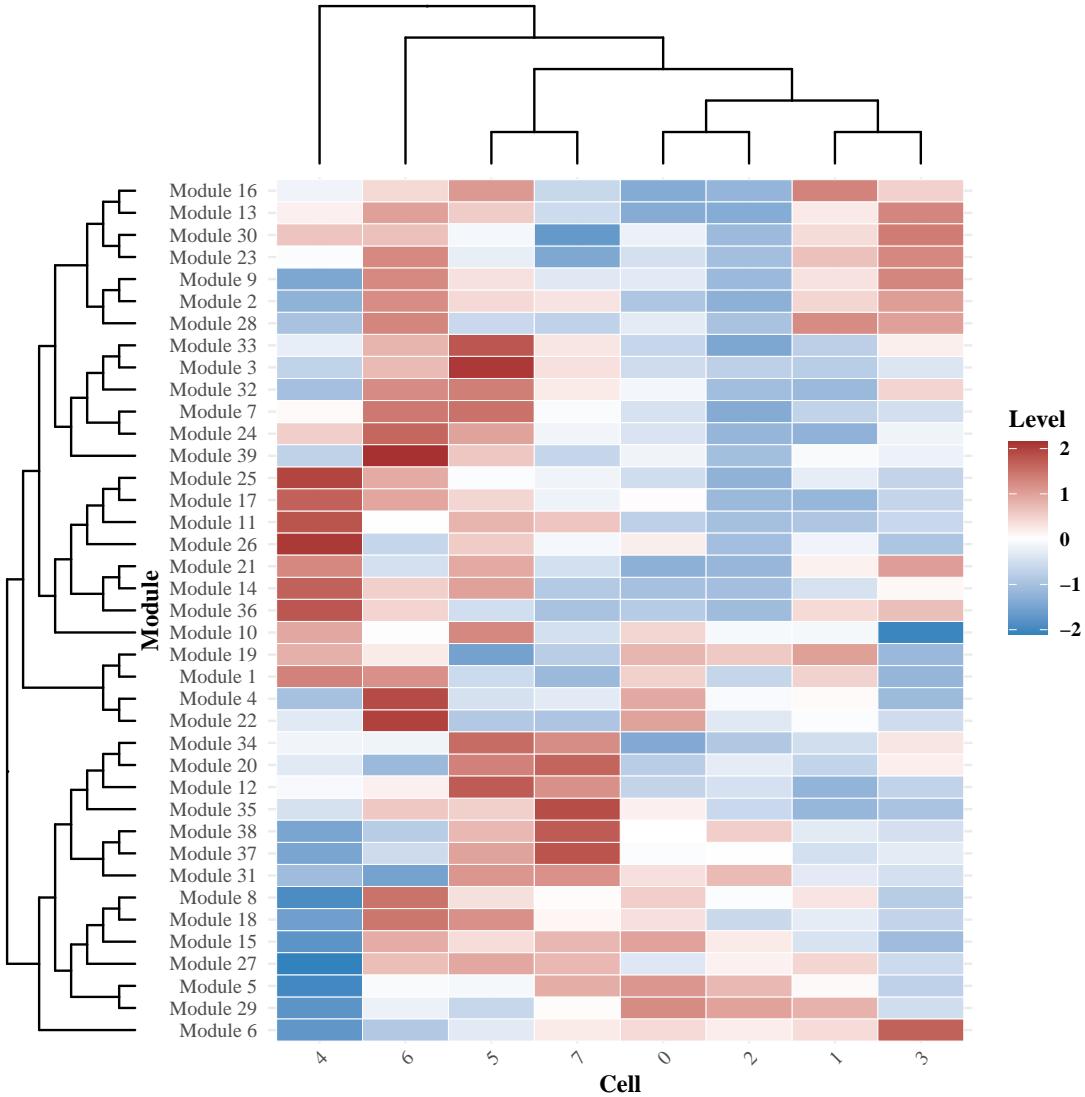


Figure 12: RCC data1 co expression modules

Figure 13为图 RCC data1 cancer subtype 概览。

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

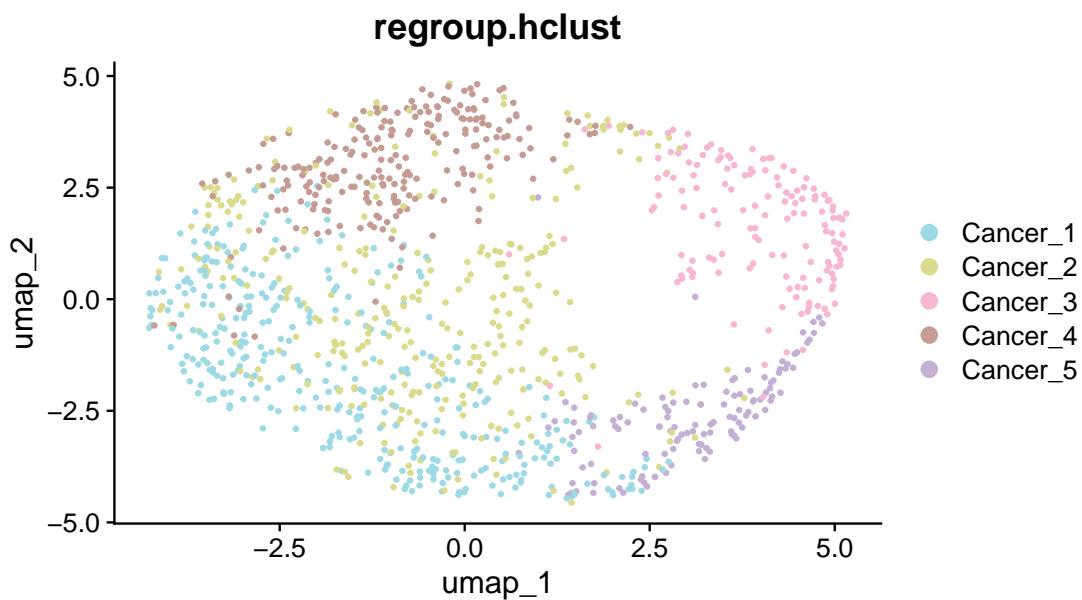


Figure 13: RCC data1 cancer subtype

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 14为图 HN data QC 概览。

(对应文件为 Figure+Table/HN-data-QC.pdf)

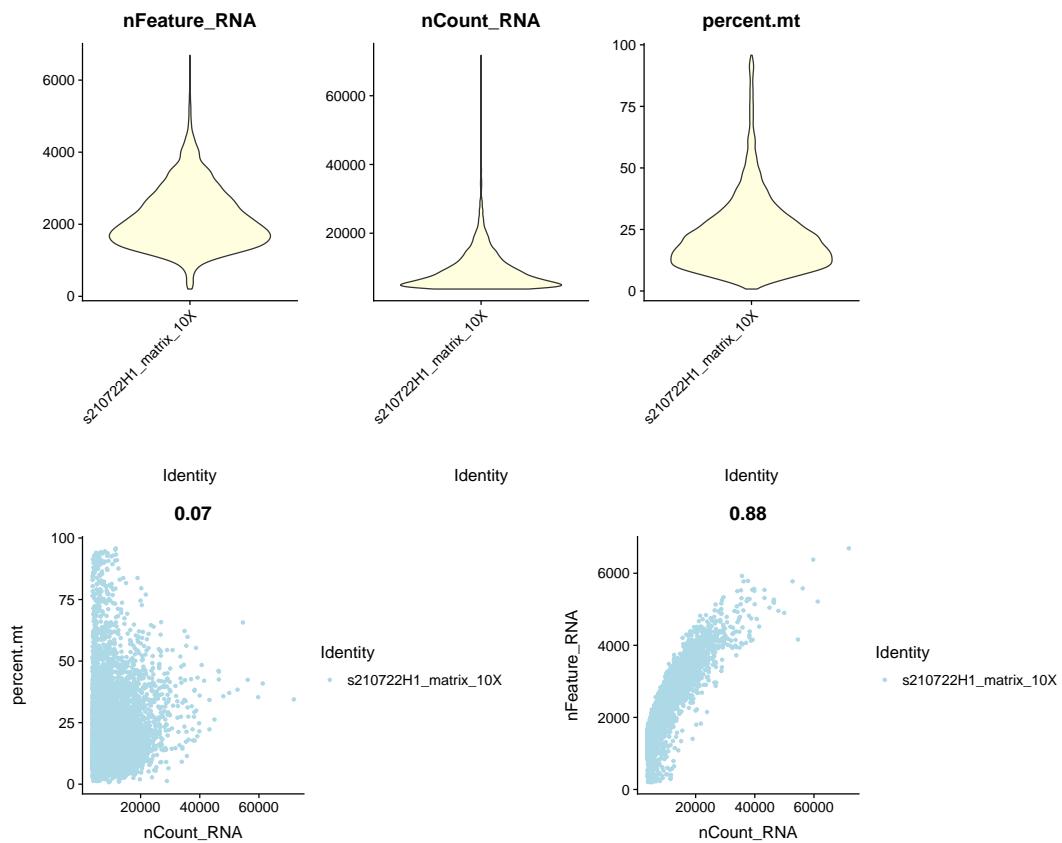


Figure 14: HN data QC

Figure 15为图 HN data SCSA mapping 概览。

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

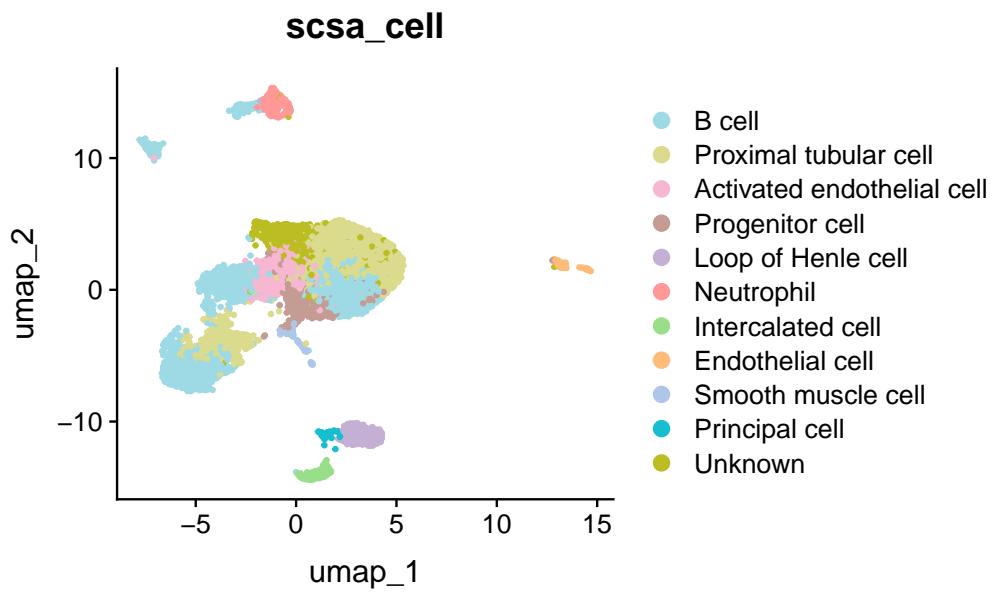


Figure 15: HN data SCSA mapping

7.2.1.2 共表达分析 Figure 16为图 HN data pseudotime 概览。

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

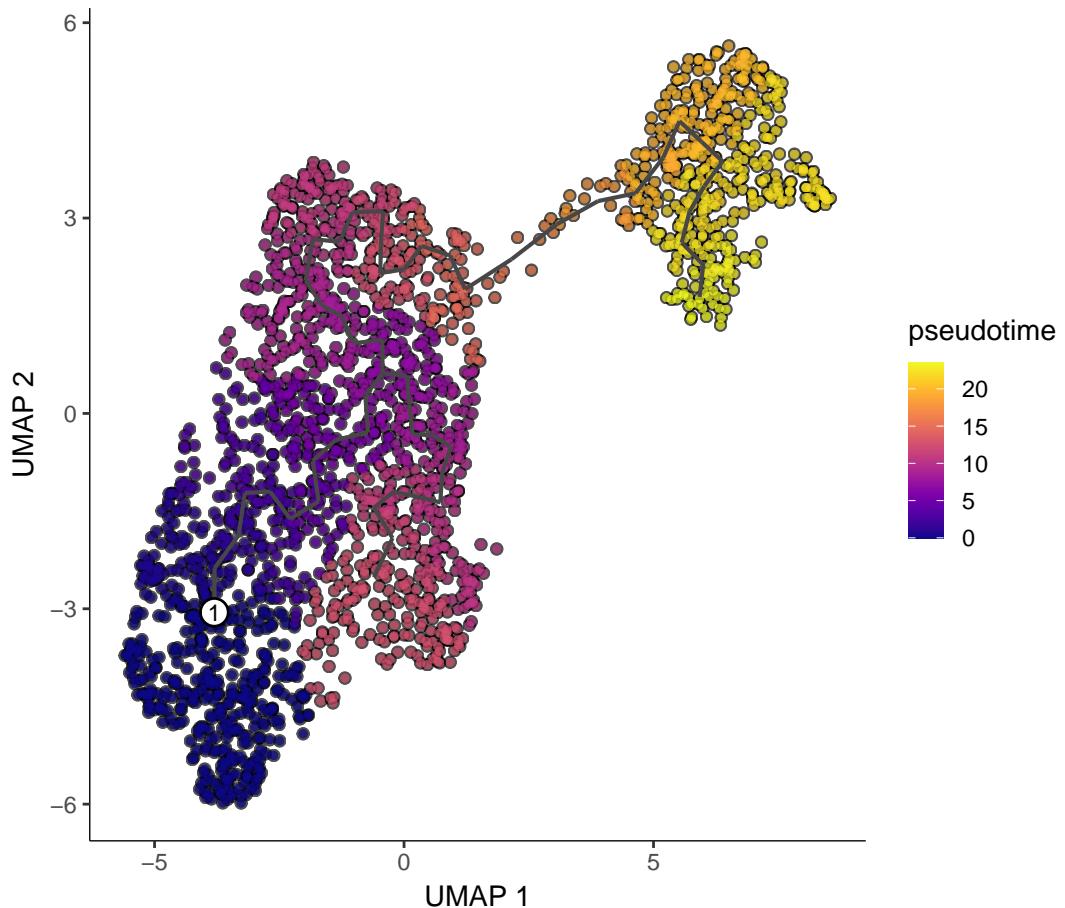


Figure 16: HN data pseudotime

Figure 17为图 HN data co expression modules 概览。

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

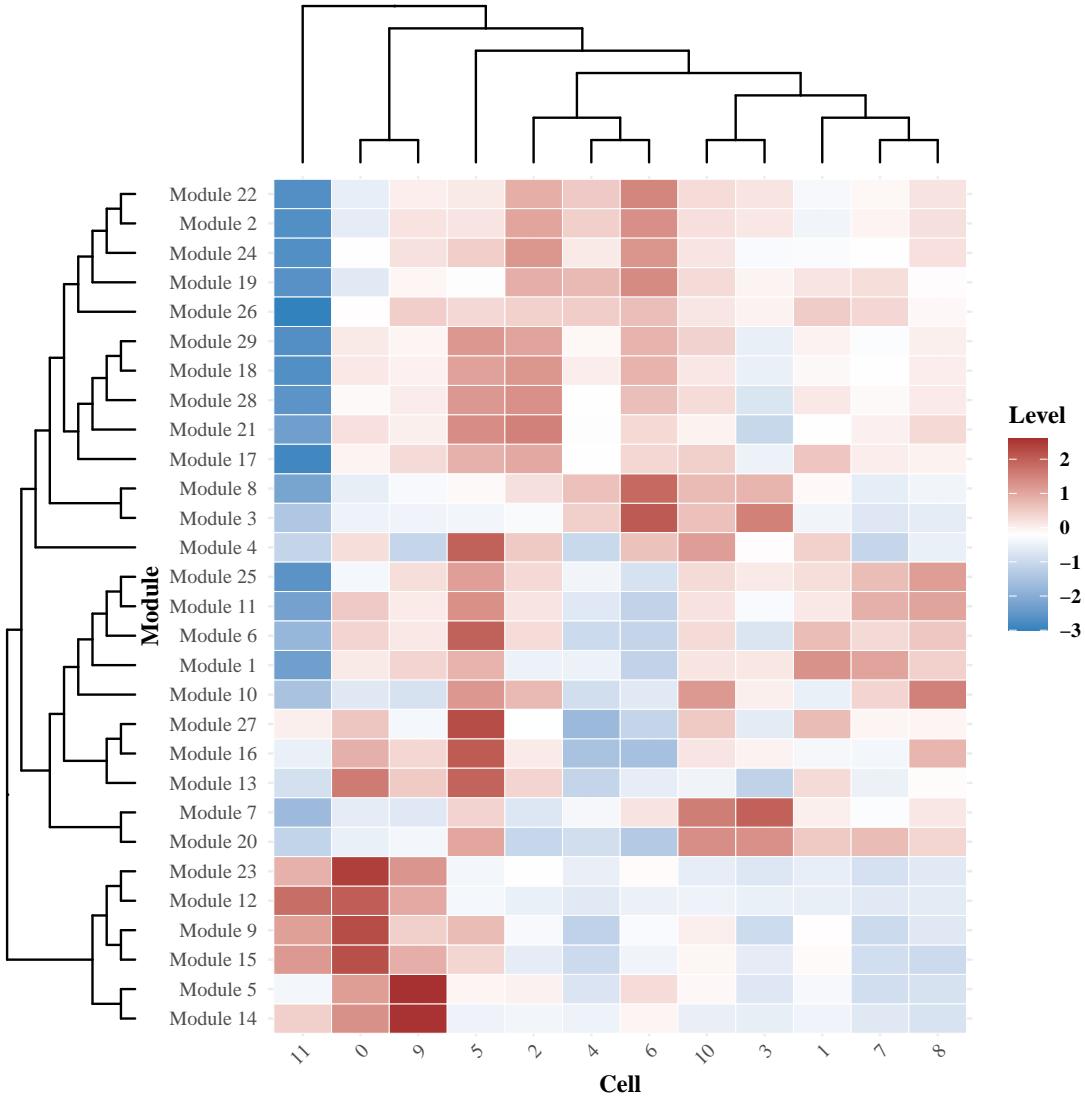


Figure 17: HN data co expression modules

Figure 18为图 HN data Proximal tubule cell subtype 概览。

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

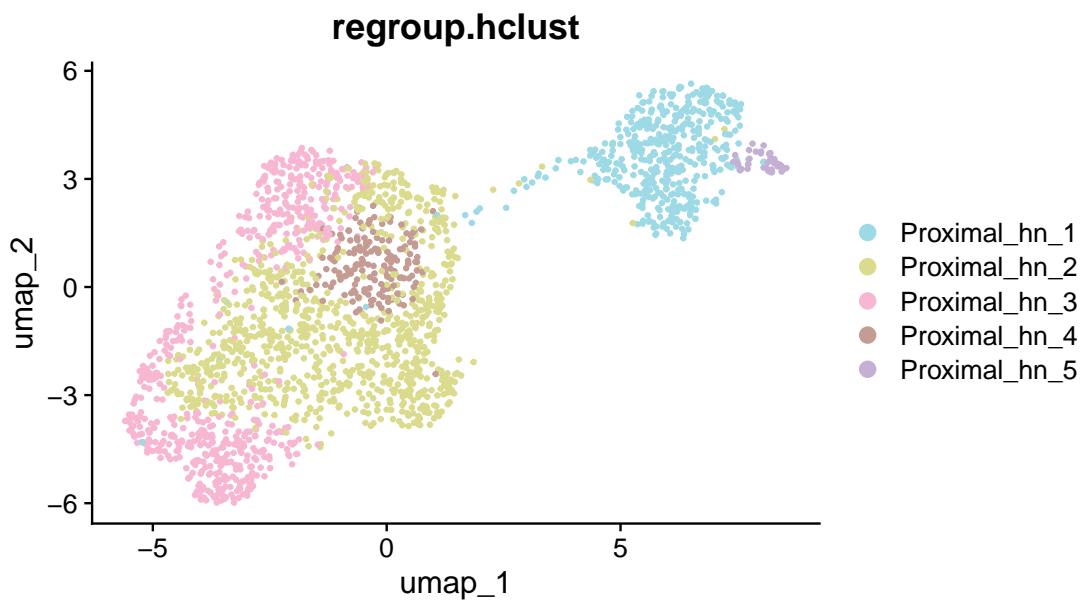


Figure 18: HN data Proximal tubule cell subtype

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 19为图 IgA data1 QC 概览。

(对应文件为 Figure+Table/IgA-data1-QC.pdf)

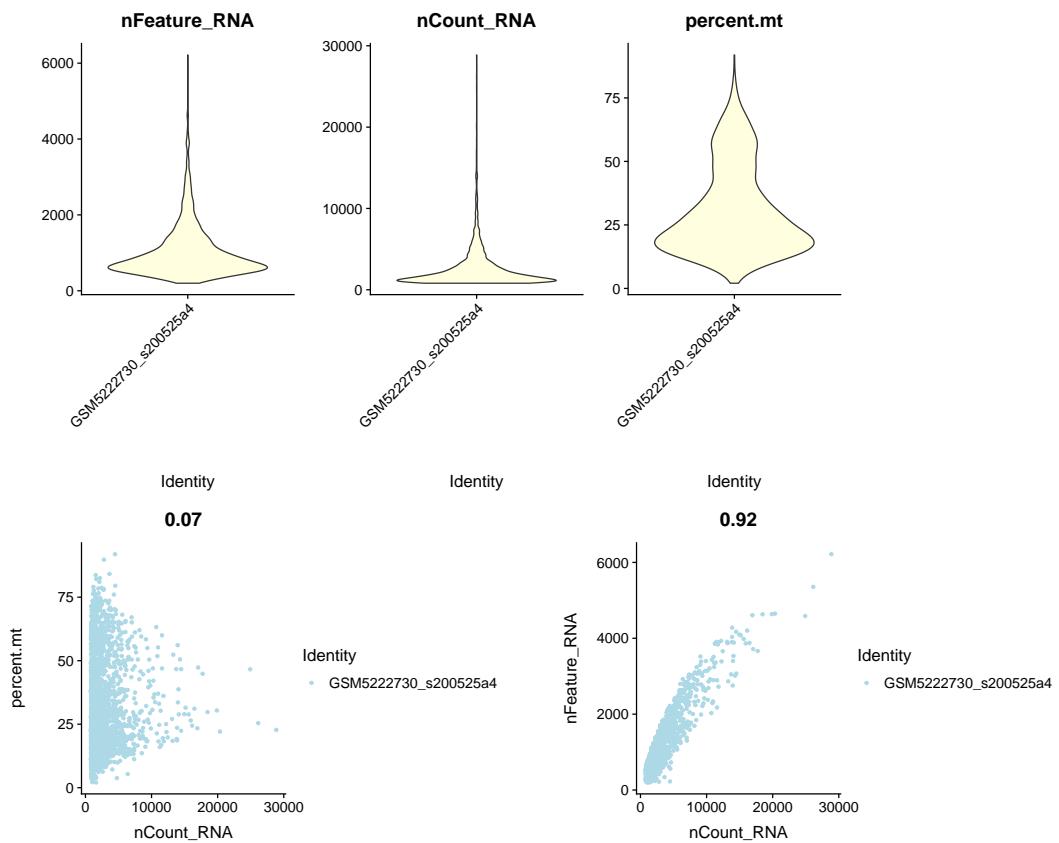


Figure 19: IgA data1 QC

Figure 20为图 IgA data1 SCSA mapping 概览。

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

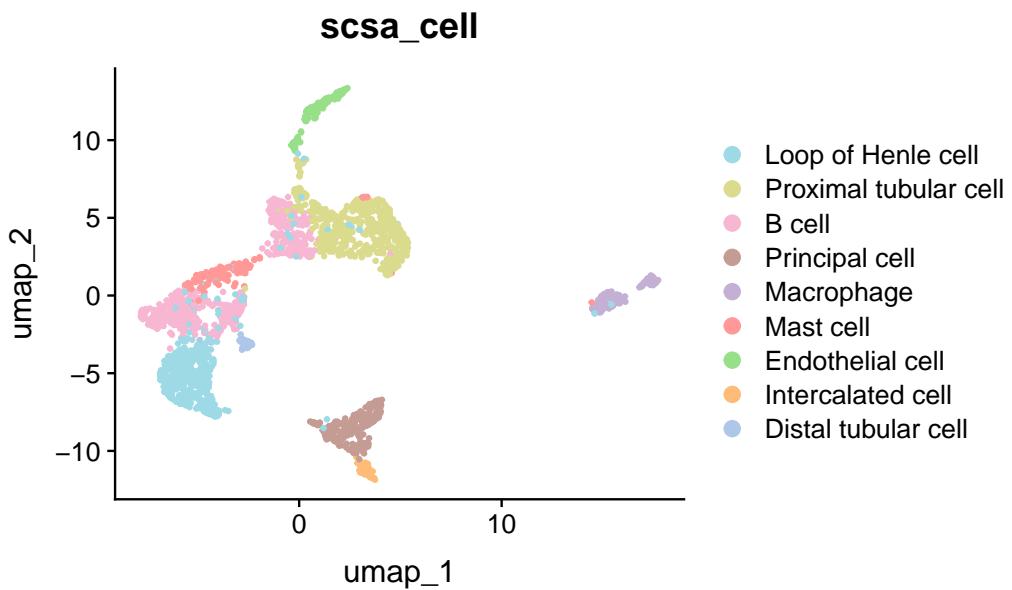


Figure 20: IgA data1 SCSA mapping

7.3.1.2 共表达分析 Figure 21为图 IgA data1 pseudotime 概览。

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

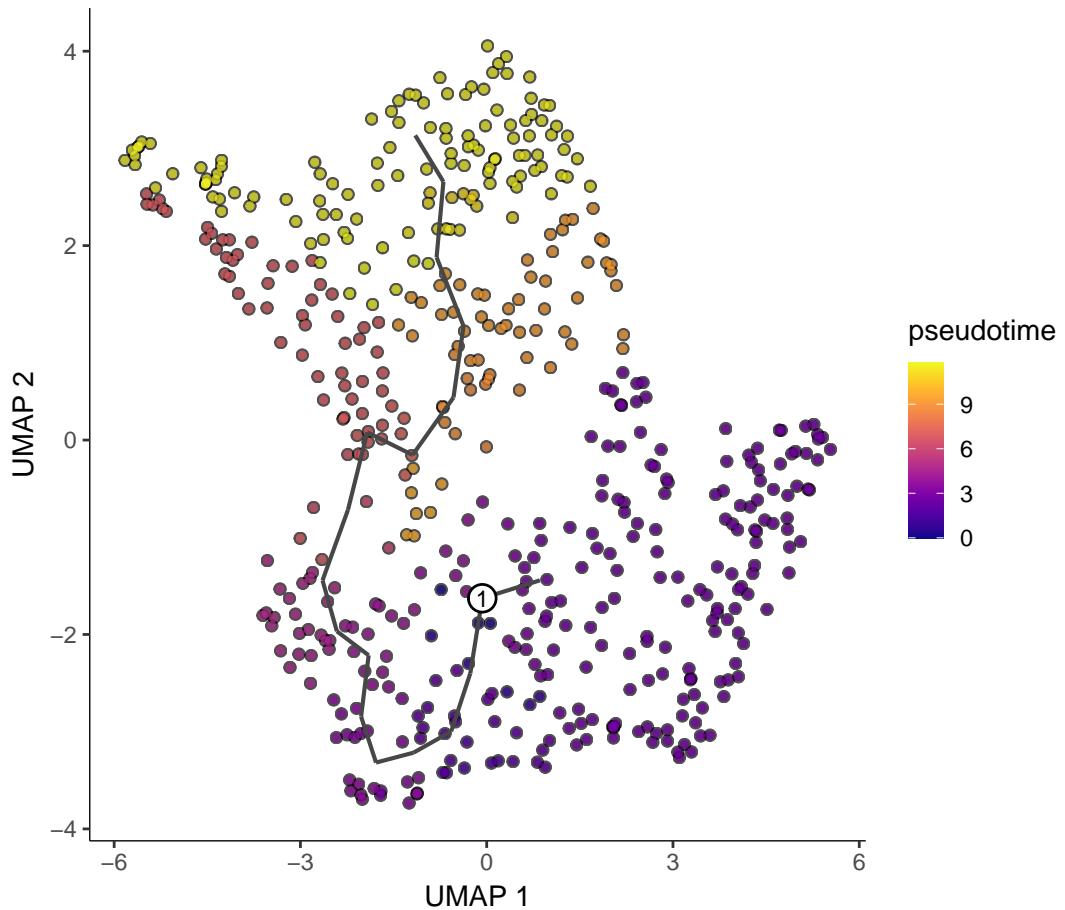


Figure 21: IgA data1 pseudotime

Figure 22为图 IgA data1 co expression modules 概览。

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

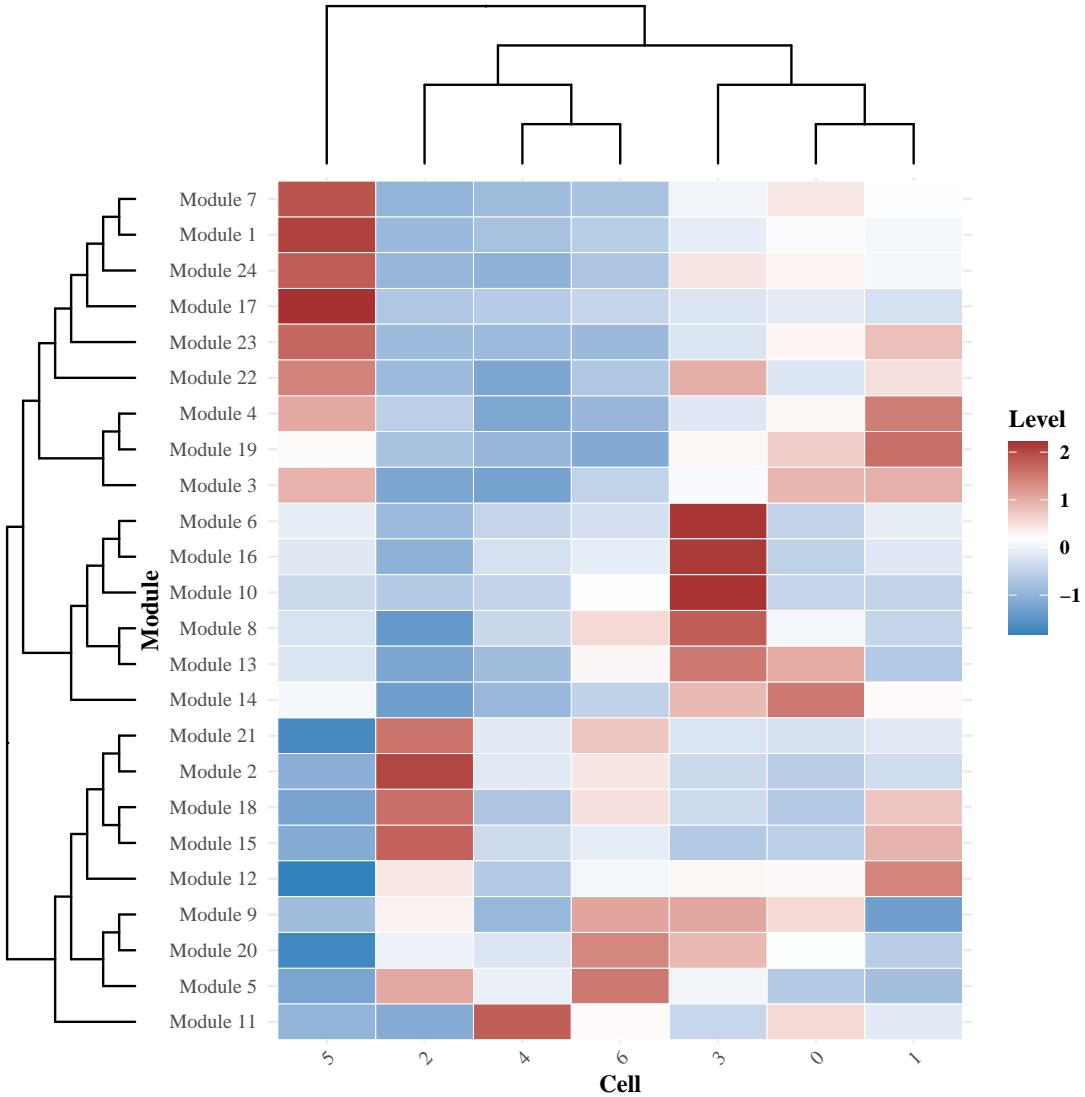


Figure 22: IgA data1 co expression modules

Figure 23为图 IgA data1 Proximal tubule cell subtype 概览。

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

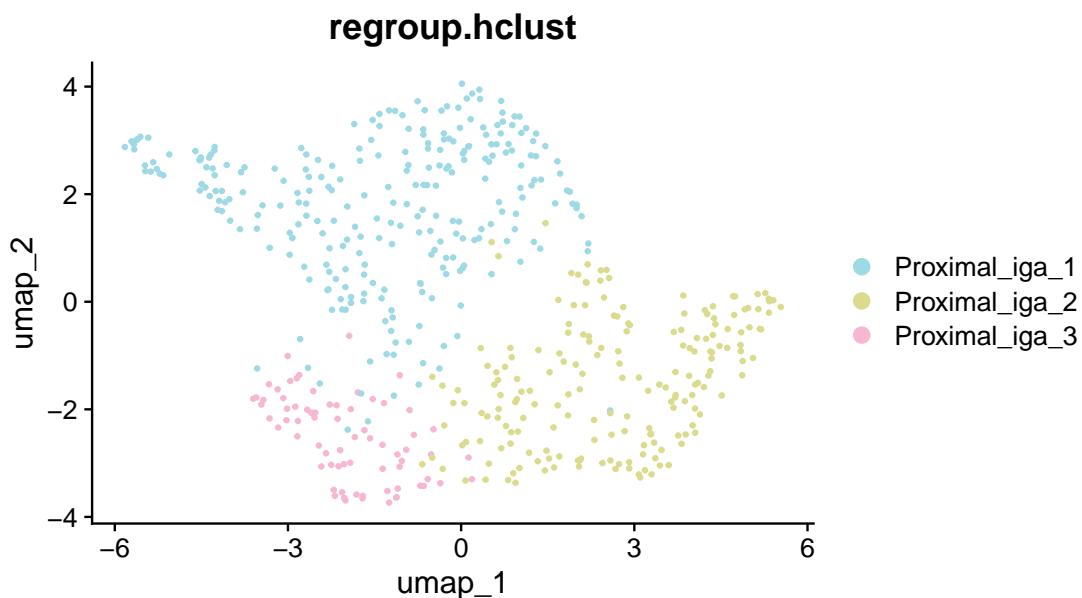


Figure 23: IgA data1 Proximal tubule cell subtype

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 24为图 IMN data QC 概览。

(对应文件为 Figure+Table/IMN-data-QC.pdf)

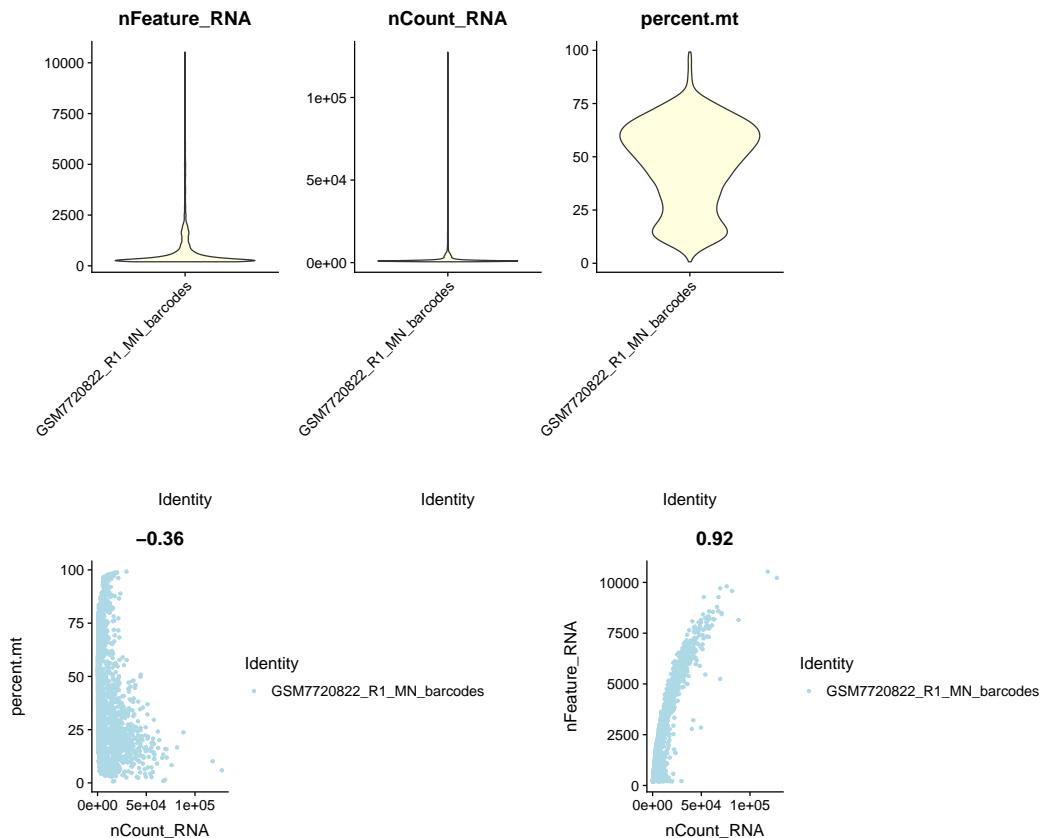


Figure 24: IMN data QC

Figure 25为图 IMN data SCSA mapping 概览。

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

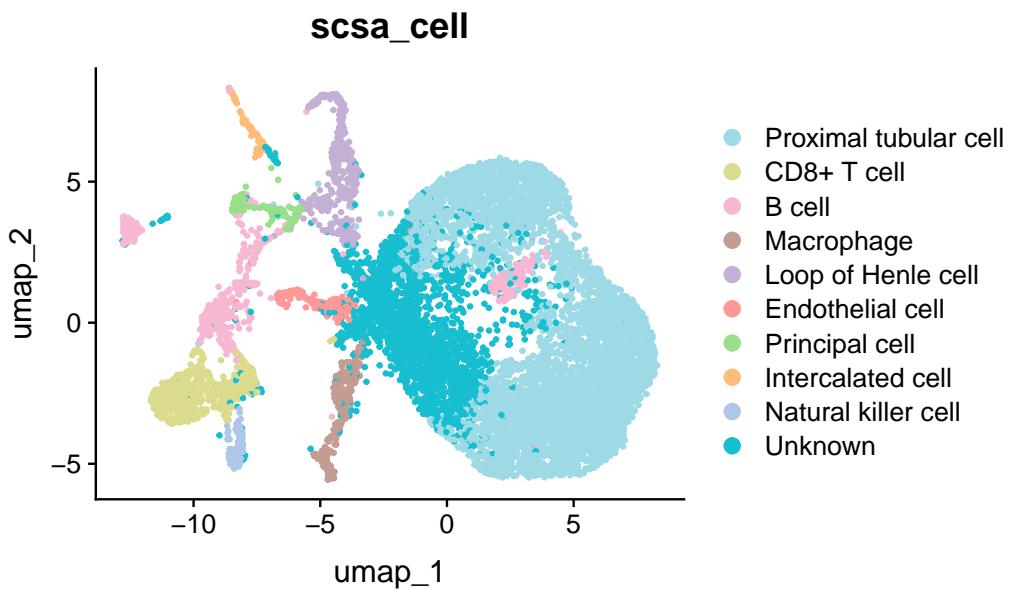


Figure 25: IMN data SCSA mapping

7.3.2.2 共表达分析 Figure 26为图 IMN data pseudotime 概览。

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

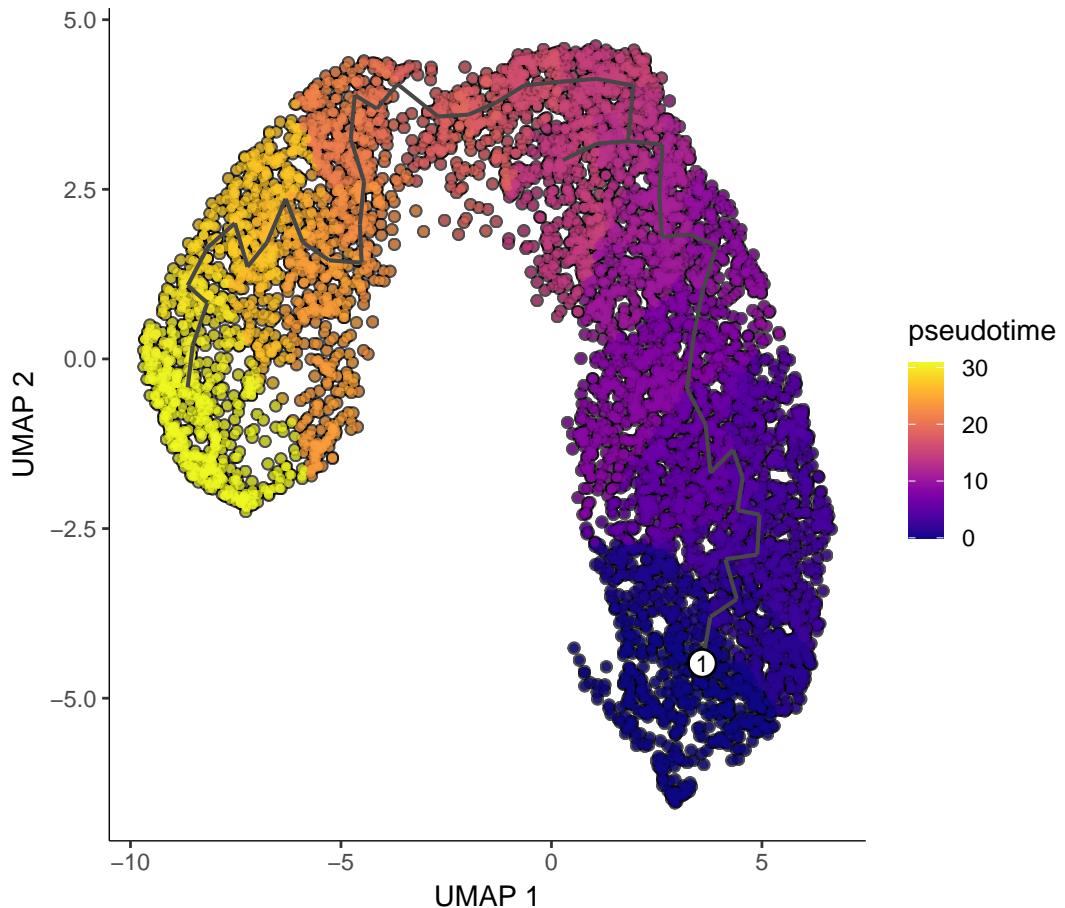


Figure 26: IMN data pseudotime

Figure 27为图 IMN data co expression modules 概览。

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

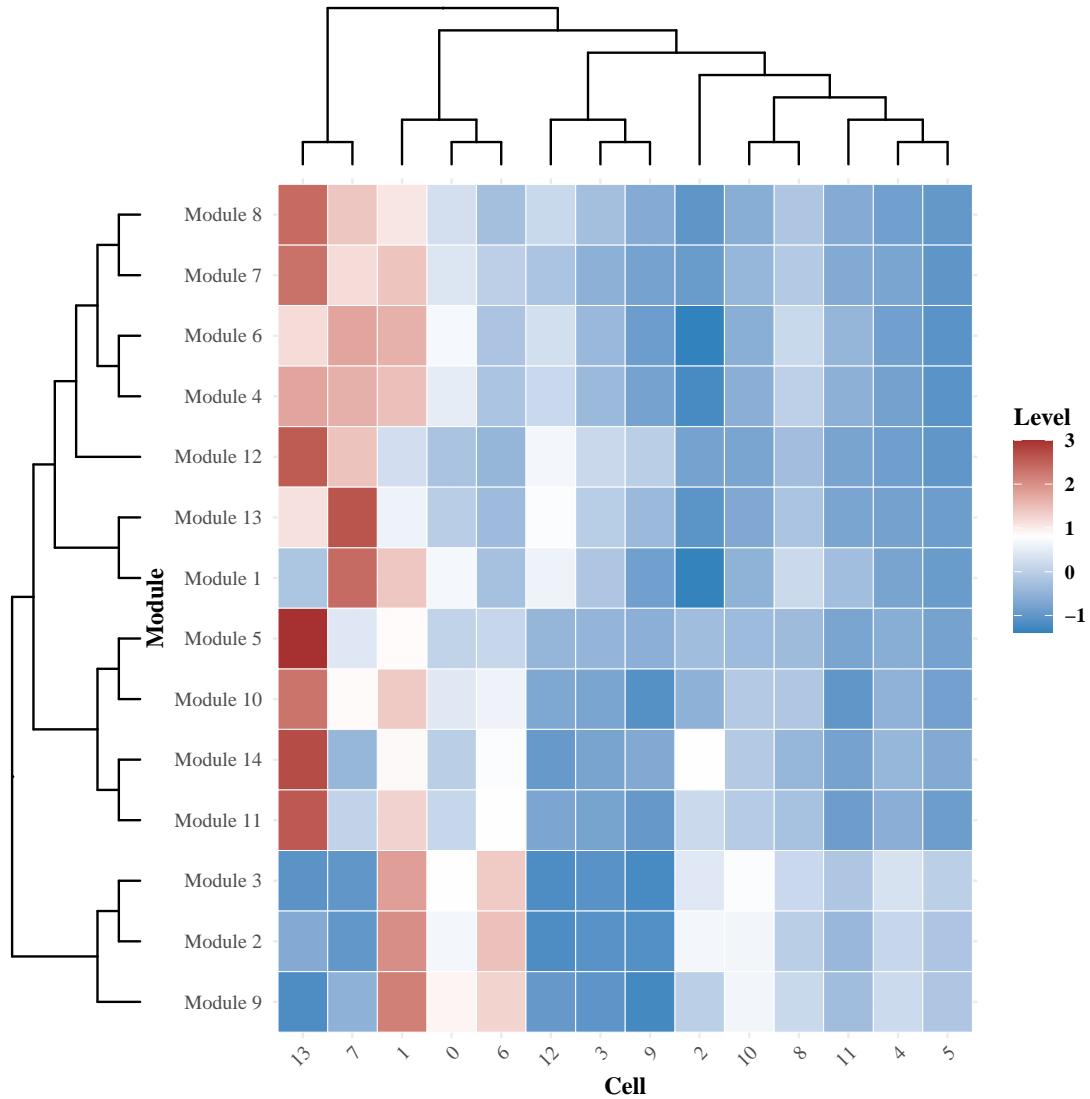


Figure 27: IMN data co expression modules

Figure 28为图 IMN data Proximal tubule cell subtype 概览。

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

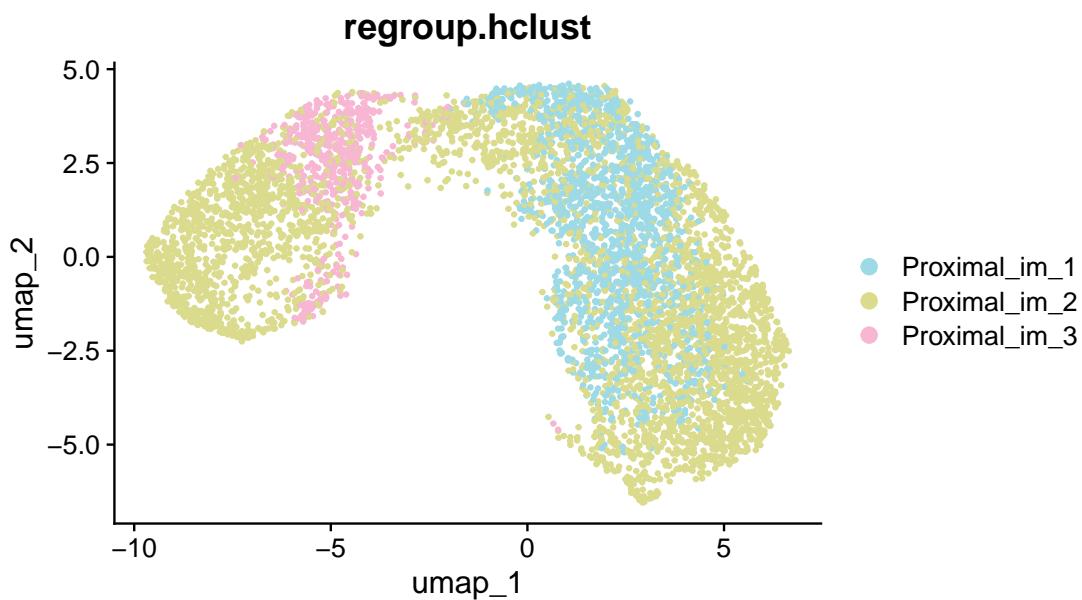


Figure 28: IMN data Proximal tubule cell subtype

7.4 正常对照 (Healthy control) (GSE171314)

7.4.0.1 细胞聚类和注释 Figure 29为图 HC data QC 概览。

(对应文件为 Figure+Table/HC-data-QC.pdf)

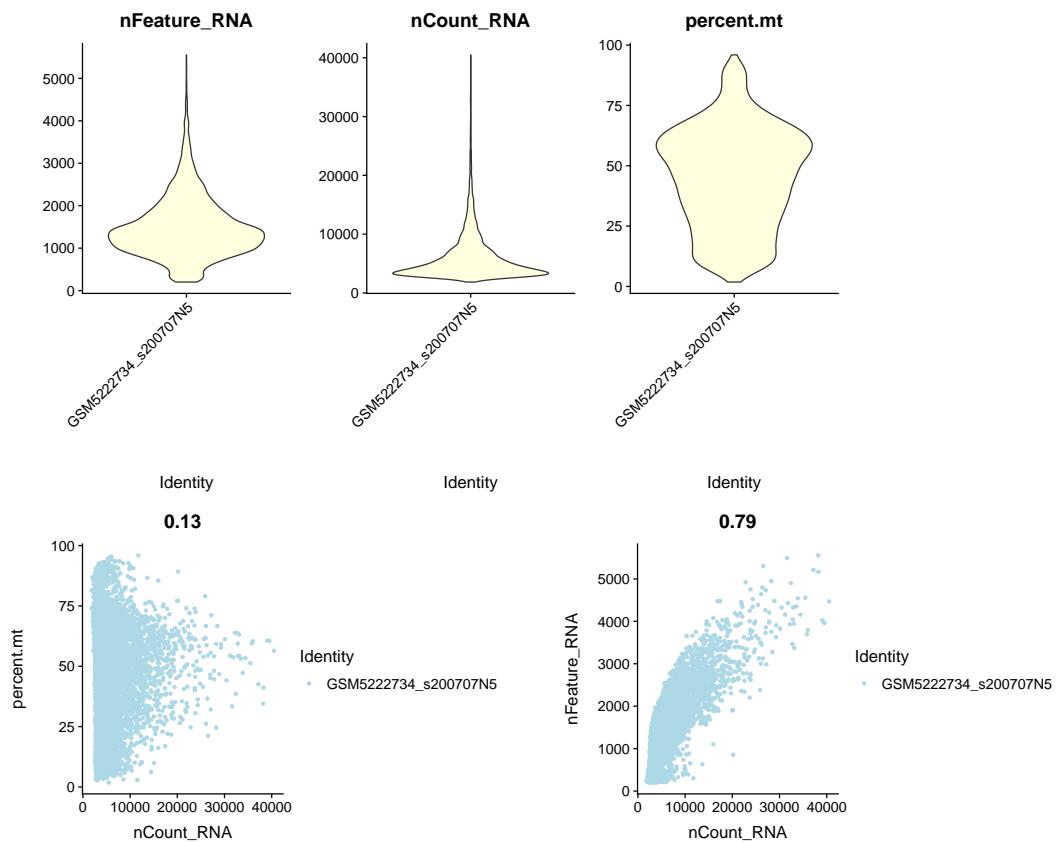


Figure 29: HC data QC

Figure 30为图 HC data SCSA mapping 概览。

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

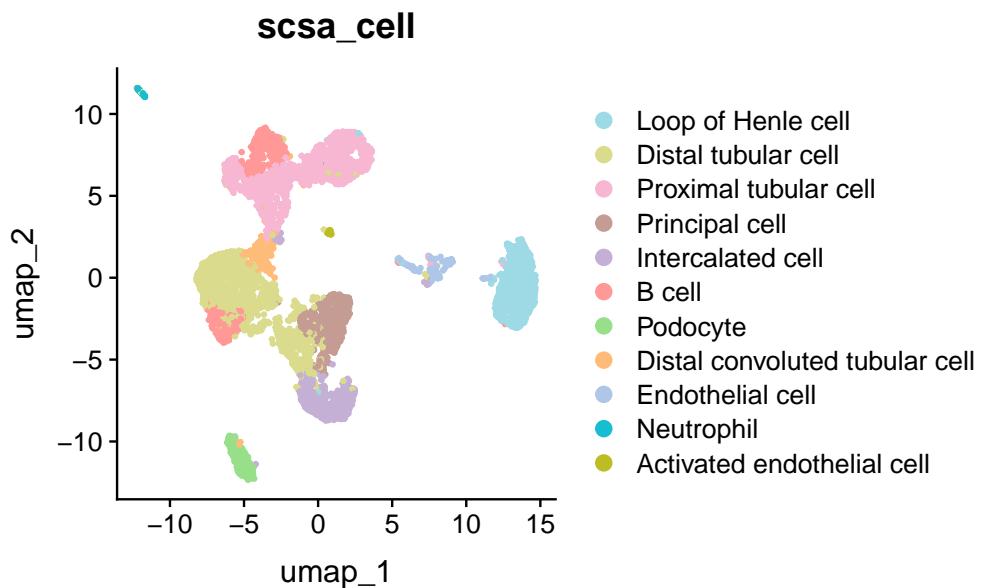


Figure 30: HC data SCSA mapping

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 31为图 SETS1 select reference dataset for integration 概览。

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

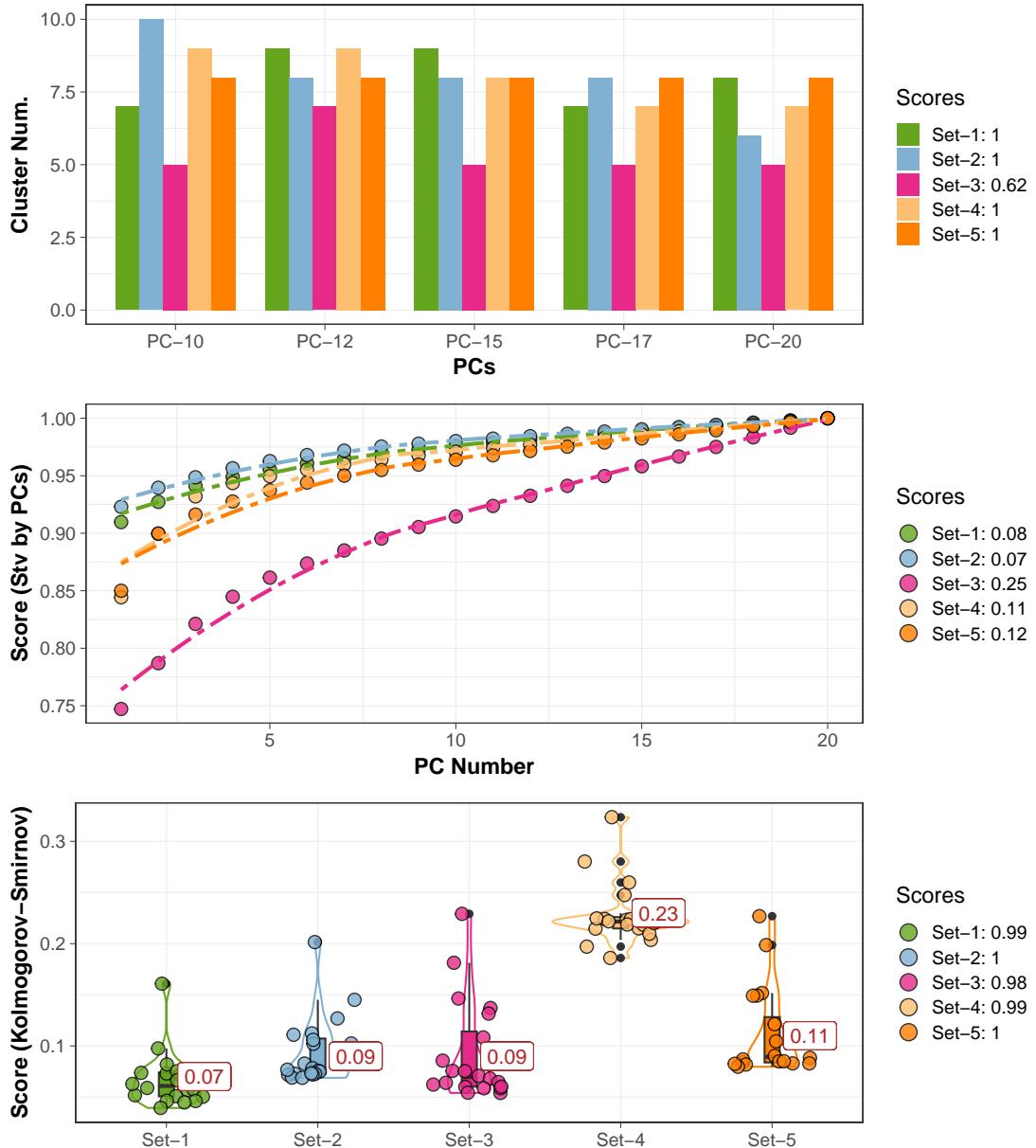


Figure 31: SETS1 select reference dataset for integration

Figure 32为图 SETS1 umap mapping 概览。

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

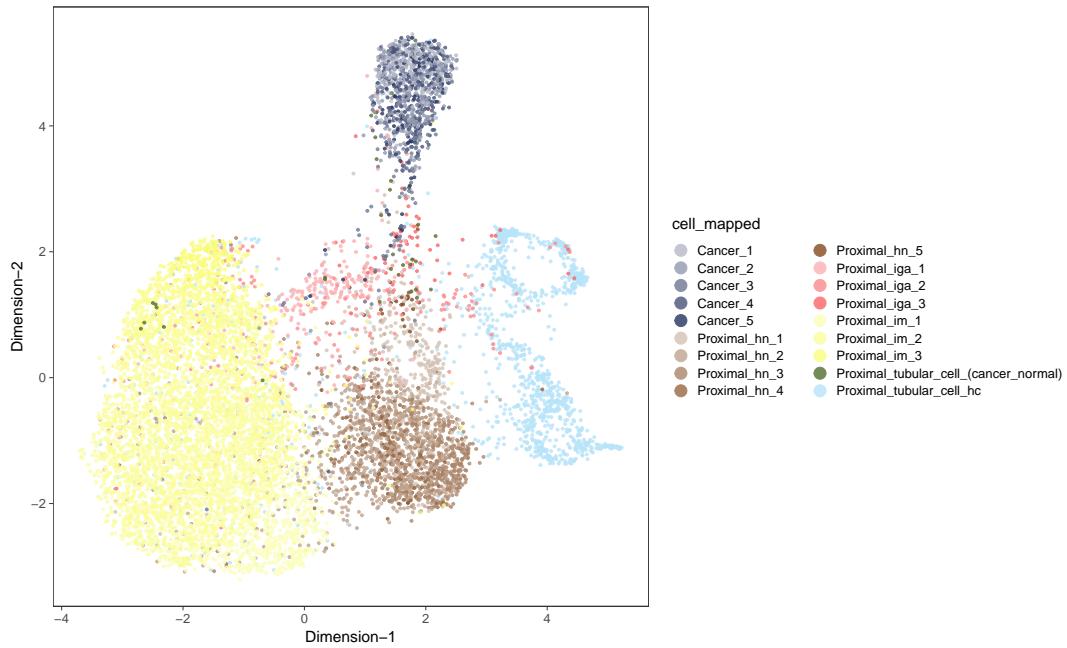


Figure 32: SETS1 umap mapping

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 33为图 SETS1 trajectory 概览。

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

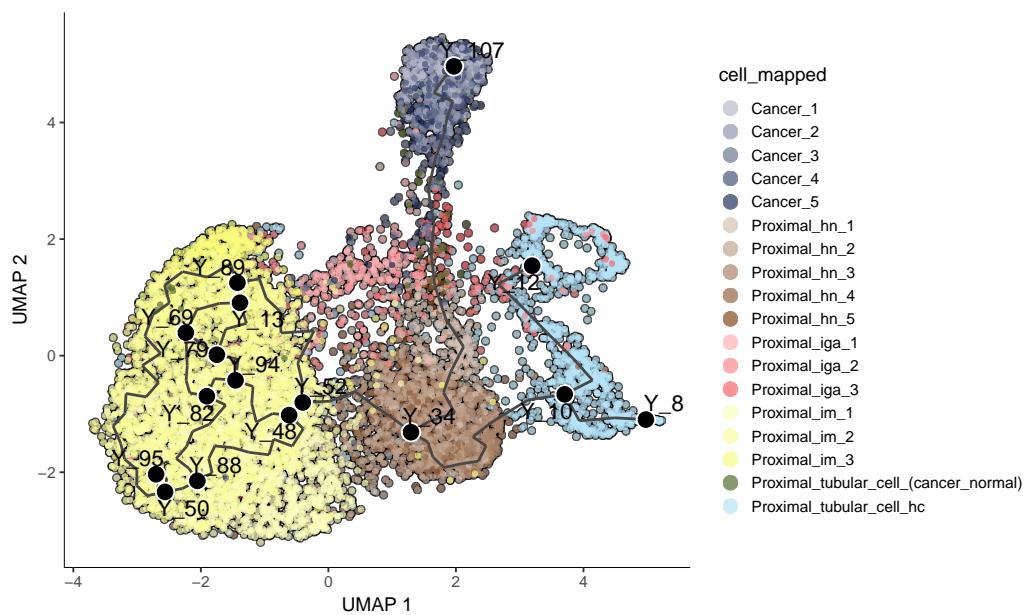


Figure 33: SETS1 trajectory

Figure 34为图 SETS1 pseudotime 概览。

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

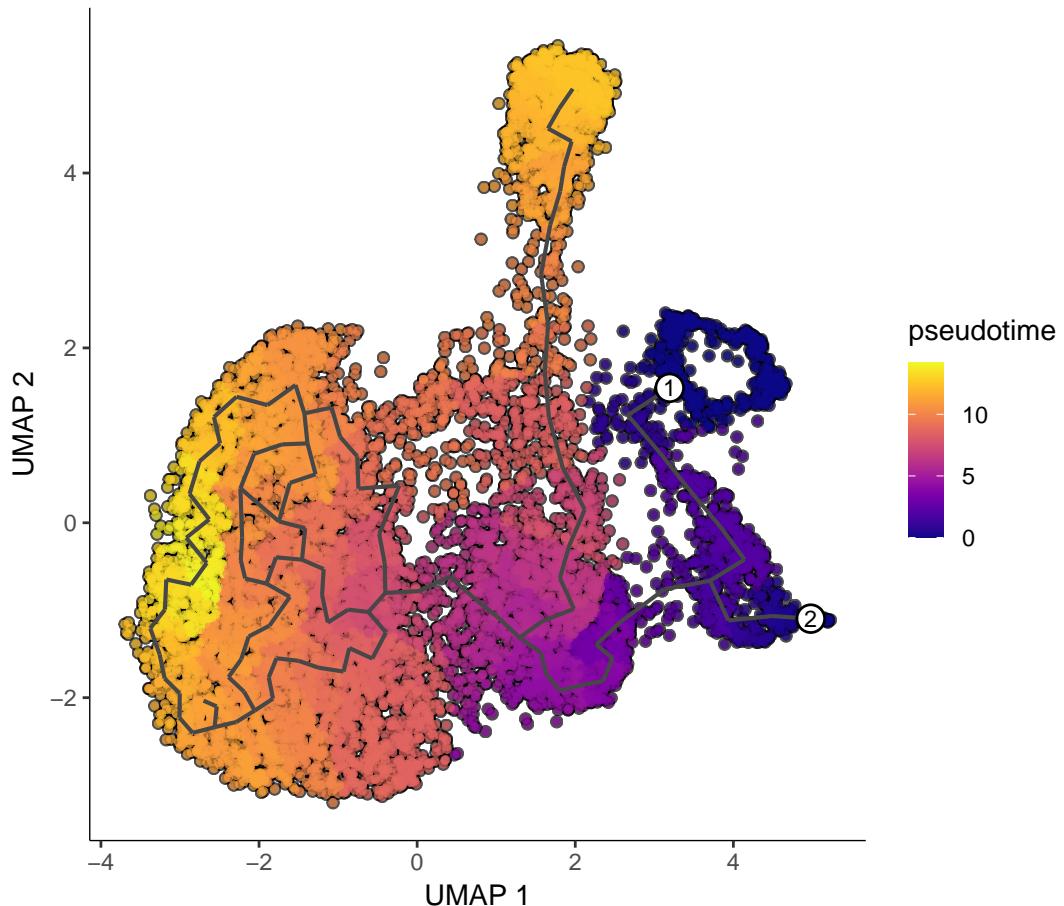


Figure 34: SETS1 pseudotime

7.5.3 IgA、Cancer、HC 差异分析

Figure 35为图 SETS1 intersection of contrasts DEGs 概览。

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

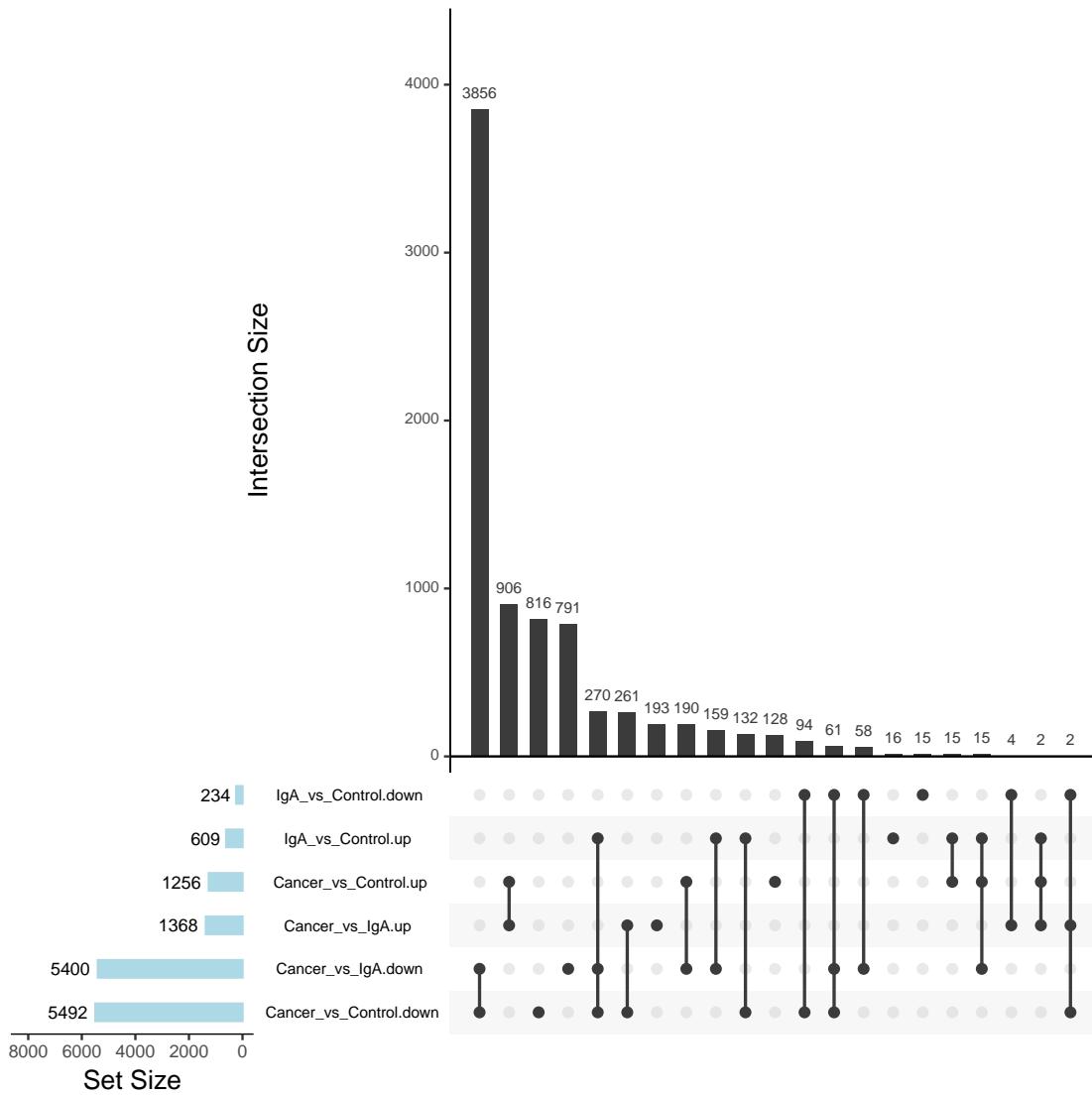


Figure 35: SETS1 intersection of contrasts DEGs

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 36为图 SETS1 GSEA enrichment of KEGG CancerVsIgA 概览。

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

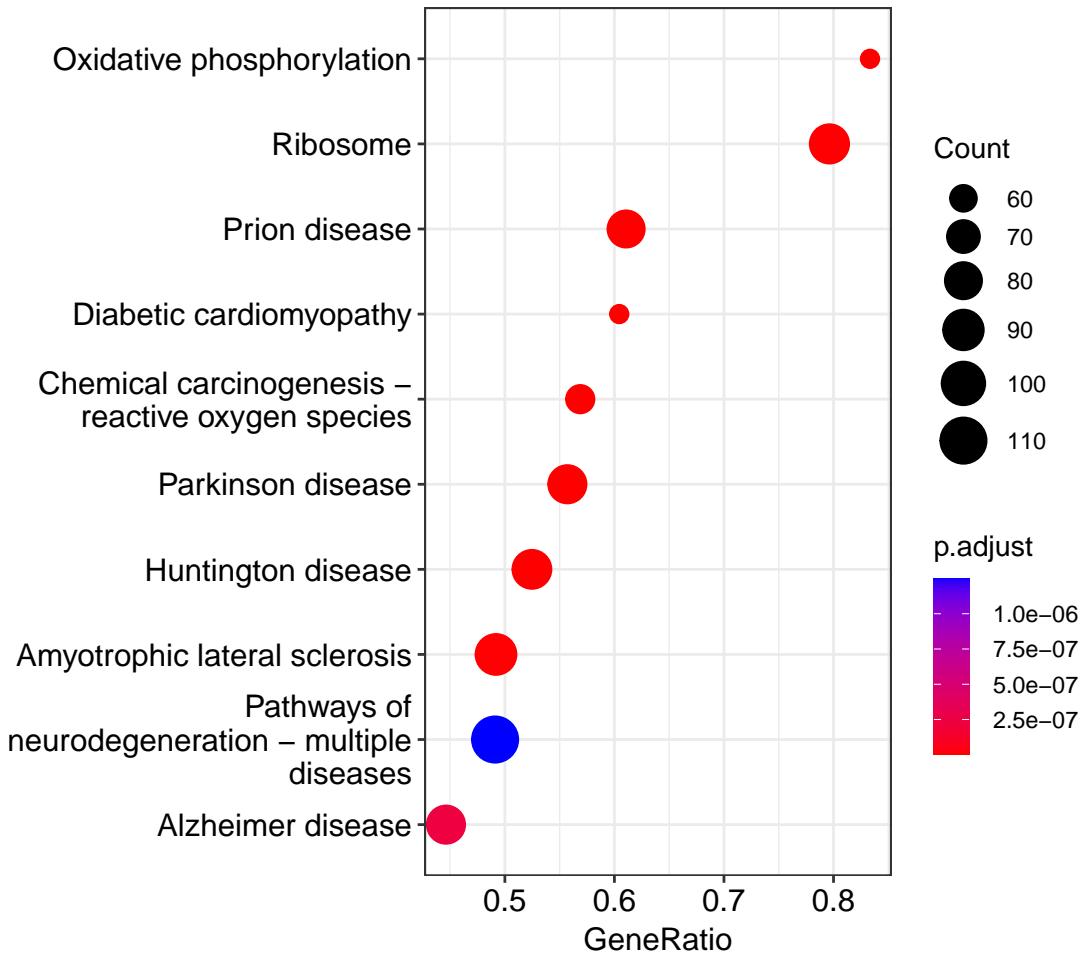


Figure 36: SETS1 GSEA enrichment of KEGG CancerVsIgA

Figure 37为图 SETS1 GSEA enrichment of GO CancerVsIgA 概览。

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

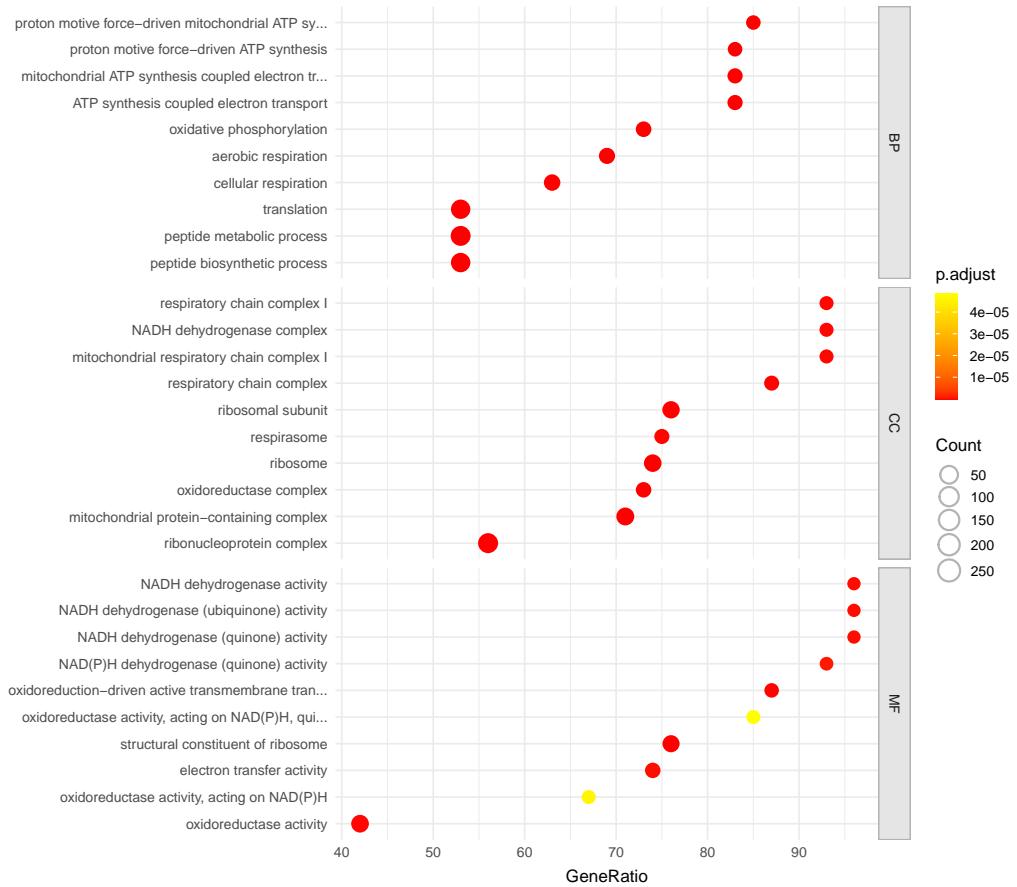


Figure 37: SETS1 GSEA enrichment of GO CancerVsIgA

Fig. ?? 富集于 OXPHOS 通路的基因表达量整体下降。

Figure 38为图 SETS1 GSEA show OXPHOS pathway CancerVsIgA 概览。

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

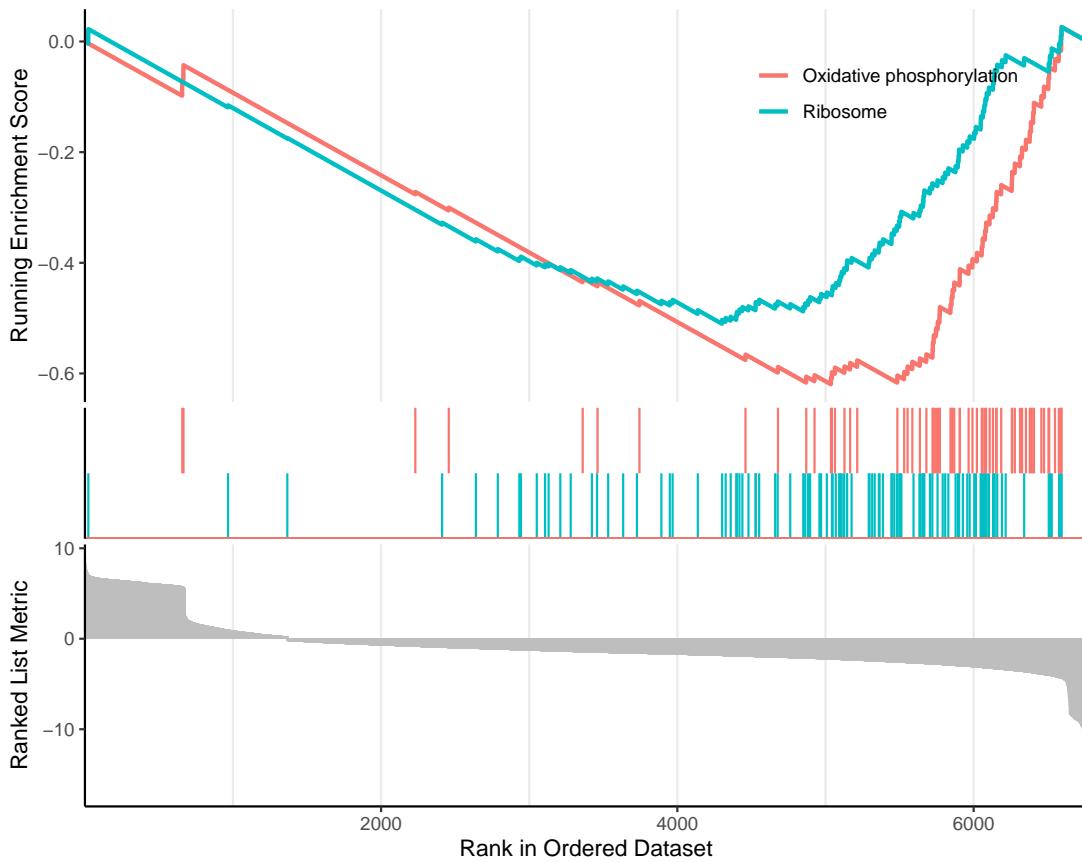


Figure 38: SETS1 GSEA show OXPHOS pathway CancerVsIgA

Figure 39为图 SETS1 GSEA show OXPHOS pathway by pathview CancerVsIgA 概览。

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

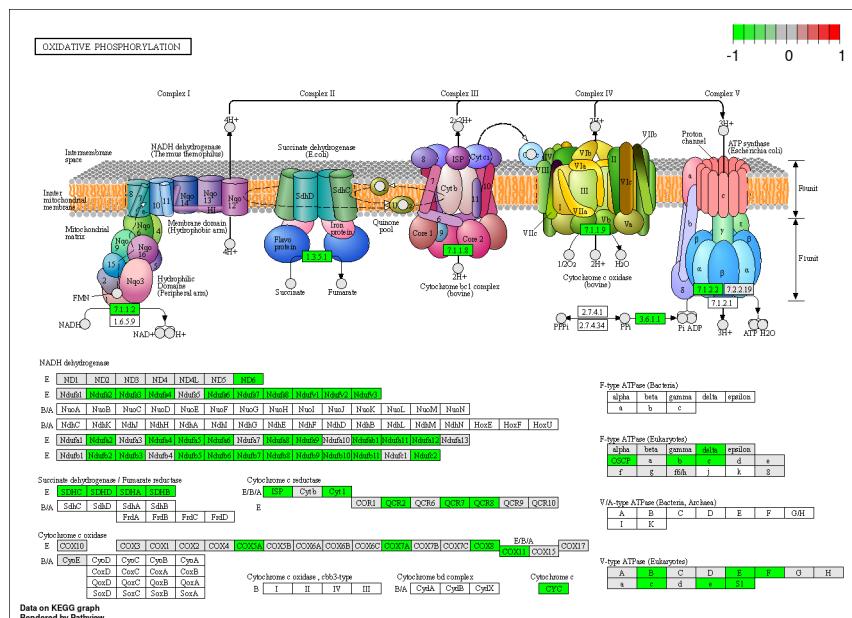


Figure 39: SETS1 GSEA show OXPHOS pathway by pathview CancerVsIgA

7.5.4.2 Cancer vs Control Fig. 40 与 Fig. 36 相比，并不富集于 OXPHOS 通路。

Figure 40为图 SETS1 GSEA enrichment of KEGG CancerVsControl 概览。

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

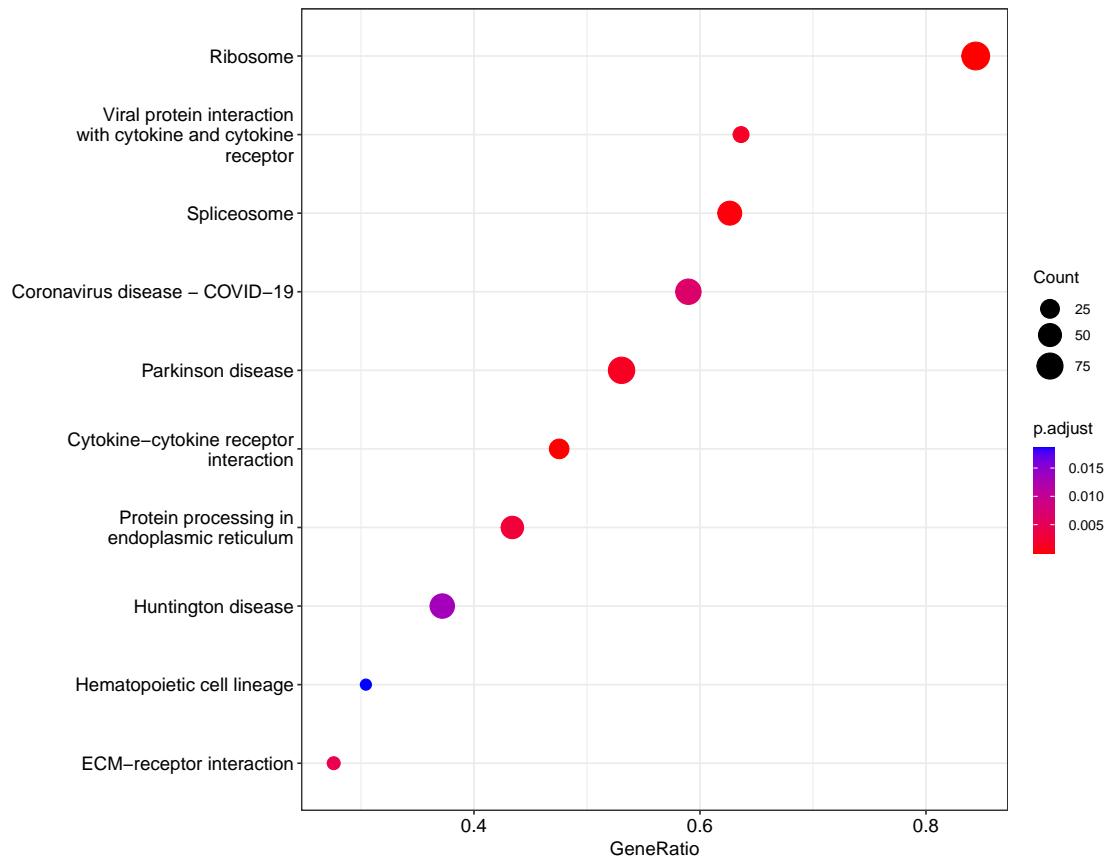


Figure 40: SETS1 GSEA enrichment of KEGG CancerVsControl

Figure 41为图 SETS1 GSEA enrichment of GO CancerVsControl 概览。

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

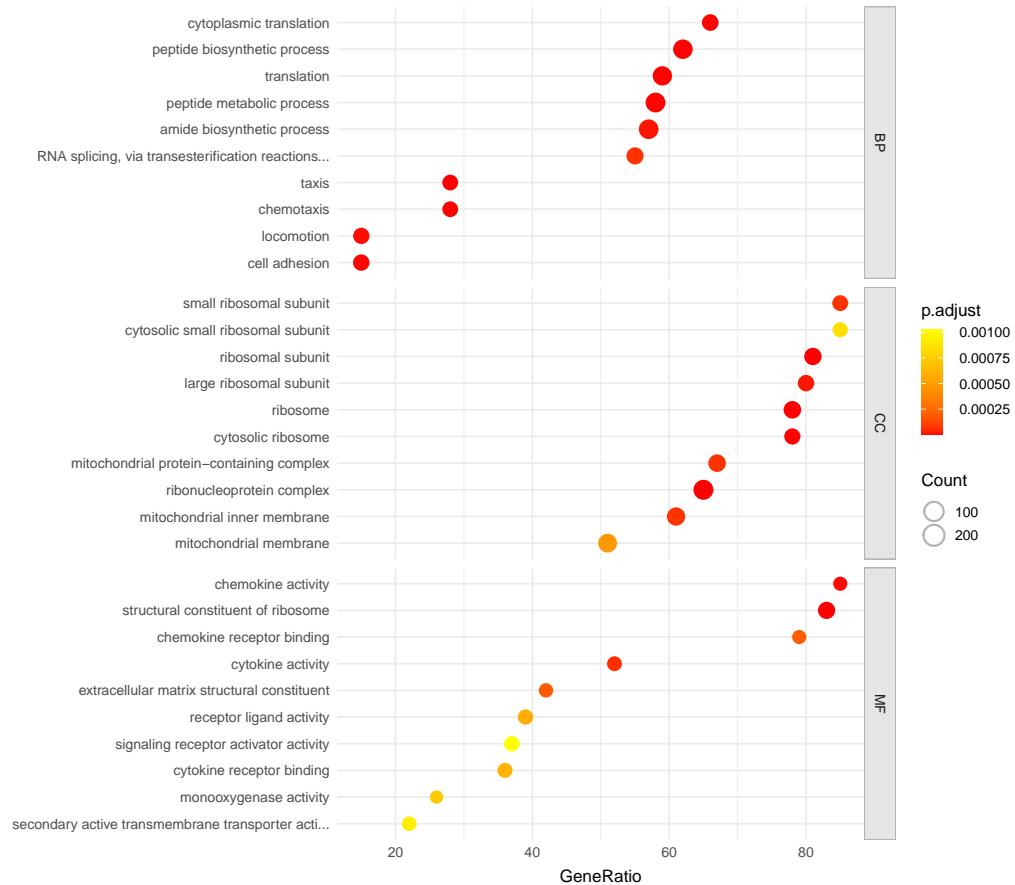


Figure 41: SETS1 GSEA enrichment of GO CancerVsControl

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 42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)

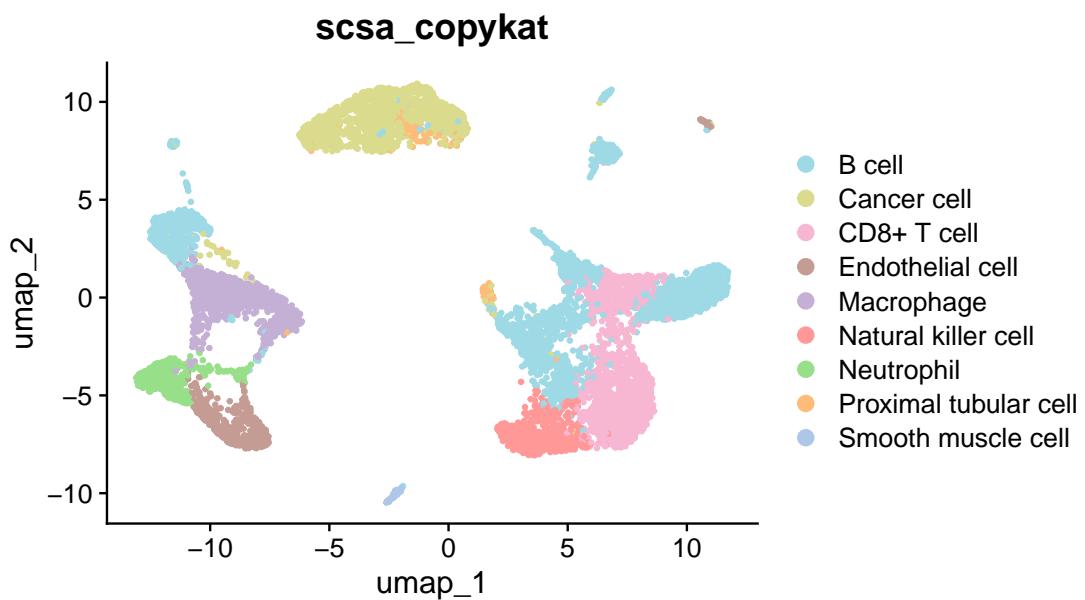


Figure 42: RCC data2 SCSA mapping with copykat prediction

7.6.3 集成验证: Cancer、IgA

7.6.3.1 RISC 集成 Figure 43为图 SETS2 select reference dataset for integration 概览。

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

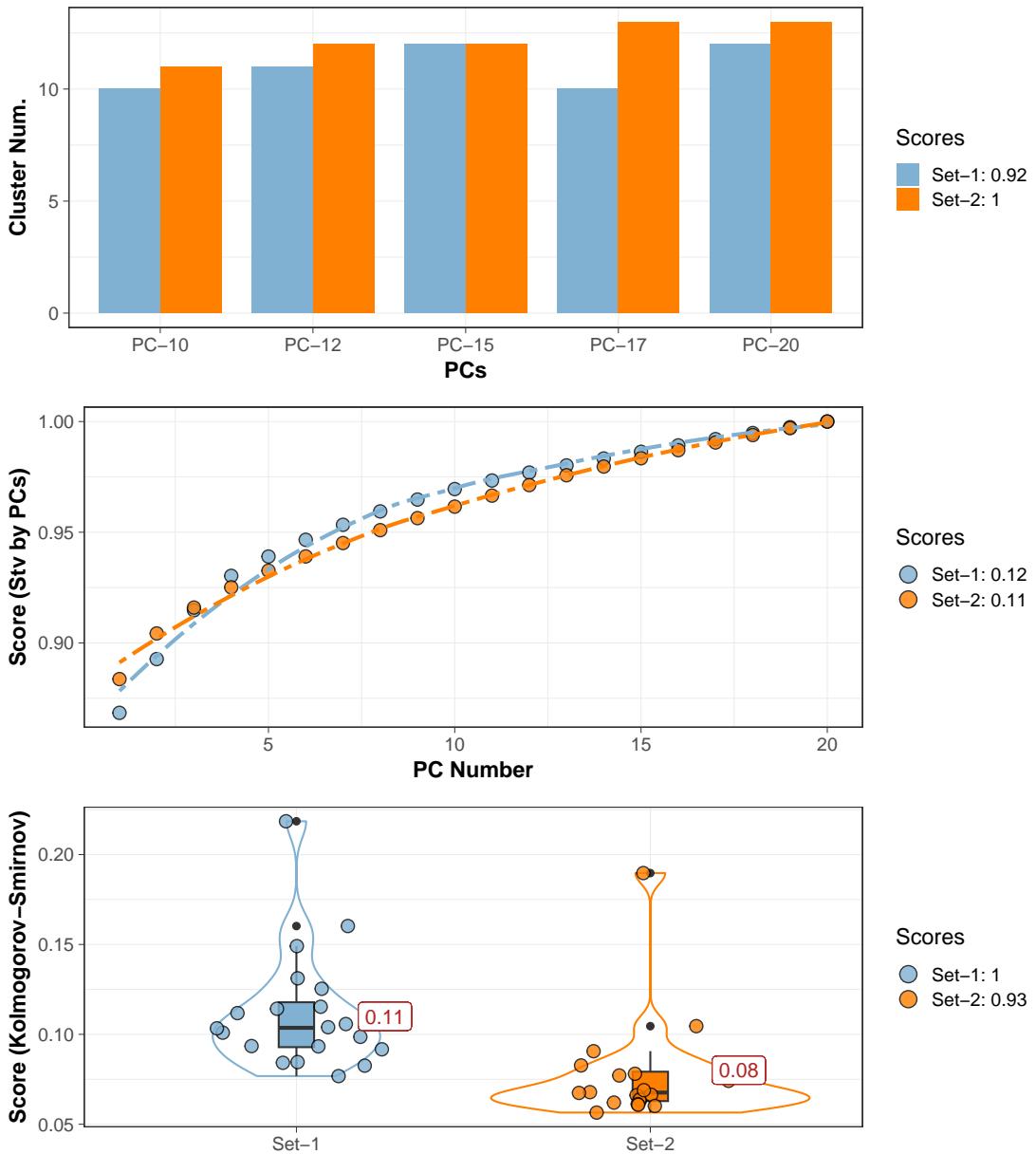


Figure 43: SETS2 select reference dataset for integration

Figure 44为图 SETS2 umap mapping 概览。

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

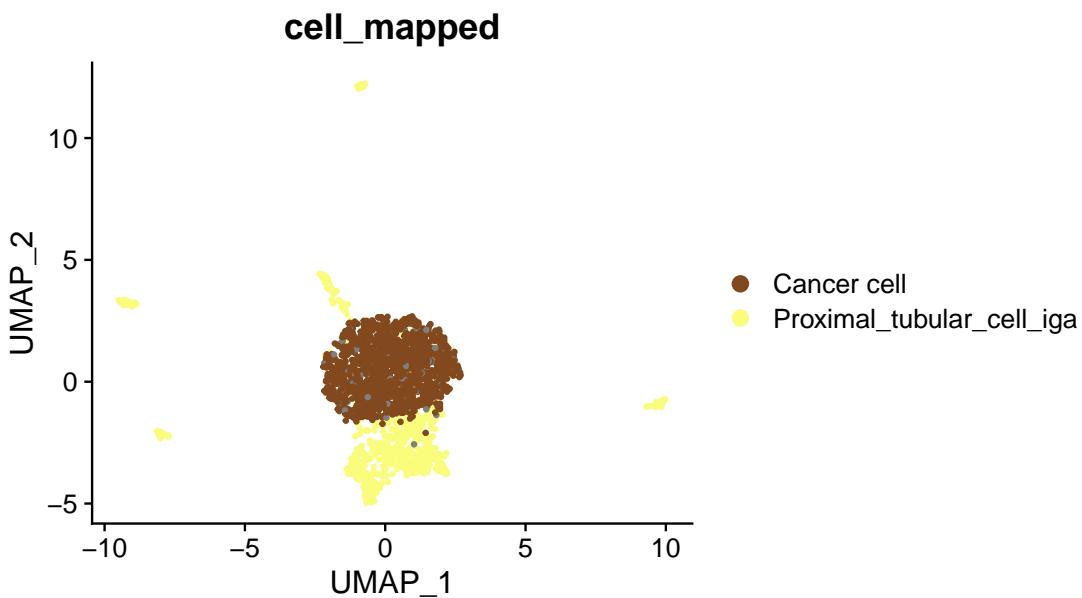


Figure 44: SETS2 umap mapping

7.6.3.2 拟时分析 依据拟时轨迹 Fig. 34, 同 Fig. 33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 45为图 SETS2 pseudotime 概览。

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

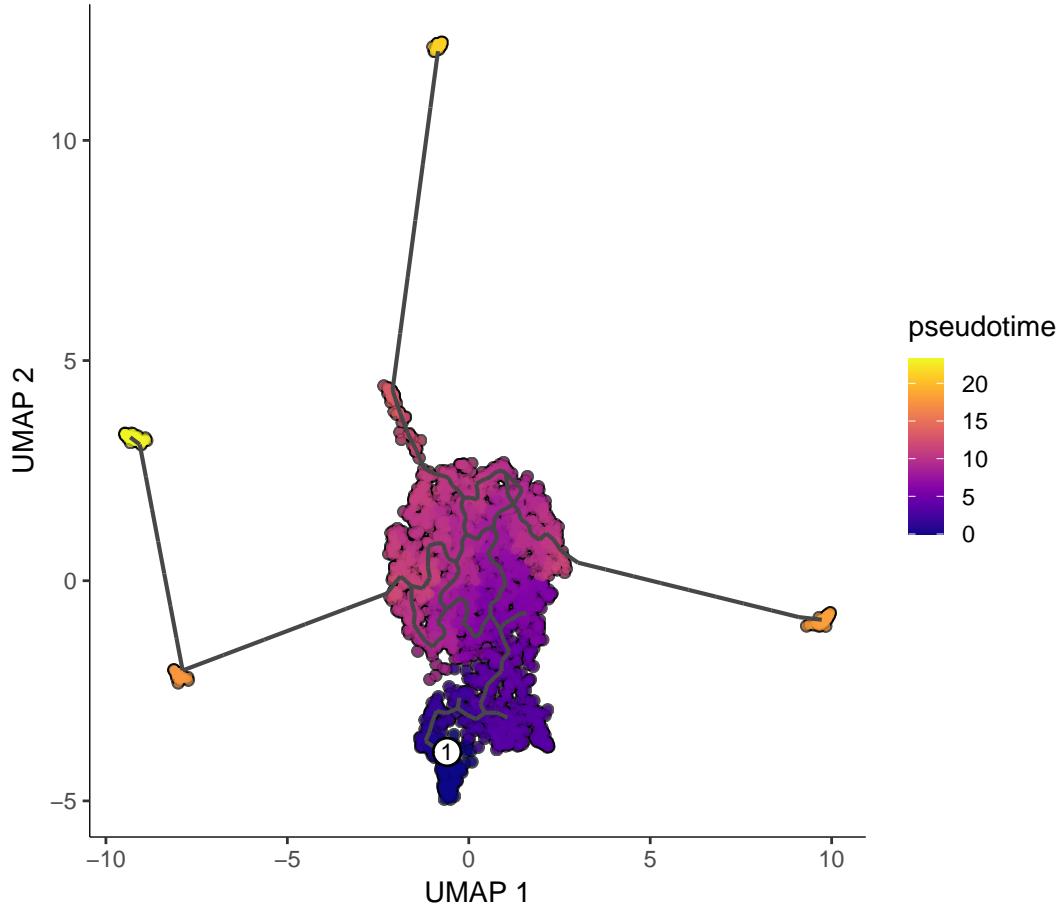


Figure 45: SETS2 pseudotime

7.6.3.3 富集分析 Figure 46为图 SETS2 GSEA enrichment of KEGG CancerVsIgA 概览。
(对应文件为 Figure+Table/SETS2-GSEA-enrichment-of-KEGG-CancerVsIgA.pdf)

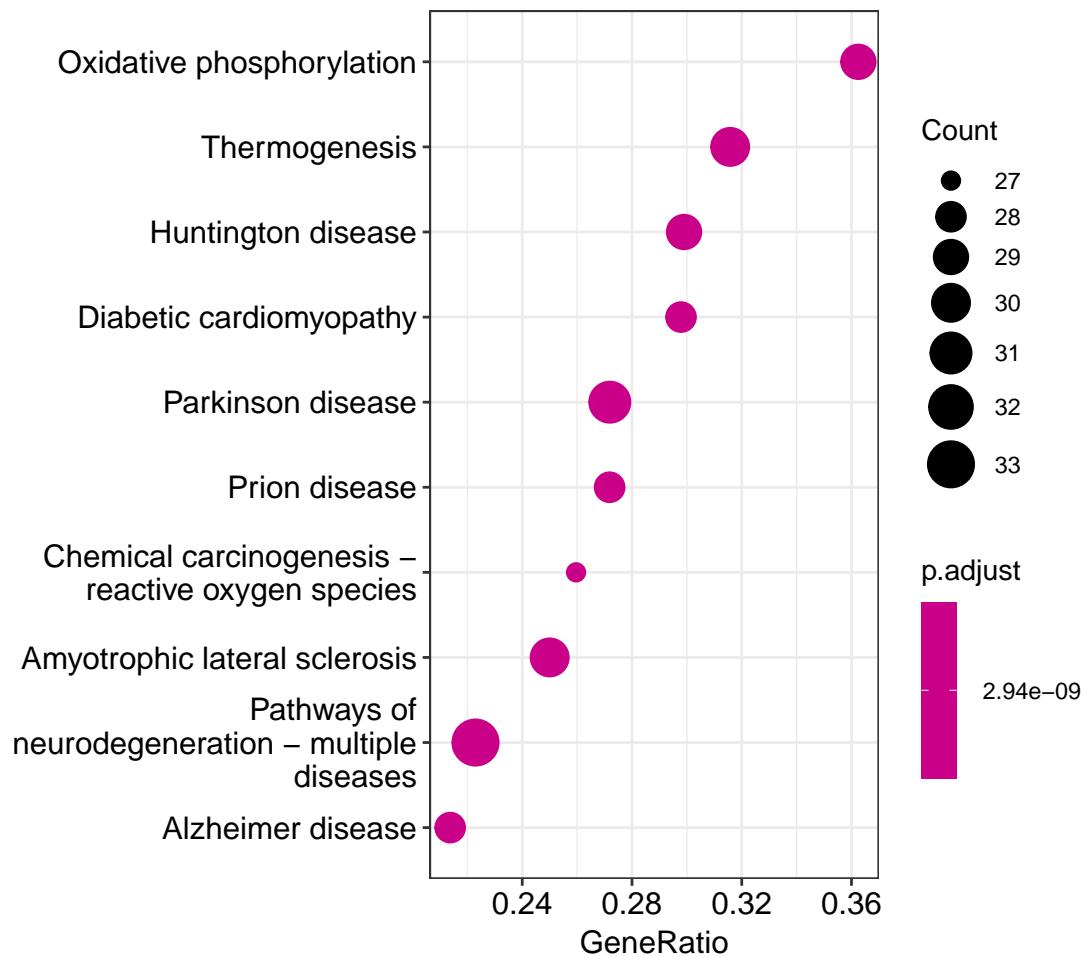


Figure 46: SETS2 GSEA enrichment of KEGG CancerVsIgA

Figure 47为图 SETS2 GSEA enrichment of GO CancerVsIgA 概览。

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

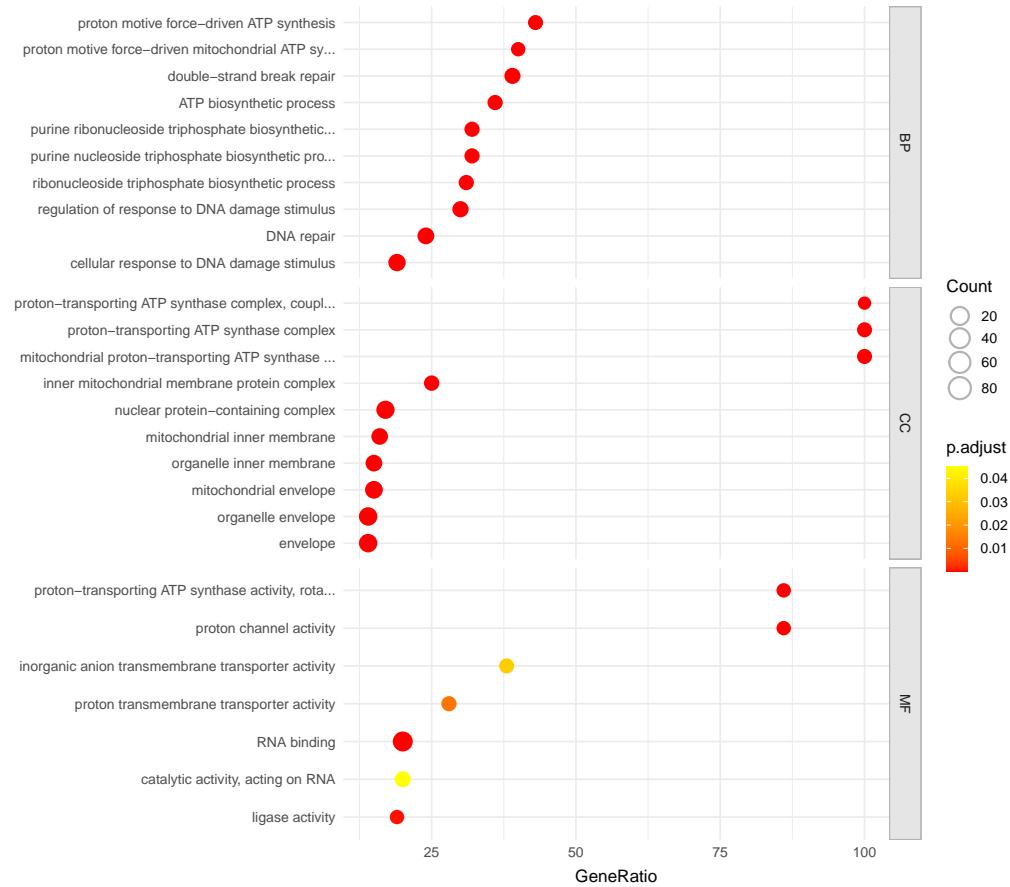


Figure 47: SETS2 GSEA enrichment of GO CancerVsIgA

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¹⁵，RCC 肿瘤可分为 OXPHOS 依赖型和非依赖型代谢增强。

Figure 48为图 SETS2 GSEA show OXPHOS pathway CancerVsIgA 概览。

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

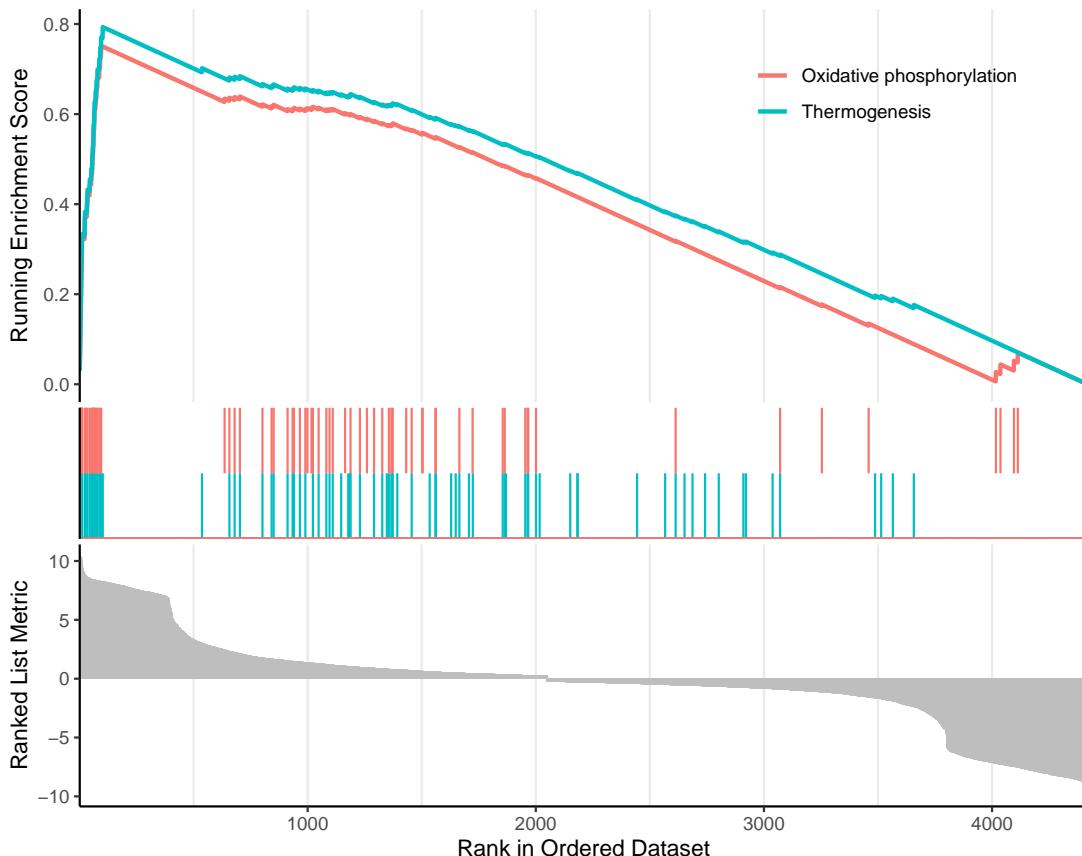


Figure 48: SETS2 GSEA show OXPHOS pathway CancerVsIgA

Figure 49为图 SETS2 GSEA show OXPHOS pathway by pathview CancerVsIgA 概览。

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

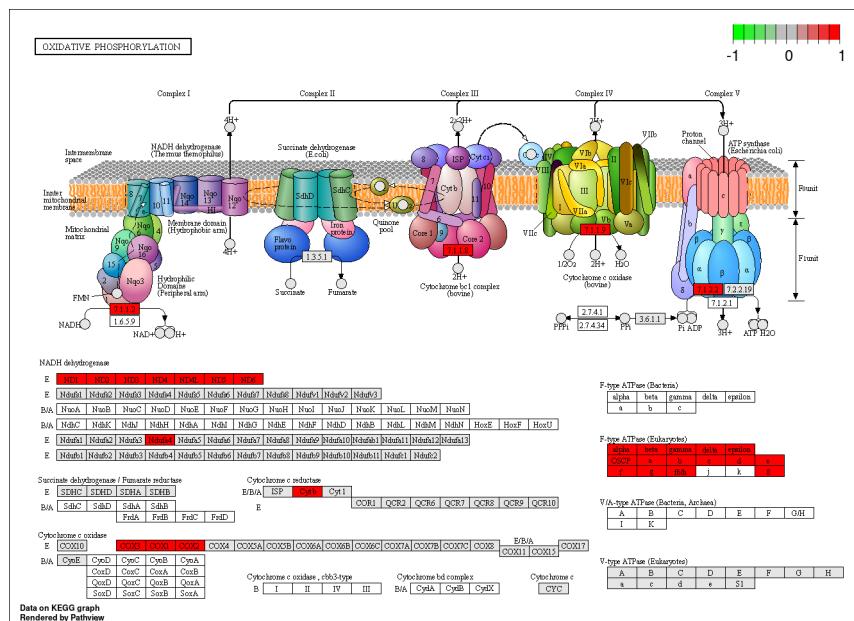


Figure 49: SETS2 GSEA show OXPHOS pathway by pathview CancerVsIgA

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. Gordon, D. J., Resio, B. & Pellman, D. Causes and consequences of aneuploidy in cancer. *Nature Reviews Genetics* **13**, 189–203 (2012).
14. Xu, Y., Xue, D., Bankhead, A. & Neamati, N. Why all the fuss about oxidative phosphorylation (oxphos)? *Journal of medicinal chemistry* **63**, 14276–14307 (2020).
15. Sancho, P., Barneda, D. & Heeschen, C. Hallmarks of cancer stem cell metabolism. *British journal of cancer* **114**, 1305–1312 (2016).