通元医学检验实验室
Tongyuan Clinical Laboratories

格物致知 识微通元

单细胞转录组分析结题报告

通元医学检验实验室

content

content 2
单细胞转录组分析结题报告 3
1 前言 3
2 实验流程 4
3 信息分析流程 5
4 细胞基因表达QC 6
4.1 细胞分布统计作图 7
5 细胞分群降维 8
5.1 cluster分群展示 9
5.2 tSNE降维分群展示 9
5.3 UMAP降维分群展示 10
6 细胞类型自动化注释 11
6.1 新格元SynEcoSys单细胞数据库展示 11
6.2 CellID软件分析展示 12
6.3 自动化注释细胞类型结果展示 13
7 富集分析 16
7.1 GO功能富集 17
7.2 KEGG通路富集分析 19
8 细胞互作分析 21
9 备注 22
9.1 结果文件解压方法 22
9.2 结果文件格式说明 23
9.3 主要分析软件列表及版本 23
10 参考文献 24

单细胞转录组分析结题报告

合同信息	合同内容
项目编号	P23032302
项目名称	单细胞转录组建库测序标准分析报告
样本名称	A140420
物种类型	mouse
组织类型	Kidney
分析流程	CeleScoot-1.0.0

Tips:表格中的组织类型参数用于自动化注释匹配数据库信息，若组织类型为all，则表明使用数据库中所有组织细胞类型进行该项目的注释。因为组织细胞的异质性和相似性，可能会存在一定的误差；若组织类型为具体组织类型，自动化注释会使用数据库中该类型的训练数据集进行细胞注释。因为项目组织类型可能存在特殊性，细胞注释结果仅供参考。

1 前言

高通量测序发展至今，转录组测序已经被广泛的应用于生物生理机制研究、遗传育种以及病理分析。传统的RNA-seq(Bulk RNA-seq)仅能反应大量细胞集合的基因表达情况，需要较高的建库起始量。并且对于一些细胞异质性较高的组织（神经细胞、肿瘤组织等）Bulk RNA-seq 并不能充分的揭示其机制，需要从单细胞层次去研究基因表达的变化以及细胞分化机制。

利用GEXSCOPE®微流控芯片捕获单细胞，并将数百万个携带独特细胞标签（Cell Barcode）的Barcoding Beads加入到芯片微孔中，确保每个微孔内只落入1个Barcoding Beads。细胞裂解后，带有独特细胞标签（Barcode）及分子标签（UMI）的Barcoding Beads通过与 mRNA上的 poly（A）尾结合捕获mRNA，对细胞及 mRNA 进行标记。收集芯片中的Barcoding Beads，将Barcoding Beads捕获的 mRNA 反转录为 cDNA 并扩增。将 cDNA 经过片段化、连接接头等步骤后构建适用于illumina测序平台的测序文库。可实现海量单细胞的快速高效标记、文库构建，应用于绘制单细胞表达图谱、进而刻画生命发育与疾病发生发展

机制。

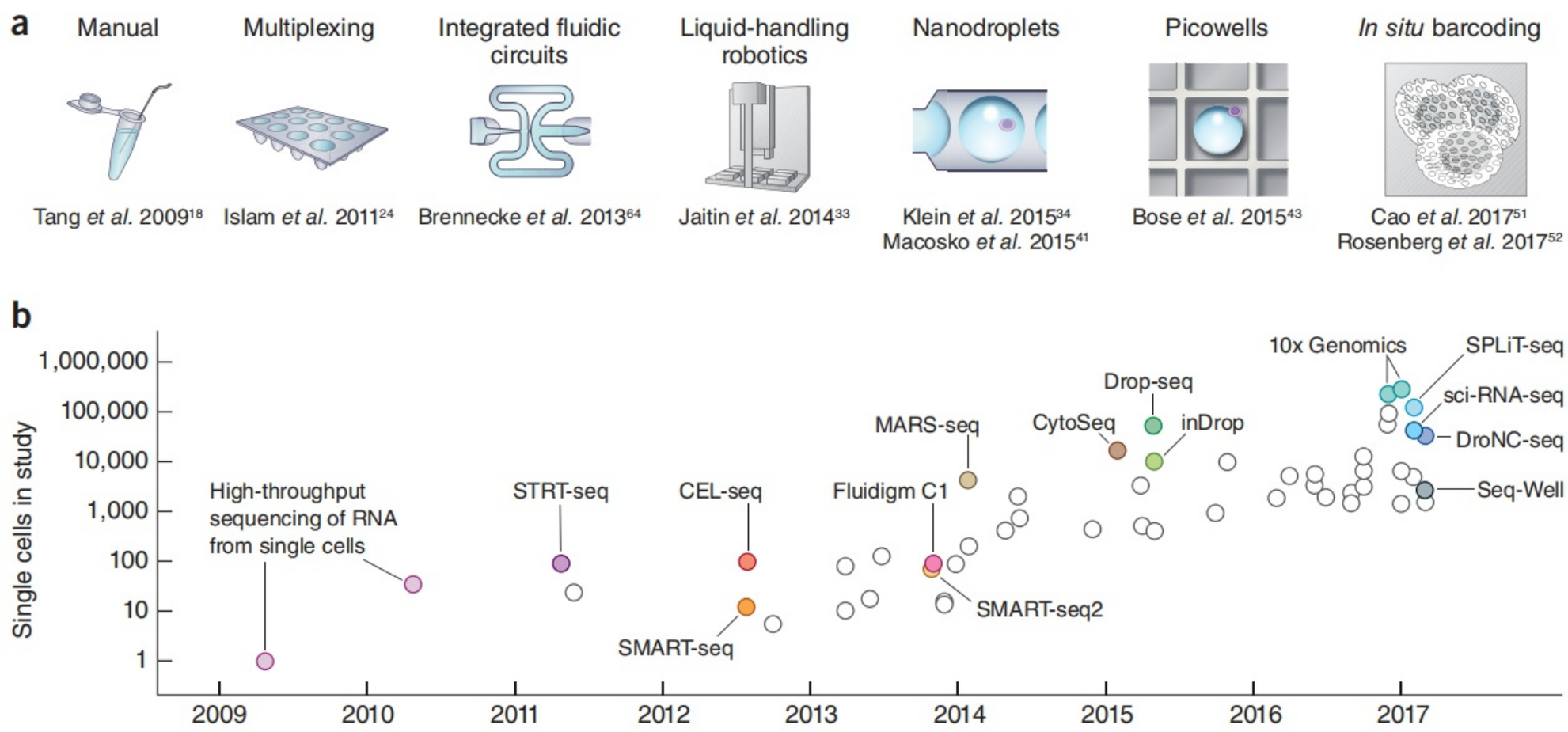


Figure1：上图为单细胞RNA-seq技术发展史。

单细胞转录组的应用：

- 单细胞转录组分析可以解决细胞异质性的问题
- 单细胞转录组分析可以发现新的细胞亚群
- 单细胞转录组分析可以构建完整的细胞谱系

单细胞转录组分析目前已经应用于肿瘤学、干细胞、发育生物学、免疫学、神经生物学、生殖医学等研究领域。

---

2 实验流程

具体实验流程如下。根据样本情况不同，整体流程会有所调整:

- （1）单细胞（核）悬液制备与质检 使用sCelLiVE™组织解离液（核分离液）将组织样品解离成单细胞（核）悬液，将细胞（核）稀释到合适浓度，一般是2.5-3.5×10^5 cells/mL（3-4×10^5 nucleus/mL）。
- （2）单细胞（核）分离与标记 细胞（核）悬液注入SCOPE-chip™微流控芯片，根据”泊松分布”的原理完成单个细胞（核）的分离。细胞（核）在重力作用下落入特殊定制的芯片微孔中， 确保每个微孔内只落入1 个细胞（核）。之后将数百万个携带独特细胞标签（Cell Barcode）的磁珠加入到芯片微孔中，确保每个微孔内只落入1 个磁珠。细胞（核）裂解后，带有独特细胞标签（Barcode）及分子标签（UMI）的磁珠通过与mRNA上的poly(A)尾结合捕获mRNA，对细胞（核）及mRNA 进行标记。
- （3）反转录与扩增 收集芯片中的磁珠，将磁珠捕获的mRNA 反转录为cDNA 并扩增。
- （4）单细胞（核）测序文库构建 将cDNA 经过片段化、连接接头等步骤后构建适用于Illumina 测序平台的测序文库。

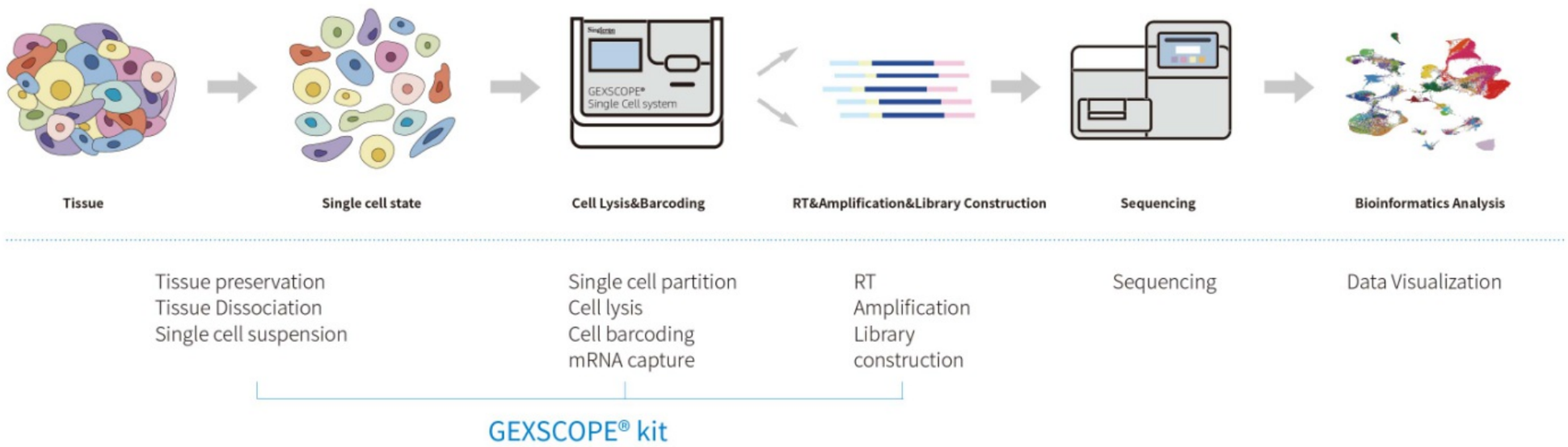


Figure2：上图为实验过程示意图。

3 信息分析流程

对于单细胞转录组文库的分析，首先使用质控分析软件对样本进行基因表达定量和细胞数目鉴定分析，然后使用标准分析流程进行后续分析，该结题报告仅展示标准分析流程结果。如下所示：

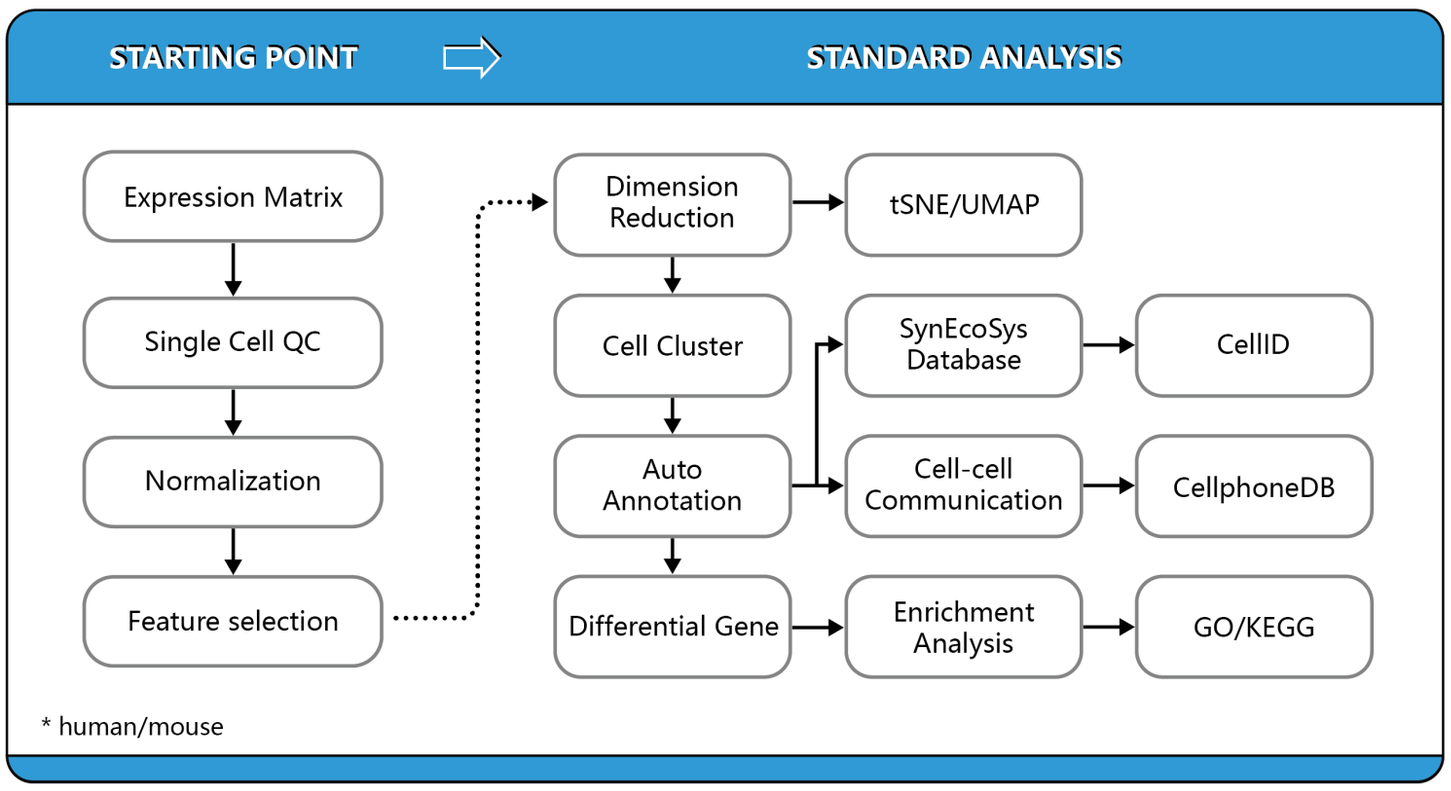


Figure3.0.1：上图展示了适用于人和小鼠类单样本的单细胞分析流程。

如流程图所示，对于单个样本主要进行数据质控分析（按照一定阈值去除doublet以及低质量的细胞）、细胞基因表达QC（展示nFeatureRNA、nCountRNA、mt.percent等）、细胞分群cluster、细胞分型降维展示(tSNE/UMAP)、自动化注释细胞类型（仅用于人鼠样本数据库中包含的组织类型）、细胞类型差异基因分析并展示细胞类型差异基因、细胞互作分析（推断不同细胞类型之间的相互作用）、富集分析（包括GO、KEGG）等。

此外对于多个样本主要进行数据整合，包括质控分析，单细胞质控指标展示，细胞分群cluster、细胞分型降维展示(tSNE/UMAP)、自动化注释细胞类型（仅用于人鼠样本数据库中包含的组织类型）、细胞类型差异基因分析、互作分析以及富集分析等。 多样本分析需要在单样本分析完成后，并且确定加测完成后，运行样本组合分析流程并产生分析报告。（不在此报告中）

4 细胞基因表达QC

原始数据通过CeleScope产品进行质控分析得到表达矩阵后，该标准分析流程会先根据细胞中基因数目、细胞中UMI数目、细胞中线粒体基因含量分布比例等去除异常的细胞后再进行后续分析。

CeleScope产品流程及使用说明可参考链接 (https://github.com/singleron-RD/CeleScope)。

4.1 细胞分布统计作图

4.1 细胞分布统计作图

首先进行细胞中基因数目（n_genes_by_counts）、细胞中UMI数目（total_counts）、细胞中线粒体基因含量（percent_mt）分布比例作图。下图展示了过滤前后的样本QC质控情况：

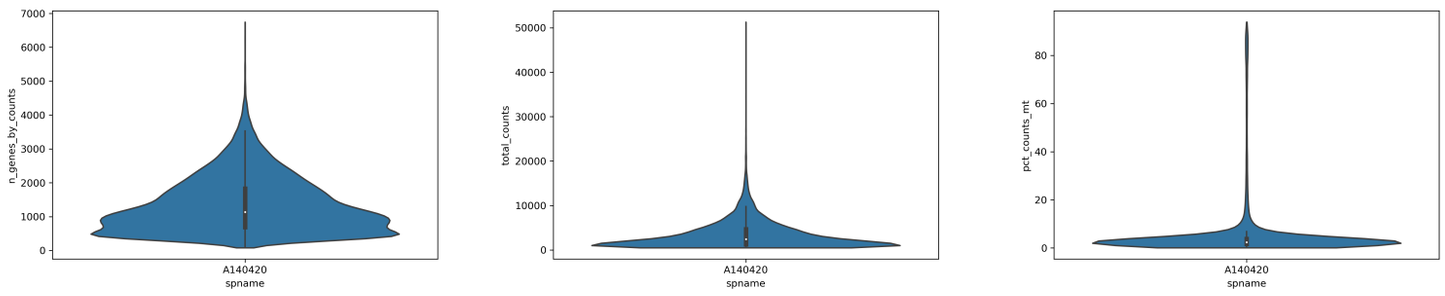


Figure3.1.1：该图为样本质量过滤之前的展示，分为3个部分,第一幅为细胞基因数目分布, 第二幅为细胞UMI数目分布, 第三幅图为细胞线粒体基因含量分布。内部Violin图表示该数目下的密度分布。

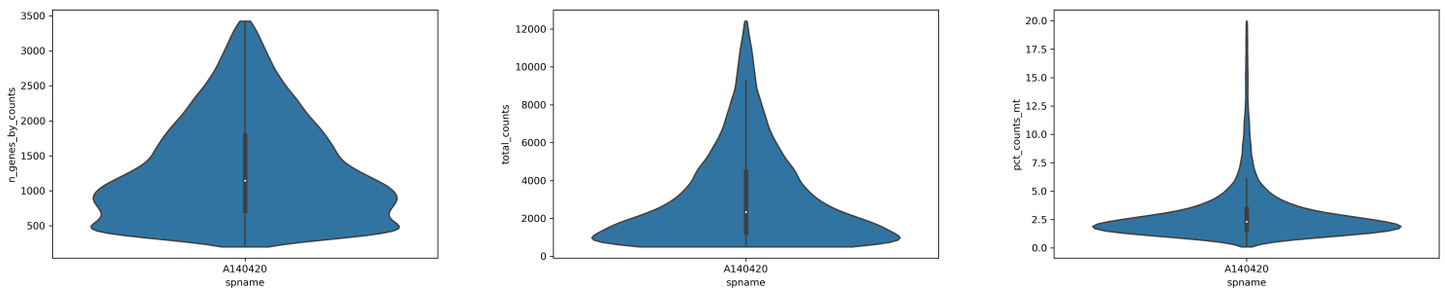


Figure3.1.2：该图为样本质量过滤之后的展示，分为3个部分,第一幅为细胞基因数目分布, 第二幅为细胞UMI数目分布, 第三幅图为细胞线粒体基因含量分布。内部Violin图表示该数目下的密度分布。

样本指标	具体数值
Estimated Number of Cells	21,956
Fraction Reads in Cells	78.65%
Mean Reads per Cell	14,479
Median UMI per Cell	2,429
Total Genes	26,637
Median Genes per Cell	1,138

Table1.1：单细胞部分质控信息，具体可看质控报告。

线粒体阈值	指标
Fraction of cells have mito gene percent>5%	16.18%
Fraction of cells have mito gene percent>10%	7.67%
Fraction of cells have mito gene percent>15%	5.83%
Fraction of cells have mito gene percent>20%	4.93%
Fraction of cells have mito gene percent>30%	3.95%
Fraction of cells have mito gene percent>50%	2.98%

Table1.2：样本线粒体含量分布情况，表示线粒体基因含量高于一定阈值的细胞占比。点击查看详细的质控结果：web_summary.html (src/web_summary.html)

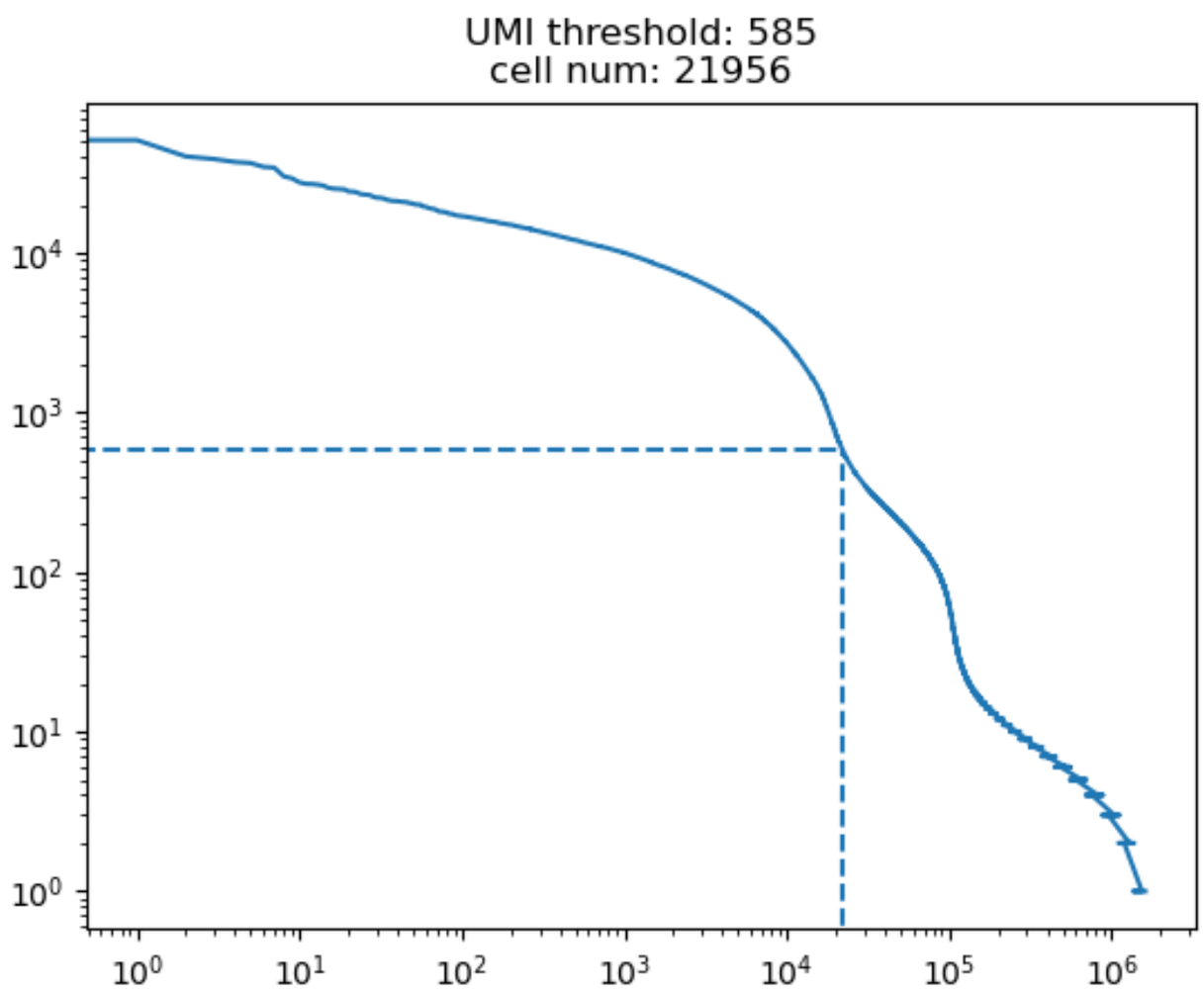


Figure3.1.3：以上曲线图显示的是barcode的计数分布。Y轴是map到每一个barcode的UMI的计数数值，X轴是对应到这个数值的barcode的数量。

该部分具体结果可见result文件夹中01.multi

5 细胞分群降维

为了进一步了解细胞之间的异质性，Scanpy 基于样本基因表达量对细胞进行分群。基本思路：首先按照

一定阈值过滤低质量的细胞、基因表达过多的细胞以及线粒体含量较高的细胞。其次，进行归一化并通过主成分分析（PCA）降低数据维度，选择前50个维度采用louvain 聚类算法进行细胞分群。 最后，进行tSNE与UMAP降维分析用于可视化。

5.1 cluster分群展示

按照上述步骤，展示cluster分群后每个cluster的细胞数目和占比情况：

A140420

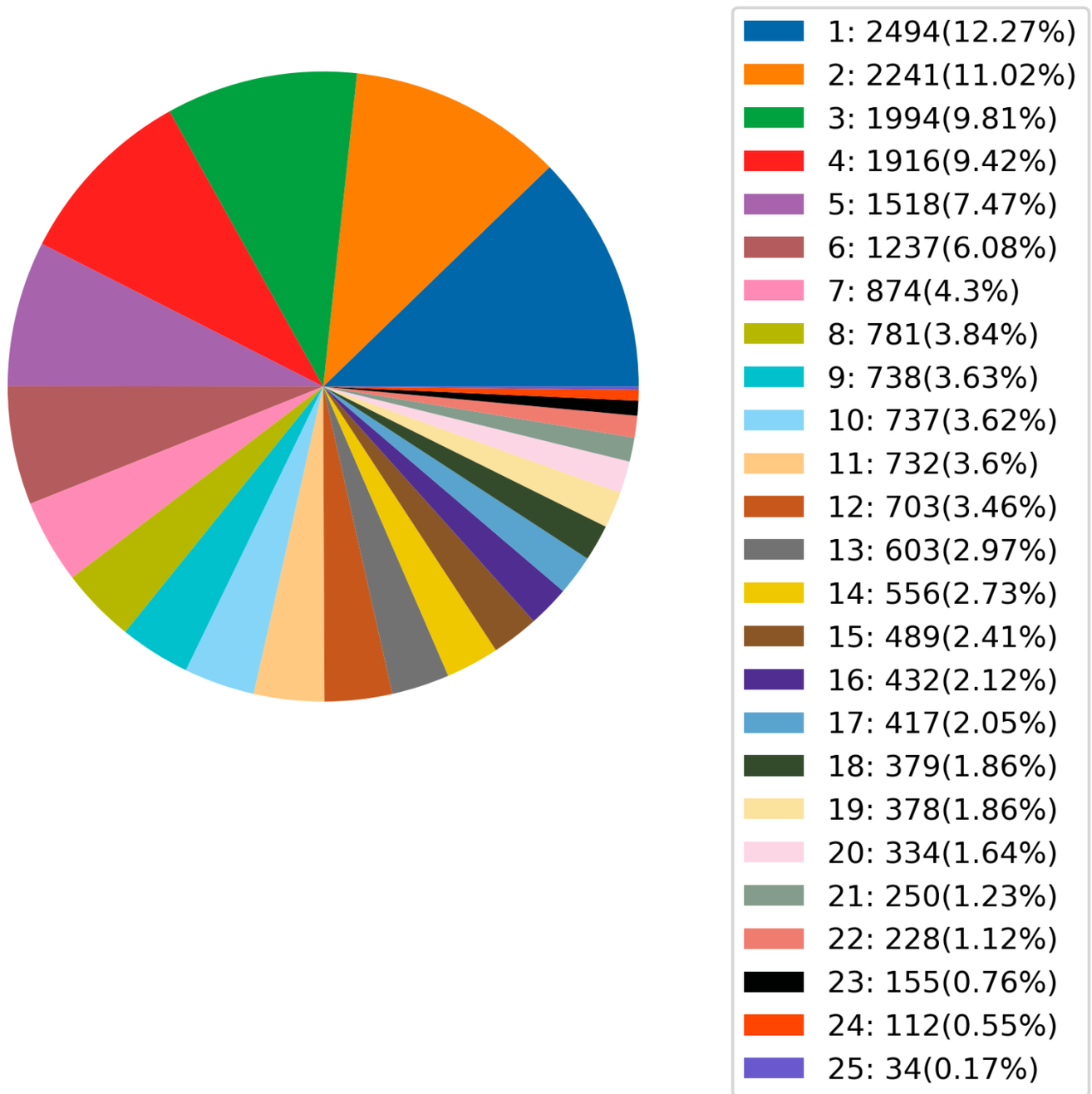


Figure3.2.1：展示不同cluster的细胞数目和占比饼图。

5.2 tSNE降维分群展示

单细胞数据降维分析的过程，主要是将高维空间的结构特征投影到低维空间中，就是用平面或立体空间内的点的疏密远近表现其在原本多维度状态下的疏密远近。在此首先进行tSNE降维分析后，展示cluster分群展示结果如下：

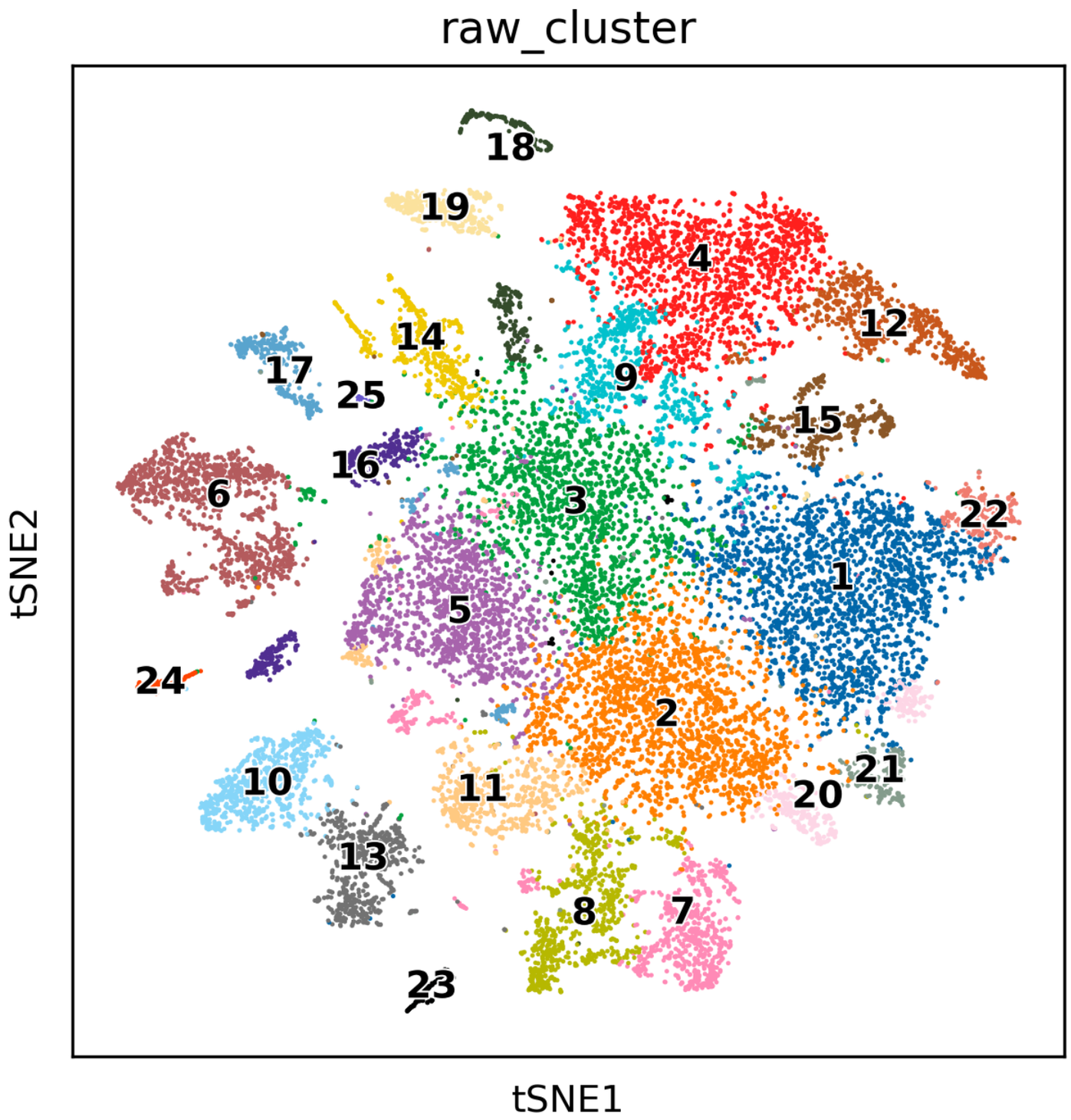


Figure3.2.2：对于各个样本tSNE分析后的坐标信息不在此展示，详见结果文件。在用tSNE降维单细胞数据时，经常会发现同一类细胞被其他细胞分隔。这是因为其损失函数（KL散度）对低维近、高维远的惩罚较轻，导致在平面上，整体差异较小的集群（cluster）可能比差异较大的集群离得更远。故而tSNE图多数情况下不能体现真实的全局结构。

5.3 UMAP降维分群展示

不同于tSNE降维算法，UMAP算法只计算各点与最近k个点之间的距离，严格限制局部的范围；而且两种算法在对信息损失的计算方法不同，tSNE使用KL散度衡量信息损失，在全局结构上存在失真的可能，而UMAP使用二元交叉熵，全局和局部结构均有保留。在此同时使用UMAP降维分析，cluster分群展示结果如下：

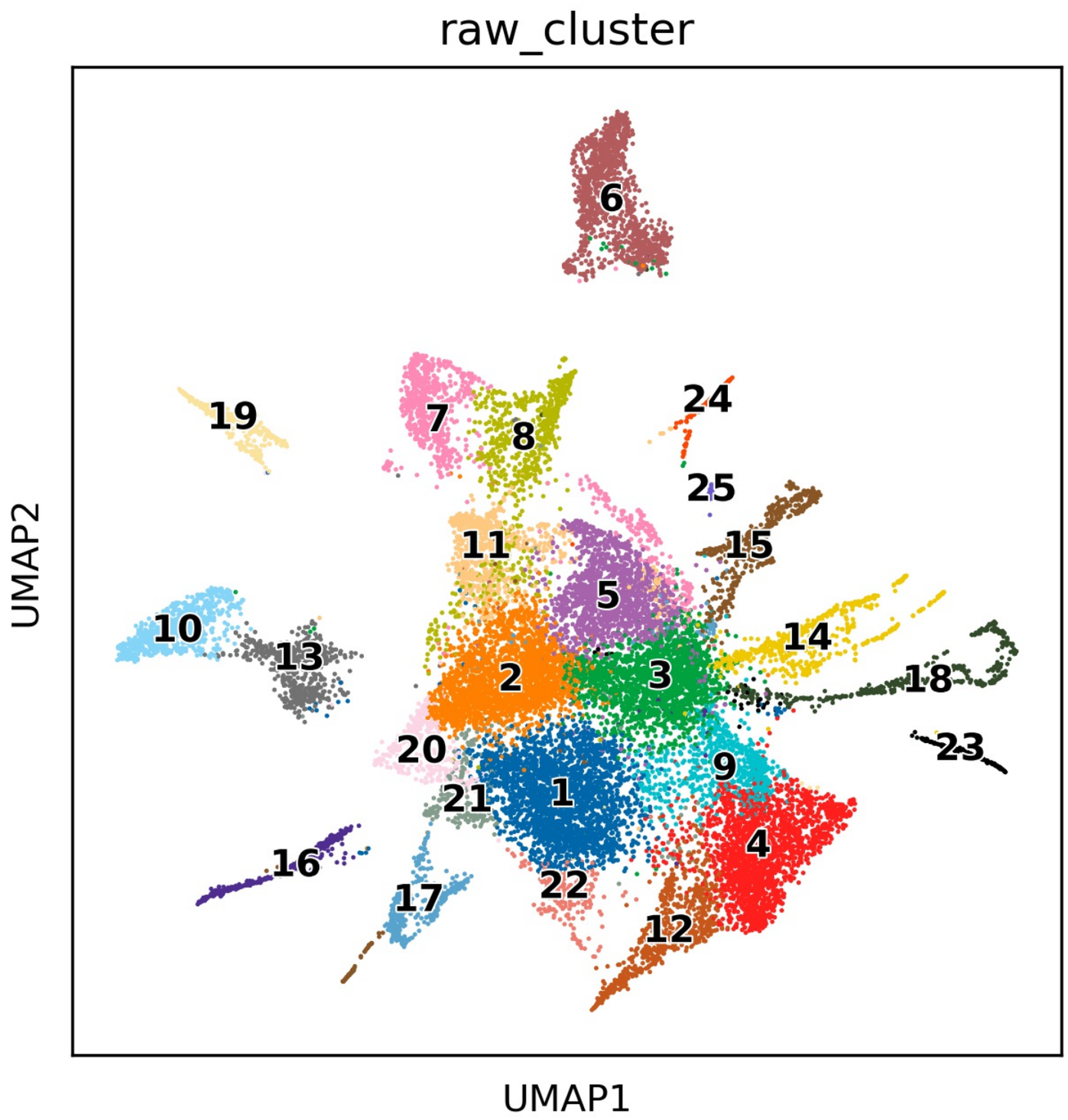


Figure3.2.3：对于各个样本UMAP分析后的坐标信息不在此展示，详见结果文件。UMAP损失函数使用的是二元交叉熵，对低维近高维远或低维远高维近的惩罚都较重，所以UMAP比tSNE更能体现真实的全局结构。

该部分具体结果可见result文件夹中01.multi

6 细胞类型自动化注释

CellID使用MCA和HGT算法，基于新格元SynEcoSys单细胞数据库，使用大量人工校验公共数据集训练的参考数据信息。对项目中的人、小鼠的几乎所有的组织、器官进行自动化注释。

6.1 新格元**SynEcoSys**单细胞数据库展示

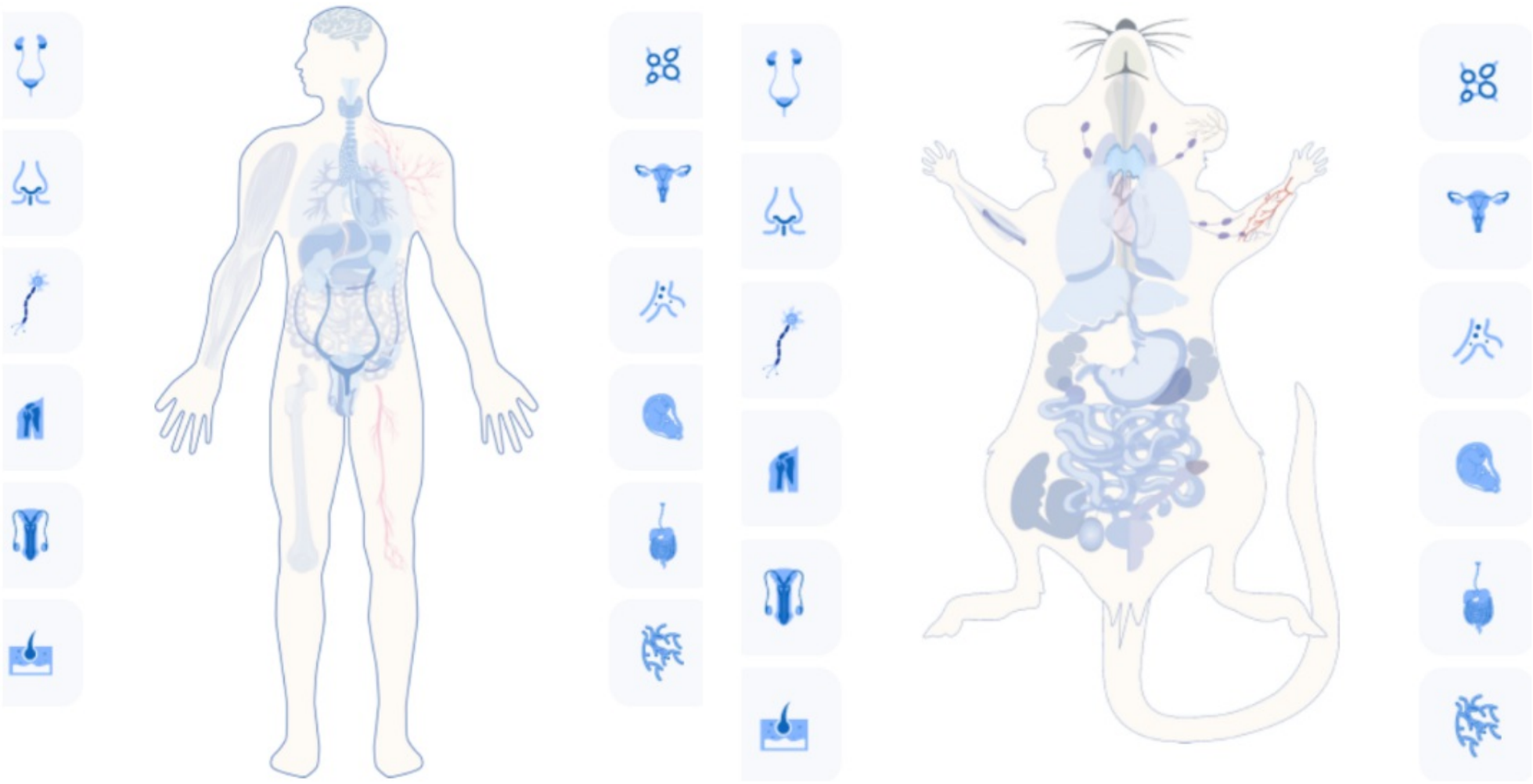


Figure3.3.1：海量的Marker基因以及丰富的数据集是新格元自动化注释的基础。SynEcoSys拥有人和小鼠各器官细胞类型几百余种，涉及Marker基因数千个，并且对细胞图谱进行注释层级细分和统一细胞类型命名，实现细胞类型注释标准化。

6.2 CellID软件分析展示

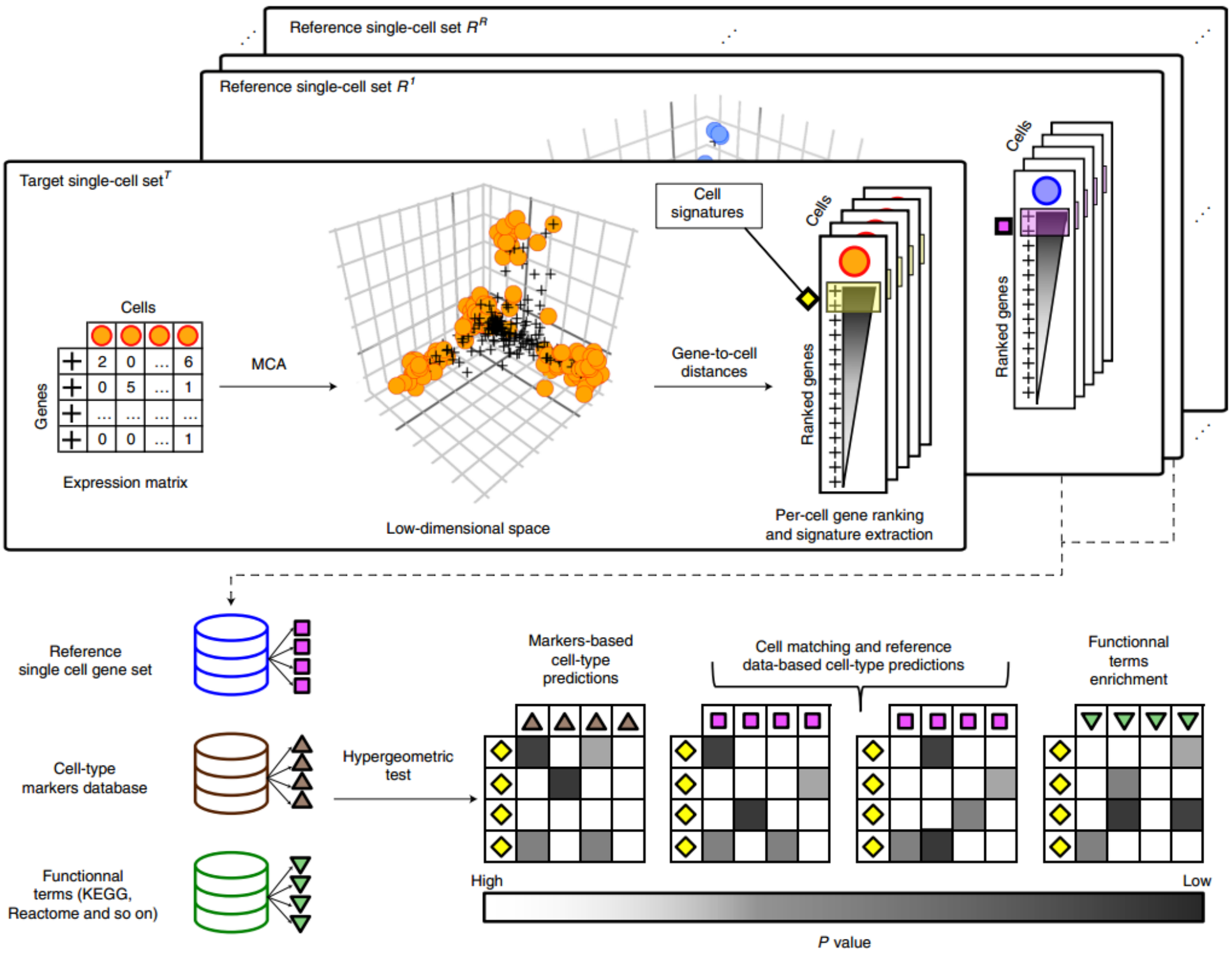


Figure3.3.2: CellID是一款可以用于单个细胞特征提取与细胞身份识别的软件，与新格元的SynEcoSys单细胞数据库结合，可实现对各种细胞类型快速、准确的自动注释。 如上图所示，CellID首先采用多重对应分析（MCA）算法，将矩阵中所有细胞与基因投影到低维空间，从而算出每个细胞与所有基因的距离，选择每个细胞距离最近的一部分基因作为细胞的特征基因；另一方面，从大量人工注释的单细胞数据训练出每个细胞类型的特征基因集合作为参考数据库，最后，采用超几何检验（HGT）算法，将每个细胞的特征基因与数据库中的细胞类型特征基因进行打分匹配，得分最高的细胞类型即为该细胞的注释信息。 由于算法本身的特点，经我们测试，该自动注释方法具有快速、准确、有效过滤污染、批次效应、能识别潜在的肿瘤细胞与新细胞等诸多优点。新格元的SynEcoSys单细胞数据库已经包含了脑、肺、胰腺、肠、骨髓等几乎所有的组织、器官的自动注释参考数据，对物种的支持上已经包含了人、小鼠等常见的物种，并且算法与数据库还在不断更新优化中。

6.3 自动化注释细胞类型结果展示

使用CellID确定细胞类型后，展示降维图及细胞类型与cluster之间的关系。报告首页表格中的组织类型参数用于自动化注释匹配数据库信息，若组织类型为all，则表明使用数据库中所有组织细胞类型进行该项目的注释。因为组织细胞的异质性和相似性，可能会存在一定的误差；若组织类型为具体组织类型，自动化注释会使用数据库中该类型的训练数据集进行细胞注释。因为项目组织类型可能存在特殊性，细胞注释结果仅供参考。

analysis_name	standard_name
TCells	T cells
MPs	Mononuclear phagocytes
PlasmaCells	Plasma cells
Fibroblasts	Fibroblasts
ECs	Endothelial cells
BCells	B cells
EpithelialCells	Epithelial cells

Table1.3：以上表格展示了细胞类型分析名称和标准名称的对应关系, 后续分析都基于分析名称进行展示。

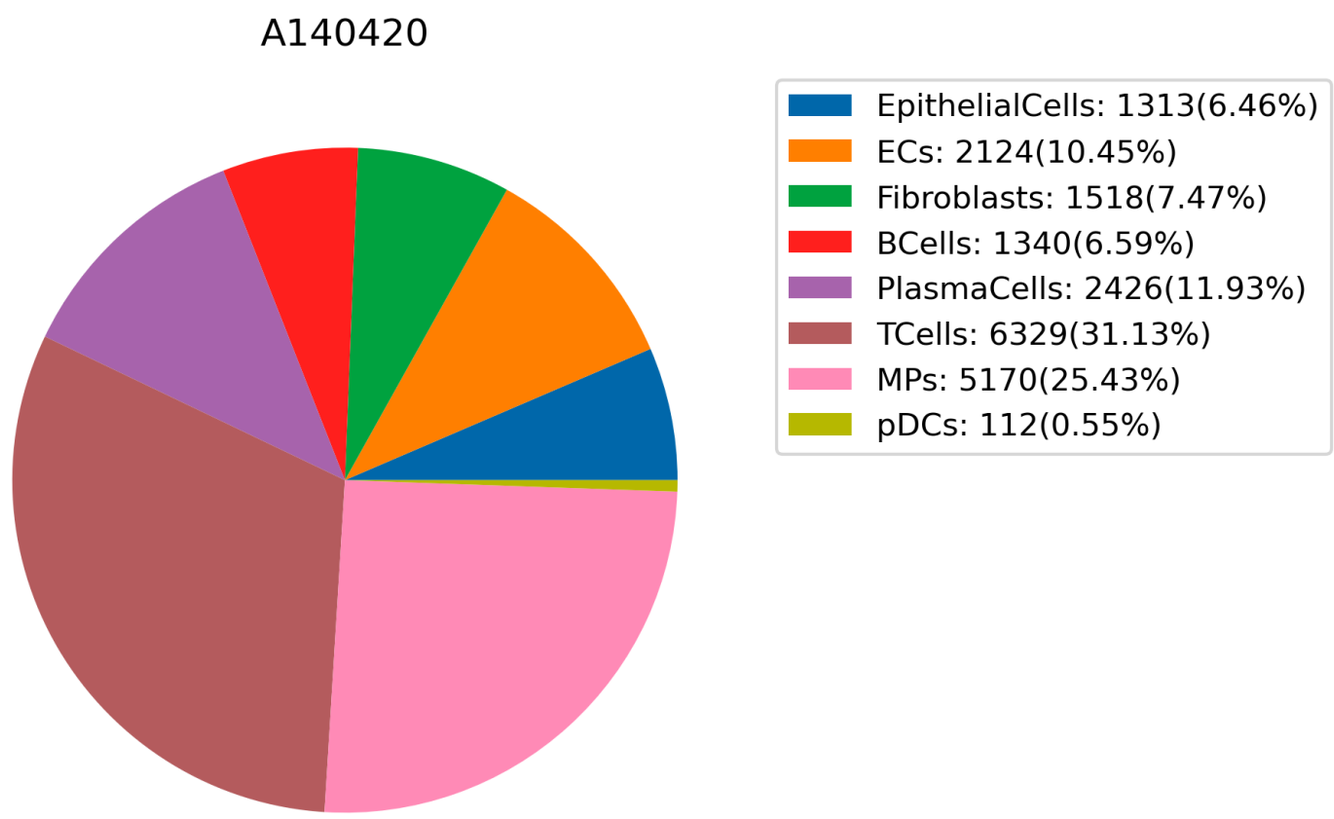


Figure3.3.3：基于自动化注释的结果，展示自动化注释后的细胞类型的细胞数目和占比饼图。

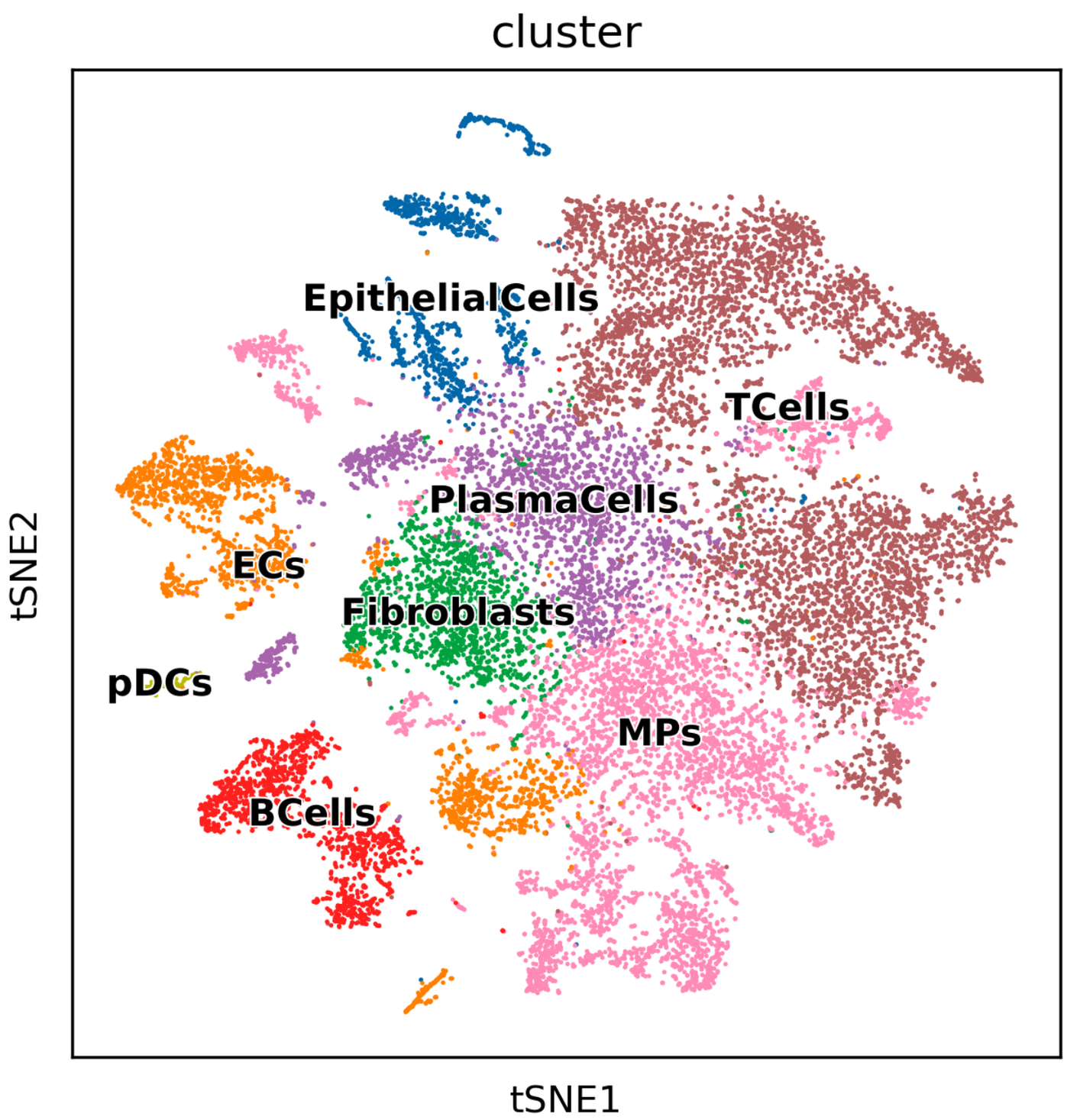


Figure3.3.4：基于自动化注释的结果，进行tSNE降维。

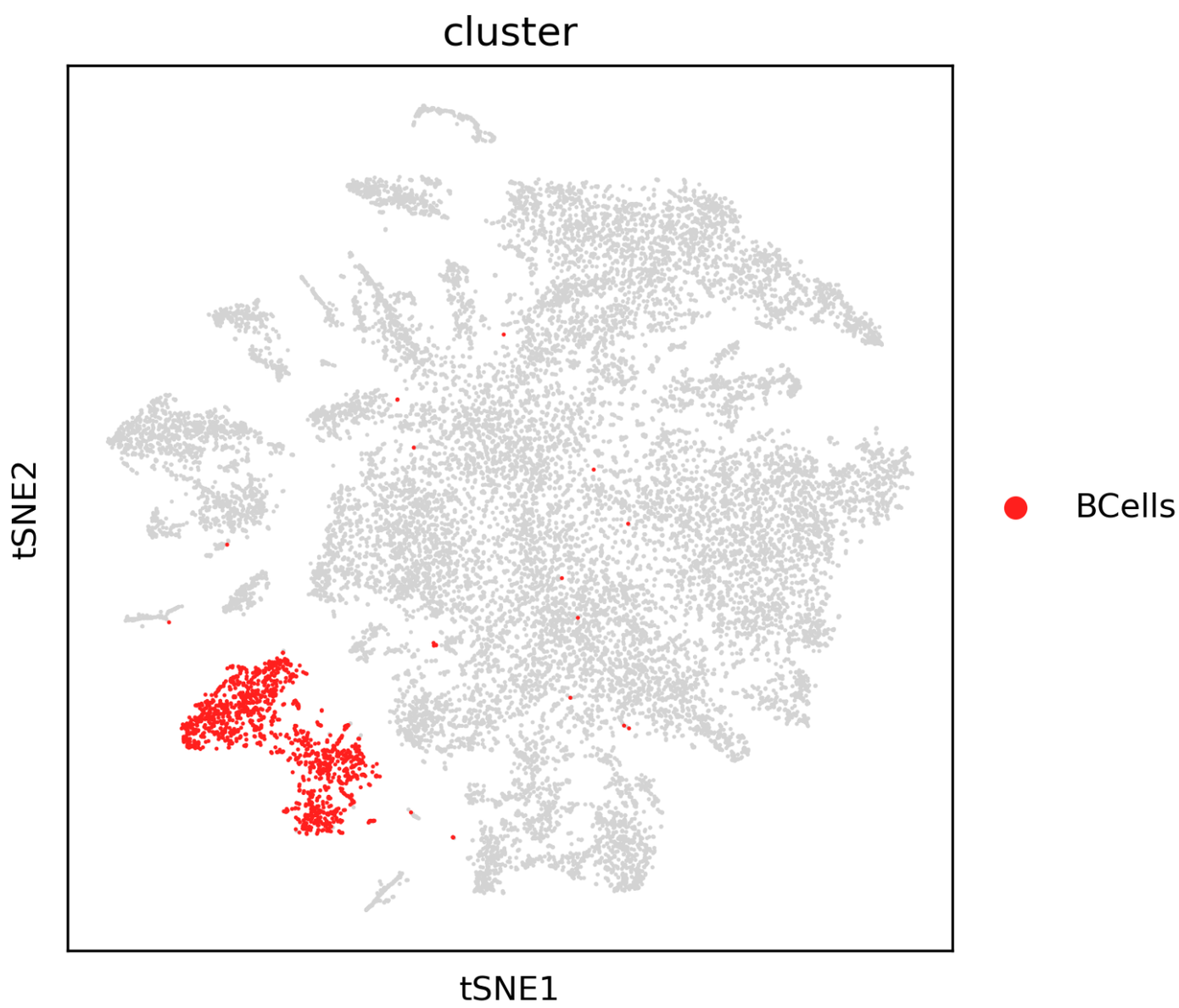


Figure3.3.5：基于自动化注释的结果，展示每个细胞类型在tSNE降维中的位置分布情况。

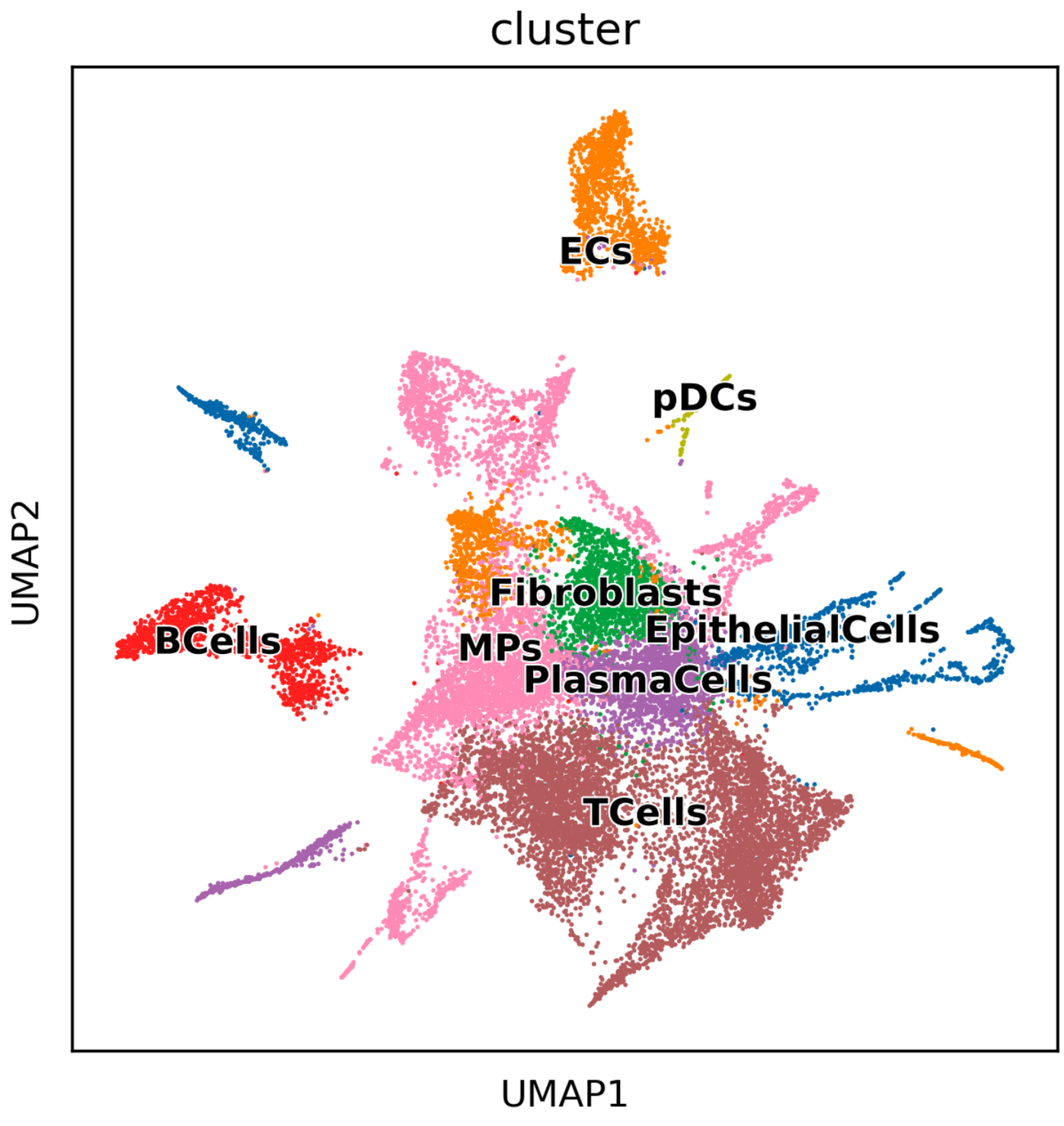


Figure3.3.6：基于自动化注释的结果，进行UMAP降维。

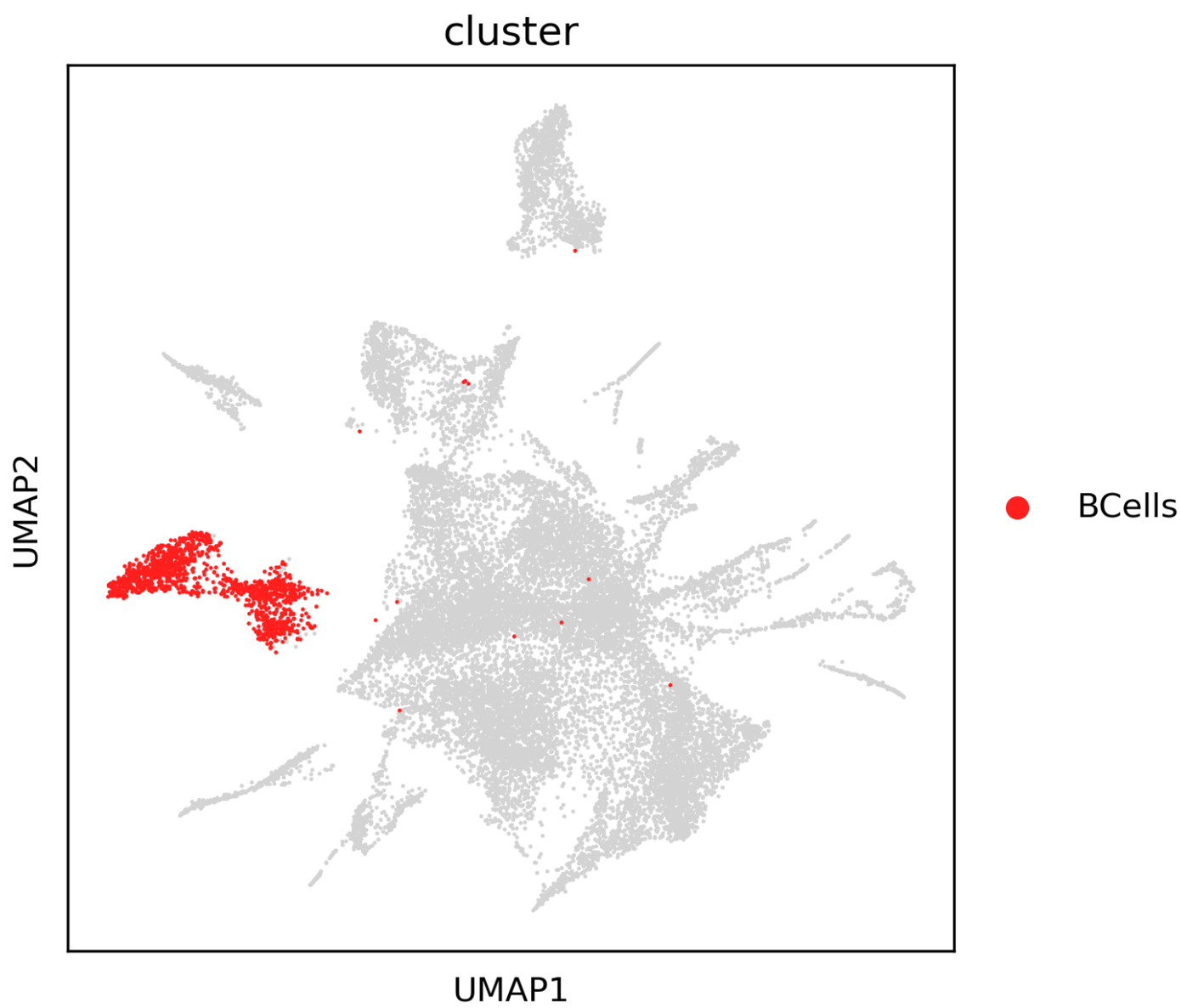


Figure3.3.7：基于自动化注释的结果，展示每个细胞类型在UMAP降维中的位置分布情况。

Figure3.3.8：基于自动化注释的结果，展示样本中细胞类型top10差异基因热图。

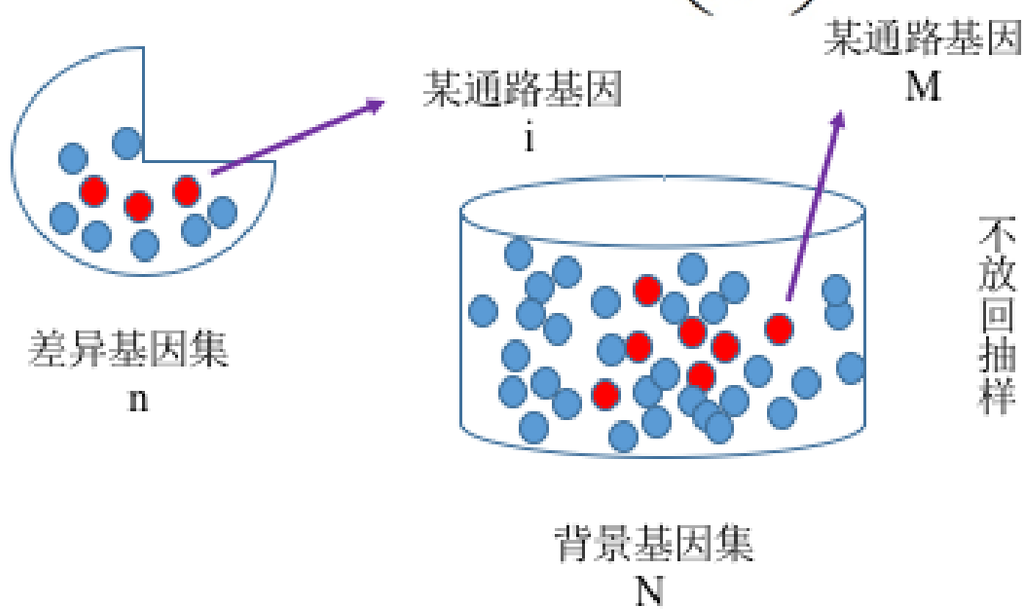
该部分具体结果可见result文件夹中02.rename

7 富集分析

通过对各个细胞类型差异上调基因进行富集分析，可以找到该细胞类型特异性高表达的基因与哪些生物学功能或通路显著性相关。使用clusterProfiler软件对基因集进行GO功能富集分析，Reactome富集分析，KEGG通路富集分析。富集分析基于超几何分布原理，如下图所示，其中差异基因集为cluster差异上调基因注释到ENTREZ数据库的基因集，背景基因集为所有cluster的差异基因并集的基因列表注释到ENTREZ数据库的基因集。结题报告展示的表格是第一个cluster上调基因富集分析结果。富集展示的图片是所有组合的所有

差异基因的富集分析结果。

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$



7.1 GO功能富集

GO(Gene Ontology)是描述基因功能的综合性数据库，可分为分子功能（Molecular Function），生物过程（biological process）和细胞组成（cellular component）三个部分。GO富集以padj小于0.05为显著富集，富集结果如下表所示：

ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID
CC	GO:0022626 (GO:0022626)	cytosolic ribosome	62/539	114/28886	0	0	0	Rps27/Rps20/Rps19
MF	GO:0003735 (GO:0003735)	structural constituent of ribosome	63/520	166/28438	0	0	0	Rps27/Rps20/Rps19

（1） Category：GO数据库
（2） ID：GO编号
（3） Description：GO编号对应的功能描述
（4） GeneRatio：注释到GO编号上的差异基因数与差异基因总数的比值
（5） BgRatio：注释到GO编号上的背景基因数与背景基因总数的比值
（6） pvalue：显著性检验p值
（7） padj：校正后的p值
（8） geneID：注释到GO编号上的基因
（9） Count：注释到GO编号上的基因数

Table2.1：仅展示单个样本其中一个细胞类型的部分富集分析结果

从GO富集分析结果中，选取最显著的30个Term绘制棒棒糖图进行展示，若不足30个，则绘制所有Term信

息，结果如下图所示：

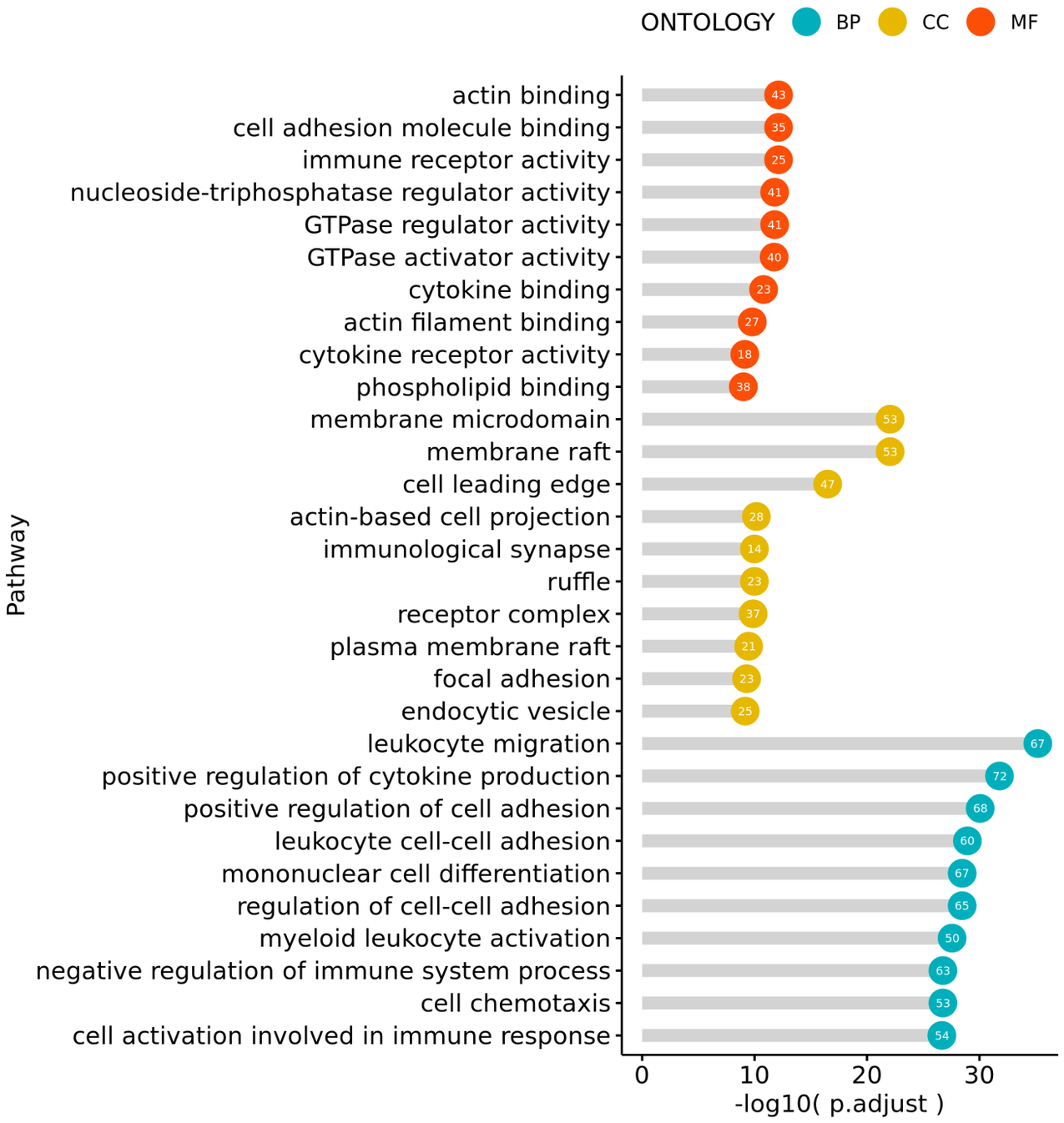


Figure3.4.1：GO富集棒棒糖图，不同颜色代表不同的功能部分（蓝：生物过程（biological process） 黄：细胞组成（cellular component） 红： 分子功能（Molecular Function）；横向为各个功能通路，纵向为该通路的富集显著程度用-log10(padj)）。

从GO富集分析结果中，选取最显著的30个Term绘制散点图进行展示，若不足30个，则绘制所有Term信息，结果如下图所示：

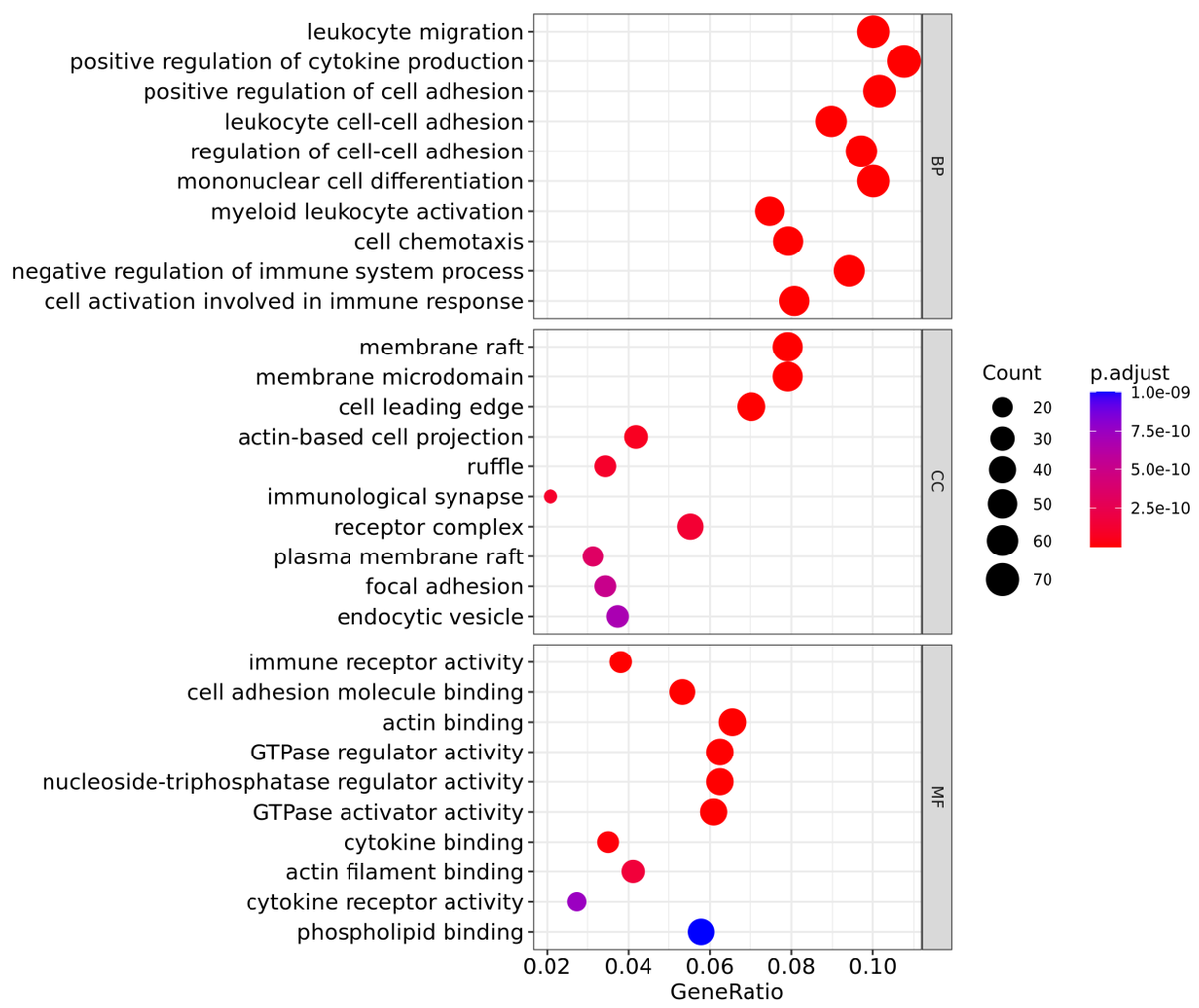


Figure3.4.2：GO富集dot图，纵向代表富集通路，横向代表注释到GO编号上的差异基因数与差异基因总数的比值；图中圆大小表示基因的多少，颜色由红到蓝表示富集显著性由高到低。

从GO富集分析结果中，选取最显著的10个Term绘制富集基因和通路之间的关联网络图进行展示，若不足10个，则绘制所有Term信息，结果如下图所示：

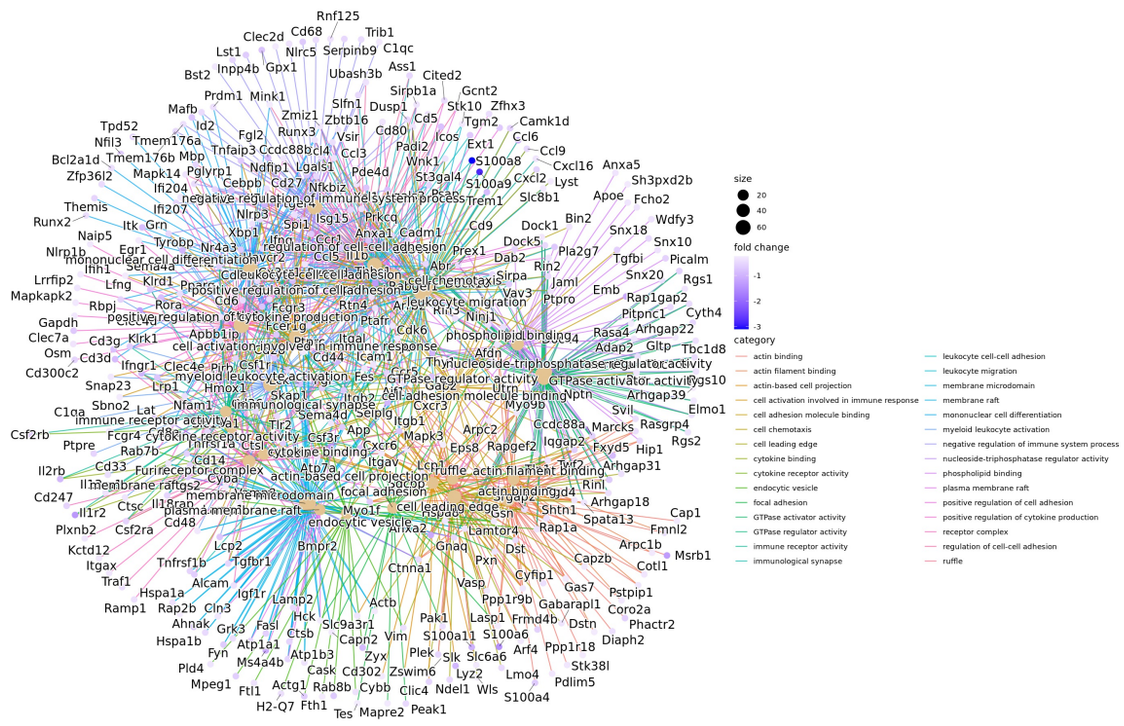


Figure3.4.3：GO富集基因和通路之间的关联网络图，黄色点为通路，点大小表示基因数量，灰色点为基因，连线表示基因属于指定通路。

7.2 KEGG通路富集分析

KEGG(Kyoto Encyclopedia of Genes and Genomes)是整合了基因组、化学和系统功能信息的综合性数据库。KEGG富集以padj小于0.05为显著富集，富集结果如下表所示：

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID
mmu03010	Ribosome	62/288	179/8930	0.0000000	0.0000000	0.0000000	Rps27/Rps20/Rps19/Rpl2
mmu05171	Coronavirus disease - COVID-19	66/288	247/8930	0.0000000	0.0000000	0.0000000	Syk/Rps27/Rps20/Prkcb/F
mmu04662	B cell receptor	16/288	81/8930	0.0000000	0.0000004	0.0000004	Cd79a/Cd79b/Cd19/Cd22

（1） Category：KEGG数据库
（2） ID：KEGG通路编号
（3） Description：KEGG通路编号对应的功能描述
（4） GeneRatio：注释到KEGG通路编号上的差异基因数与差异基因总数的比值
（5） BgRatio：注释到KEGG通路编号上的背景基因数与背景基因总数的比值
（6） pvalue：显著性检验p值
（7） padj：校正后的p值
（8） geneID：注释到KEGG通路编号上的基因
（9） Count：注释到KEGG通路编号上的基因数

Table2.2：仅展示单个样本其中一个细胞类型的部分富集分析结果

从KEGG富集结果中，选取最显著的20个通路绘制棒棒糖图进行展示，若不足20个，则绘制所有通路信息，结果如下图所示：

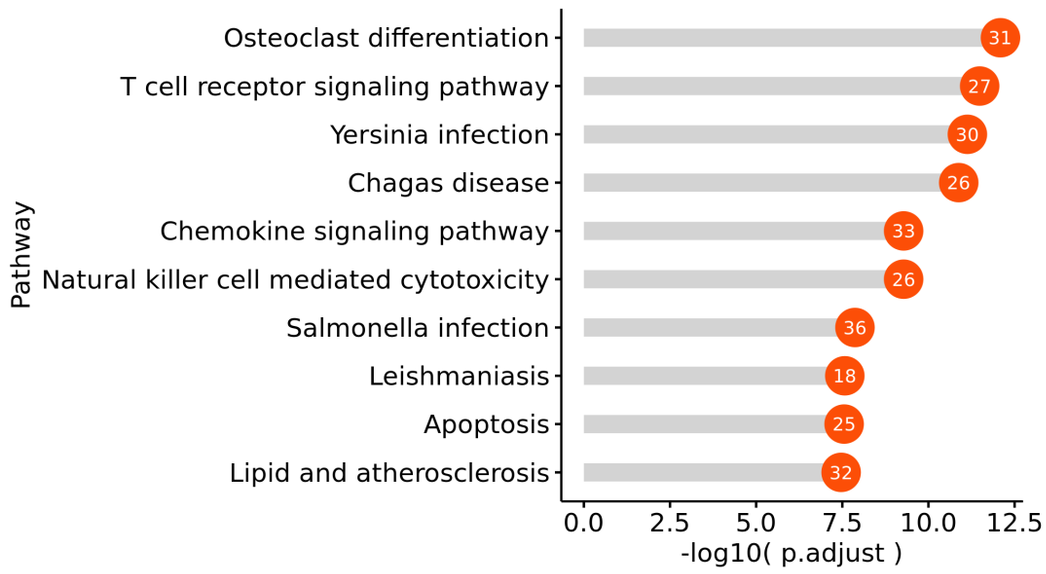


Figure3.4.4：KEGG富集棒棒糖图，横向代表各个通路，纵向代表该通路的显著程度

从KEGG富集结果中，选取最显著的20个通路绘制散点图进行展示，若不足20个，则绘制所有通路信息，结果如下图所示：

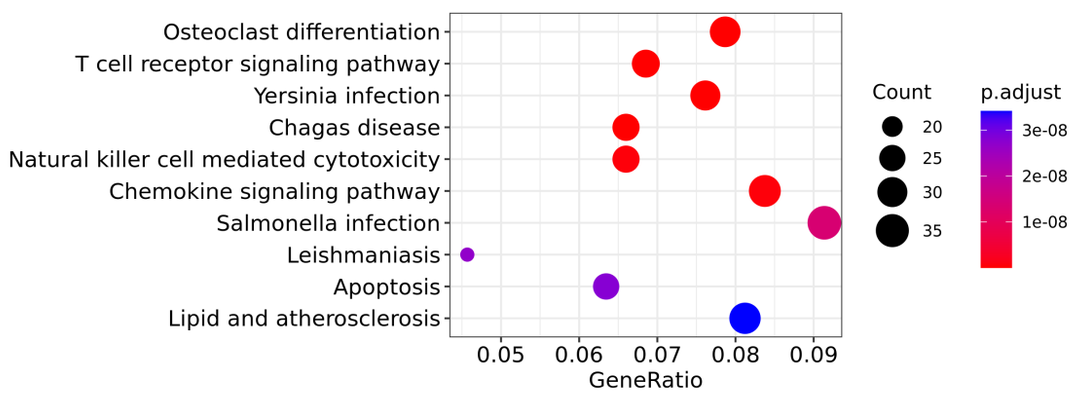


Figure3.4.5：KEGG富集dot图，纵向代表富集通路，横向代表注释到KEGG通路编号上的差异基因数与差异基因总数的比值；图中圆大小表示基因的多少，颜色由红到蓝表示富集显著性由高到低。

从KEGG富集分析结果中，选取最显著的10个Term绘制富集基因和通路之间的关联网络图进行展示，若不足10个，则绘制所有Term信息，结果如下图所示：

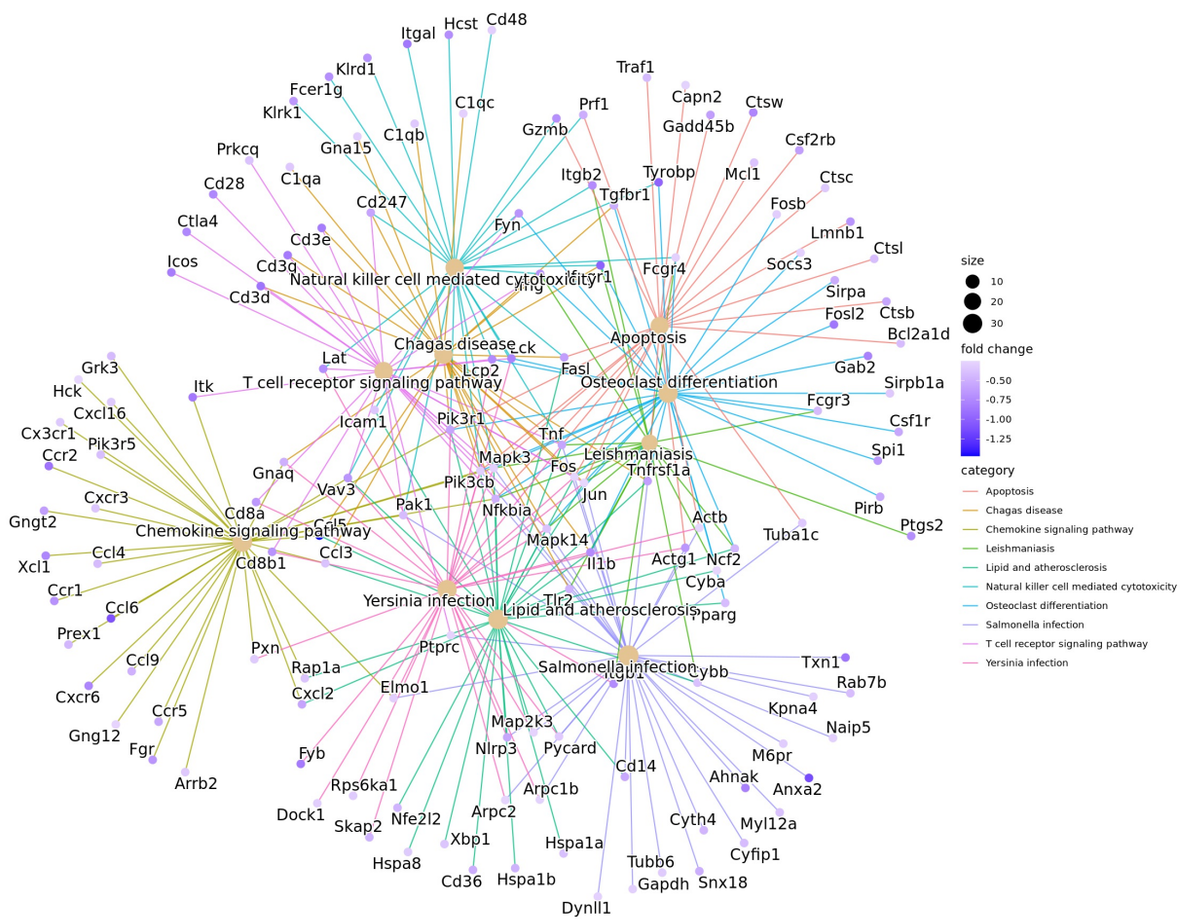


Figure3.4.6：KEGG富集基因和通路之间的关联网络图，黄色点为通路，点大小表示基因数量，灰色点为基因，连线表示基因属于指定通路。

该部分具体结果可见result文件夹中03.enrich

8 细胞互作分析

以细胞群为相互作用的对象，计算受配体在每个细胞群的平均表达量，然后计算两个细胞群之间受配体的平均表达量作为相互作用的指标。不同于之前的其他相互作用分析方法，CellphoneDB 考虑蛋白复合体的情况，例如受体如果有两个亚基 Sub.1 和 Sub.2，取表达量较低的亚基作为受体的表达量。

CellphoneDB 用置换检验计算某一对相互作用的 P 值。随机打乱细胞类型标签，计算相互作用平均表达量；重复 1000 次，得到平均表达量的分布；如果用鉴定细胞类型计算的平均表达量较高，大于分布的 top50，那么我们可以认为这一对相互作用的 P 值小于 50/1000=0.05。

CellphoneDB互作分析软件推断的细胞类型之间的互作关系如下图所示：

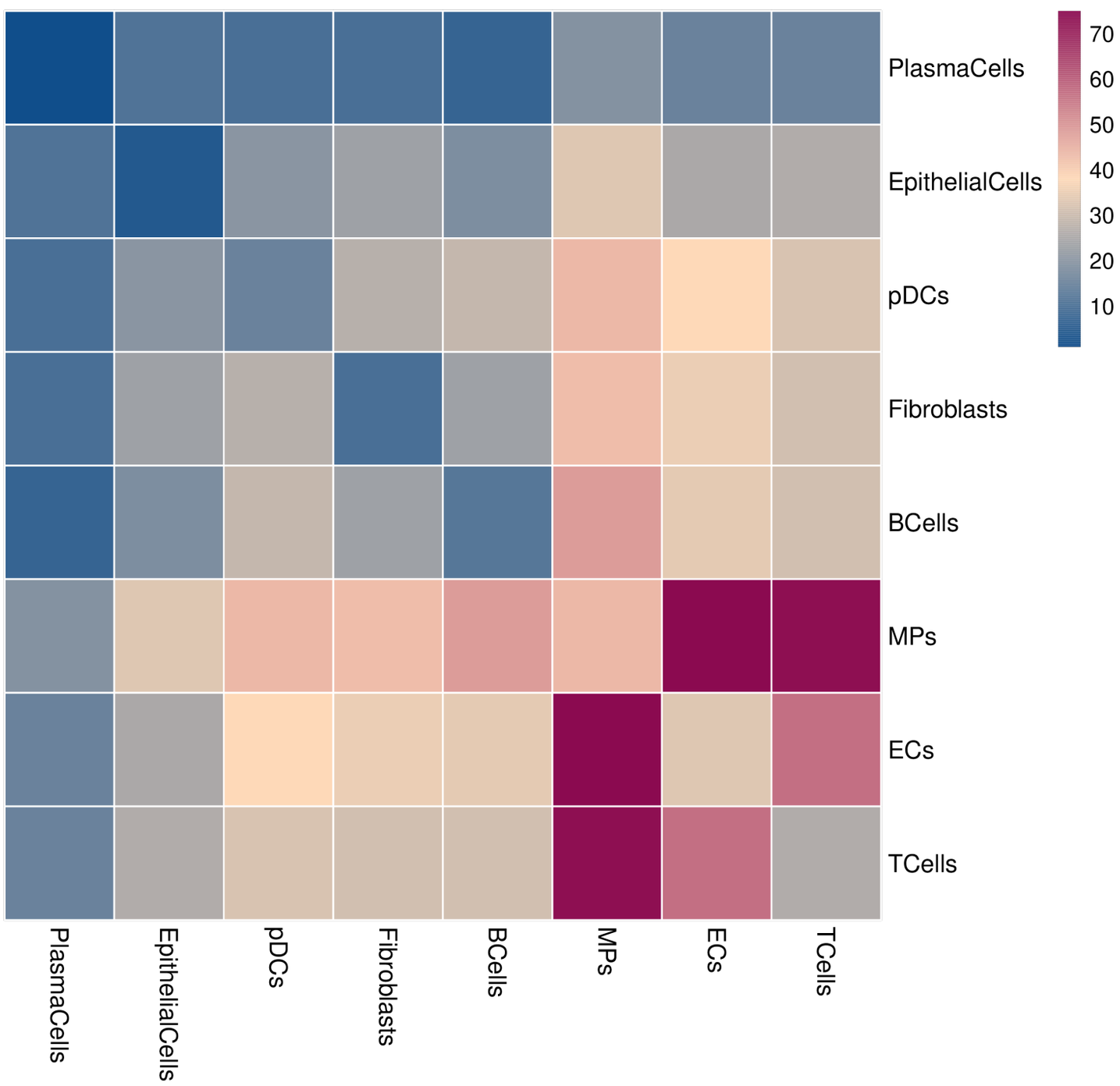


Figure3.5.1：展示了两个细胞类型之间的相互作用对数，颜色越深，代表两个细胞类型之间互作对数越多。

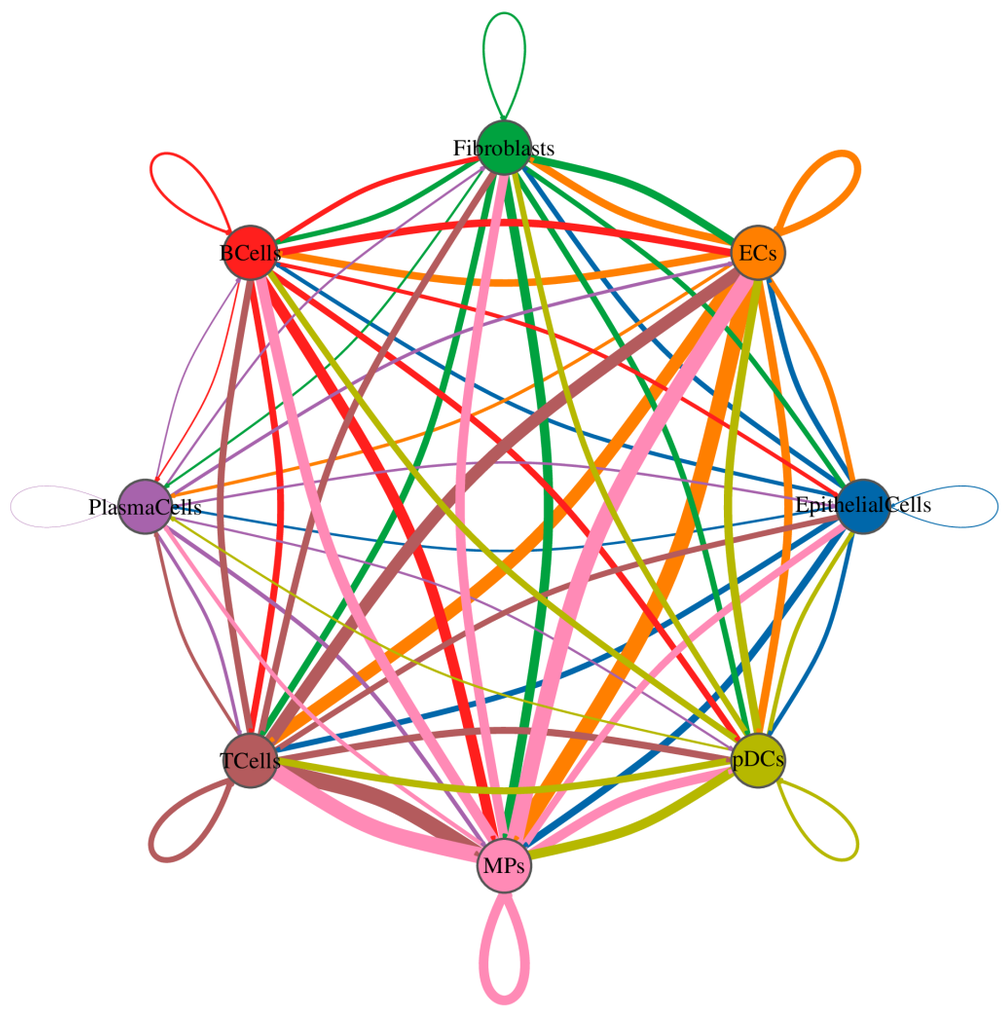


Figure3.5.2：所有细胞类型间的互作网络图，其中网络节点为细胞类型，网络边粗细为配体与受体对总数，线条颜色与配体细胞类型一致。

该部分具体结果可见result文件夹中04.cellphonedb

9 备注

9.1 结果文件解压方法

压缩格式	用户类型	方法
***.tar.gz形式的压缩文件**	Unix/Linux/Mac用户	使用tar -zxvf *.tar.gz命令
	Windows用户	使用解压缩软件如WinRAR、7-Zip等
***.gz形式的压缩文件**	Unix/Linux/Mac用户	使用gzip -d *.gz命令
	Windows用户	使用解压缩软件如WinRAR、7-Zip等

压缩格式	用户类型	方法
***.zip形式的压缩文件**	Unix/Linux/Mac用户	使用unzip *.zip命令
	Windows用户	使用解压缩软件如WinRAR、7-Zip等

9.2 结果文件格式说明

文件类型	文件描述	打开方式
***.fasta**	序列文件，fasta格式，一般为基因序列或者基因组序列。因文件一般较大，打开较为困难	unix/Linux/Mac用户使用less或more命令
		windows用户使用高级文本编辑器Editplus/Notepad++等
***.fq/fastq**	序列文件，fastq格式，一般为reads序列；因文件一般较大，打开较为困难	unix/Linux/Mac用户使用 less 或 more 命令
		windows用户使用高级文本编辑器Editplus/Notepad++等
***.xls*,.txt**	结果数据表格文件；文件以制表符Tab分隔	unix/Linux/Mac用户使用 less 或 more 命令
		windows用户使用高级文本编辑器Editplus/Notepad++ 等，也可以用Microsoft Excel打开
***.png**	结果图像文件；位图,无损压缩	unix/Linux/Mac用户使用display命令打开
		windows用户可以使用图片浏览器打开，如photoshop等。
***.pdf**	结果图像文件；矢量图，可以放大和缩小而不失真，方便用户查看和编辑处理，可使用Adobe Illustrator进行图片编辑，用于文章发表等	windows/Mac用户可以使用Adobe Reader/福昕阅读器/网页浏览器等打开
		unix/Linux用户使用evince命令打开

9.3 主要分析软件列表及版本

分析	软件	版本
单细胞分析软件	Seurat	4.0.6
	Scanpy	1.8.2
	CellphoneDB	2.1.7
	clusterProfiler	3.8.1
	CellID	0.1.0
	samtools	1.14

分析	软件	版本
其他分析软件	velocyto.py	0.17.17
	seuratdisk	0.0.9019
	ggplot2	3.1.0
	R	4.1.2
	python	3.6.7
	perl	5.32.1

10 参考文献

- 1.Wolf, F., Angerer, P. & Theis, F. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 19, 15 (2018). (scanpy)
- 2.Andrew Butler, Efthymia Papalexi ,Rahul Satija et al.,Integrating single-cell transcriptomic data across different conditions, technologies, and species Nature Biotechnology volume 36, pages 411–420 (2018) (Seurat)
- 3.Heiser CN, Wang VM, Chen B, Hughey JJ, Lau KS. Automated quality control and cell identification of droplet-based single-cell data using dropkick. Genome Res. 2021 Oct;31(10):1742-1752. (dropkick)
- 4.Etienne Becht, Charles-Antoine Dutertre, Immanuel W.H. Kwok, Lai Guan Ng, Florent Ginhoux, Evan W Newell.Evaluation of UMAP as an alternative to t-SNE for single-cell data. BioRxiv .(UMAP for single-cell data)
- 5.Leland McInnes, John Healy. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:1802.03426. (UMAP)
- 6.Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL (2014). The dynamics and regulators of cell fate decisions are revealed by pseudo-temporal ordering of single cells. Nature Biotechnology. (monocle)
- 7.Qiu X, Hill A, Packer J, Lin D, Ma Y, Trapnell C (2017). Single-cell mRNA quantification and differential analysis with Census. Nature Methods. (monocle)
- 8.Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner H, Trapnell C (2017). Reverse graph embedding resolves complex single-cell developmental trajectories. BioRxiv. (monocle)
- 9.Wang et al., Single-Cell RNA Sequencing Analysis Reveals Sequential Cell Fate Transition during Human Spermatogenesis, Cell Stem Cell (2018), https://doi.org/10.1016/j.stem.2018.08.007(monocle2 (https://doi.org/10.1016/j.stem.2018.08.007(monocle2) use gene set)
- 10.Bergen,V.,Lange,M.,Peidli,S.etal.Generalizing RNAvelocity to transient cell states through dynamical modeling.NatBiotechnol38,1408–1414 (scvelo)
- 11.Wilk,A.J.,Rustagi,A.,Zhao,N.Q.etal.A single-cell atlas of theperipheral immune response in patients with severe COVID-19.NatMed26,1070–1076 (scvelo)
- 12.VandeSandeB,FlerinC,DavieK,DeWaegeneerM,HulselmansG,AibarS,SeurinckR,SaelensW,CannoodtR,RouchonQ,VerbeirenT,DeMaeyerD, SCENIC workflow for single-cell generegulatory network analysis.NatProtoc.2020Jul;15(7):2247-2276. (pyscenic)
- 13.Zhang Yang,Liu Tianyuan,Hu Xuesong et al. CellCall: integrating paired ligand-receptor and transcription factor activities for cell-cell communication.[J] .Nucleic Acids Res, 2021 (cellcall)
- 14.Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes[J]. Nucleic acids research, 2000, 28(1): 27-30.(KEGG)
- 15.https://www.genome.jp/kegg/pathway.html(kegg pathway)
- 16.David Croft, Gavin O'Kelly et al.,Reactome: a database of reactions, pathways and biological processes.Nucleic Acids Res. 2011

Jan; 39(Database issue): D691–D697.(Reactome database)

- 17.Yu G, He Q (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Molecular BioSystems, 12(12), 477-479. doi: 10.1039/C5MB00663E, http://pubs.rsc.org/en/Content/ArticleLanding/2015/MB/C5MB00663E (http://pubs.rsc.org/en/Content/ArticleLanding/2015/MB/C5MB00663E).(Reactome)
- 18.Yu G, Wang L, Han Y and He Q*. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS: A Journal of Integrative Biology. 2012, 16(5):284-287.(clusterProfiler)
- 19.Cortal, A., Martignetti, L., Six, E. et al. Gene signature extraction and cell identity recognition at the single-cell level with Cell-ID. Nat Biotechnol 39, 1095–1102 (2021). (cellid)
- 20.Young, Matthew D., et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. Science 361.6402 (2018): 594-599.
- 21.Chen Z, Zhou L, Liu L, et al., Single-cell RNA sequencing highlights the role of inflammatory cancer-associated fibroblasts in bladder urothelial carcinoma. Nature Communications. 2020; 11(1): 5077 (CellphoneDB)