

**基于单细胞和转录组测序数据鉴定胃癌患者中TAM相关基因及细胞生态型**

**项目编号：BJTC-277**

**目录**

[1 背景 2](#_Toc103936942)

[2 分析方法 3](#_Toc103936943)

[3 分析结果 5](#_Toc103936944)

[3.1 单细胞质控分析 5](#_Toc103936945)

[3.2 单细胞数据标准化和PCA分析 6](#_Toc103936946)

[3.3 聚类与细胞类型鉴定分析 7](#_Toc103936947)

[3.4 胃癌内细胞的异质性分析 10](#_Toc103936948)

[3.5 胃癌内细胞EMT、血管新生、细胞干性之间的相关性 11](#_Toc103936949)

[3.6 胃癌内Monocyte/Macrophage谱系分析 12](#_Toc103936950)

[3.7 胃癌内Monocyte/Macrophage细胞EcoTyper分析 16](#_Toc103936951)

[3.8 不同生态型间M1和M2的差异分析 18](#_Toc103936952)

[3.9 WGCNA分析筛选TAM相关模块基因 21](#_Toc103936953)

[3.10 通过TAM相关标记基因对样本进行分子分型 28](#_Toc103936954)

[4 总结 33](#_Toc103936955)

[5 软件列表 34](#_Toc103936956)

[6 结果文件列表 35](#_Toc103936957)

[References 72](#_Toc103936958)

# 1 背景

胃癌 (GC) 是全球第五大常见癌症，也是癌症相关死亡的第三大原因，仍然是主要的公共卫生问题[ 1-3 ]。GC 的侵袭性生物学特征包括高发病率、早期复发和远处转移，是导致预后不良和 5 年随访时间较短的原因 [ 4 ]。因此，在胃癌预后的可预测性和新靶向治疗方法的开发方面，需要更有效和可靠的生物标志物。

肿瘤相关巨噬细胞 (TAM) 是浸润肿瘤组织的巨噬细胞，是肿瘤微环境 (TME) 中最丰富的免疫细胞之一。TAM 分为典型的激活 M1 亚型和交替激活的 M2 亚型，它们具有不同的受体、细胞因子和趋化因子表达[5-7]。M1 的特点是表达大量诱导型一氧化氮合酶和 TNF-α，并通过促进促炎和免疫反应发挥抗肿瘤活性[8-10]。 M2 通常表达精氨酸酶 1 和高水平的细胞因子、生长因子和蛋白酶，以支持其致癌功能[11,12]。此外，M2参与刺激肿瘤血管生成、基质重塑、肿瘤细胞迁移和侵袭，促进免疫抑制[13,14]。TAMs 的抗炎特性受肿瘤细胞控制，这对 GC 患者的治疗策略具有重要意义，尤其是针对癌细胞和巨噬细胞的联合治疗，可以产生协同作用[15]。此外，在免疫治疗的研究中，PD-L1/PD-1 抗体在 GC 中的功效需要 M1 样 TAM，因为它们通过释放 CXCL9、10 和 11来招募更多浸润的CD8 + T 细胞[16]。M1样TAMs不仅提高了GC的客观反应率，而且增加了PD-L1/PD-1抗体的应用。这表明 TAM 在 GC 发展中发挥关键作用，并为抗癌治疗提供了良好的靶点[17]。之前有研究表明，高水平的 TAM 浸润与 GC 的侵袭性特征有关，并且是 GC 患者的独立不良预后因素[18,19]。因此，我们推测TAMs的数量和分布是影响癌细胞和TAMs共同进化的关键因素。

因此，本研究通过对单个肿瘤细胞、大量RNA序列数据和临床数据的进行生物信息学分析， 研究BC中肿瘤的异质性以及巨噬细胞与临床结果的关系。

# 2 分析方法

1、 数据来源：胃癌和转移肿瘤组织样品的8个10 x单细胞转录组测序数据以及TCGA-GC队列。

2、 利用R包Seurat对单细胞测序数据质控，过滤除掉表达基因数异常或线粒体基因表达占比过高的细胞；标准化之后筛选出高变异的基因，归一化之后利用PCA和tSNE对8个样品的单细胞测序数据进行降维聚类。

3、 利用R包SingleR、CellMarker数据库对细胞进行注释，以识别不同的细胞类型。

4、 利用GSVA包分析肿瘤样本中各类型细胞的异质性，并对EMT、血管新生和细胞干性的相关性进行分析。

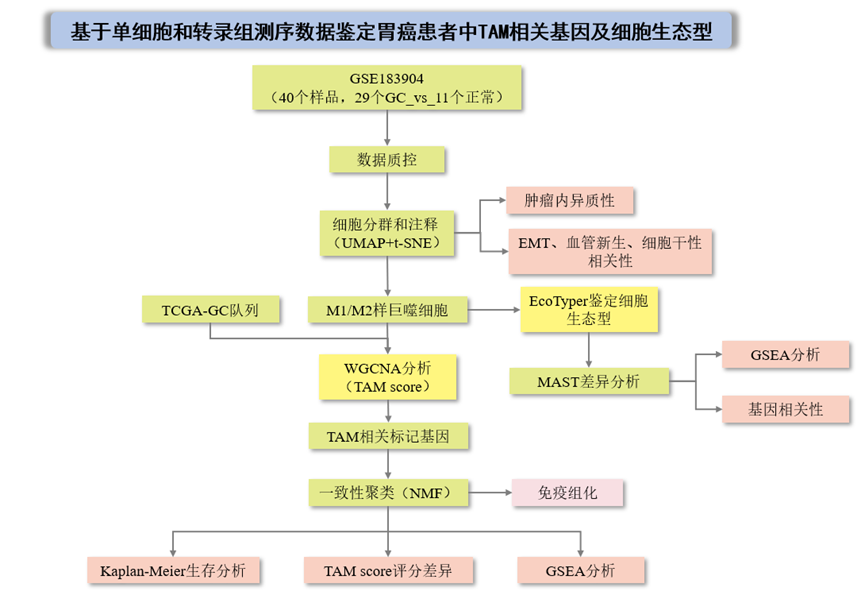
5、 运用机器学习EcoTyper对细胞的生态系统进行分析。

6、 利用R包MAST在肿瘤不同生态型间鉴定差异基因，之后计算spearman相关性，进行GSEA分析。

7、 利用TCGA-GC数据集进行WGCNA共表达网络分析，筛选TAM相关标记基因。

8、 根据TAM相关差异基因对样本进行一致性聚类，并对不同簇样本进行分析。

**研究思路**



# 3 分析结果

## 3.1 单细胞质控分析

从GEO 下载GSE163558 单细胞10x数据([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/))。对6 名患者的8份新鲜人体组织样本进行了 scRNA-seq 分析，包括 3 个原发性肿瘤样本（PT；即 PT1、PT2 和 PT3），5个转移性肿瘤样本样品 (metastasis)。在 metastasis 中，我们获得了两个肝转移样本（Li；即 Li1 和 Li2）、两个淋巴结转移样本（LN；即 LN1 和 LN2）、一个腹膜转移样本（P；即 P1）。

使用 Seurat R包（4.1.0）将上述样本原始基因表达矩阵导入 R（4.1.1），保留基因至少在3个细胞中表达以及每个细胞至少表达200个基因的基因和细胞，符合标准的有45388个细胞和25366个基因。图1所示，我们可视化nFeature\_RNA（每个细胞检测到的基因数目）、nCount\_RNA（基因数目一共测到的count数目,也就是总的UMI数目）、percent.mt（线粒体在样本中的占比）。

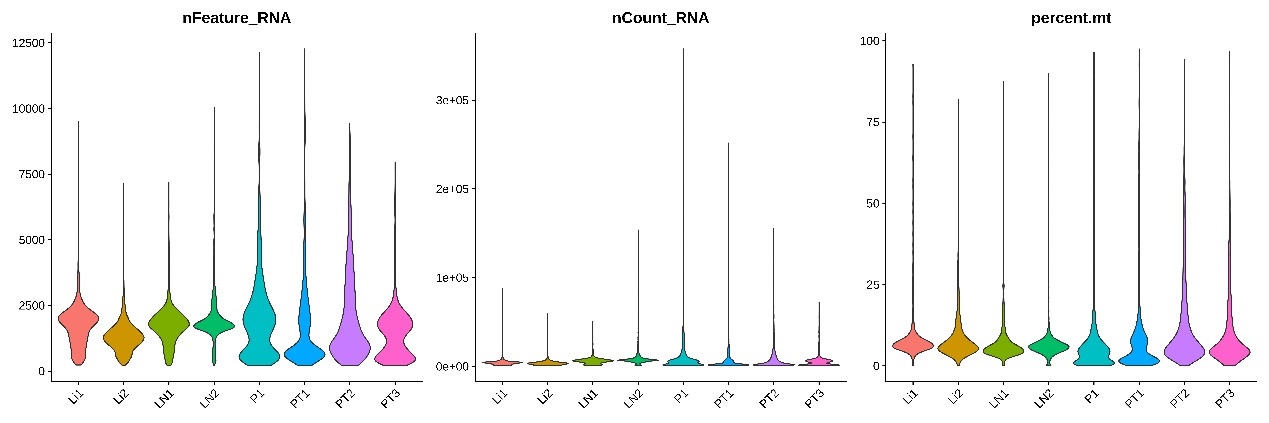


图1. 质控前单细胞nFeature\_RNA、nCount\_RNA和percent.mt指标图

我们设置质控的标准为nFeature\_RNA数目（大于2%（316）和小于98%（5952））、nCount\_RNA数目（小于95%（20752））、percent.mt小于20%，最后得到38148个细胞和20752个基因。如图2所示，可视化质控后nFeature\_RNA（每个细胞检测到的基因数目）、nCount\_RNA（基因数目一共测到的count数目,也就是总的UMI数目）、percent.mt（线粒体在样本中的占比）。

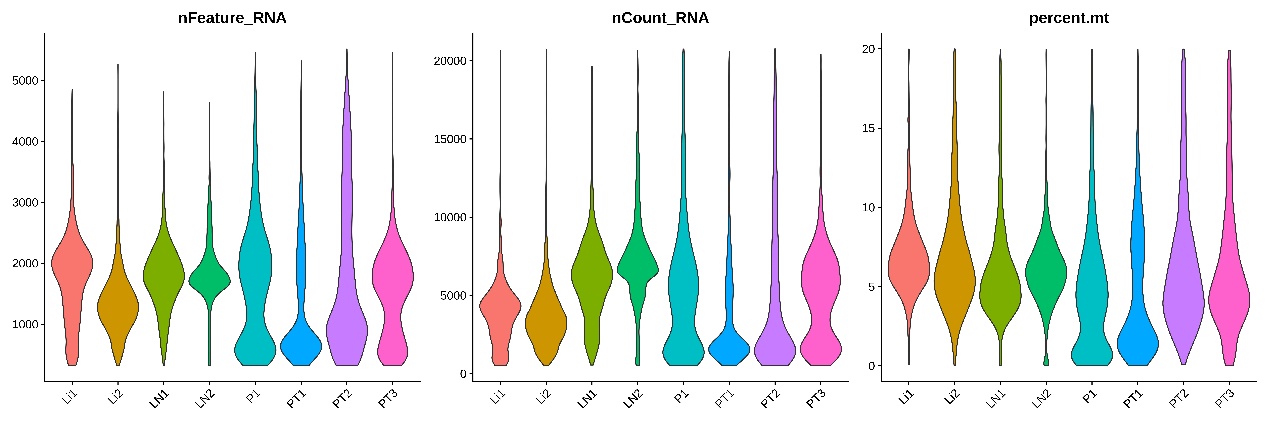


图2. 质控后单细胞nFeature\_RNA、nCount\_RNA和percent.mt指标图

## 3.2 单细胞数据标准化和PCA分析

使用Seurat包中的SCTransform 函数对数据集RNA的表达标准化和缩放，然后对SCTransformed RNA (SCT)进行PCA降维，根据PCA碎石图选取PC数目纳入后续分析。

如图3所示，我们可视化PCA碎石图前40的PC，选取前36个PC纳入后续分析。

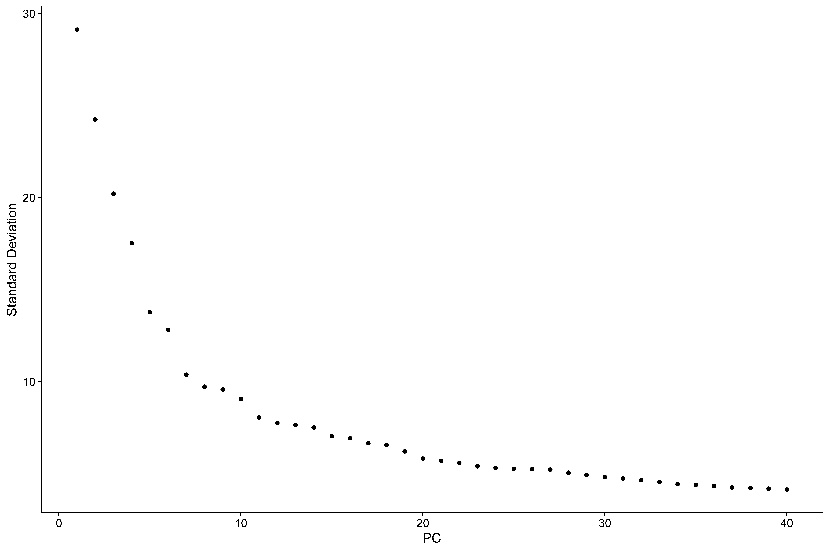


图3.PCA分析碎石图

## 3.3 聚类与细胞类型鉴定分析

通过主成分分析选择了最显着的 36 个主成分，并使用它们进行 tSNE 降维。使用 Seurat 中的“FindClusters”功能进行细胞聚类，并通过表1标记基因的表达对细胞簇进行注释。

表1.胃癌单细胞数据中不同细胞类型的标记基因

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 细胞类型 | | 标记基因 | | 来源 | |
| Epithelial | | EPCAM | | 胃癌单细胞文献报道 | |
| Endothelial | | PECAM1 | | 胃癌单细胞文献报道 | |
| Monocyte/Macrophage | | CD163、CD68 | | 胃癌单细胞文献报道 | |
| Neutrophils | | CSF3R、FCGR3A | | Cellmarker以及文献 | |
| B cell | | MS4A1、CD79A | | 胃癌单细胞文献报道 | |
| T cell | | CD3D、CD2 | | 胃癌单细胞文献报道 | |
| NK cell | | KLRD1 | | 胃癌单细胞文献报道 | |
| Fibroblast | | FN1 | | 胃癌单细胞文献报道 | |
| Proliferative | | MKI67 | | 胃癌单细胞文献报道 | |
| mast cell | | KIT | | 胃癌单细胞文献报道 | |
| pDC | | LILRA4 | | 胃癌单细胞文献报道 | |
| Plasma cell | | MZB1 | | 胃癌单细胞文献报道 | |

如图4所示，我们可视化tSNE聚类分析得到的细胞簇和不同细胞簇来源的患者以及原发胃癌和转移胃癌细胞的分布，通过观察样本的混合状况，并不需要进行批次校正，来自不同样本的细胞聚类效果好不存在明显的批次效应。可在02.TSNE中获取原始的高DPI图。

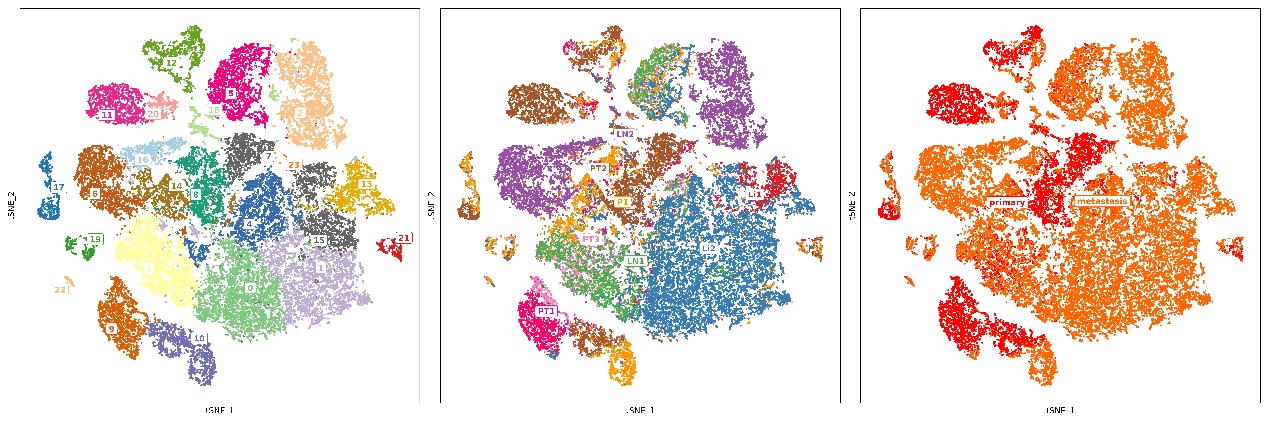


图4.胃癌原发与转移的单细胞细胞簇、患者来源、原发与转移的分类图

通过表1的标记基因，如图5所示，我们可视化原发以及转移胃癌单细胞tSNE图。可在03.celltype中获取原图。其中B cell 有5555、Endothelial有363、Epithelial有1853、Monocyte/Macrophage有1358、Neutrophils有3219、T cell有20535、Fibroblast有695、mast cell有129、NK cell有3329、pDC有49、Plasma cell有446、Proliferative有617。

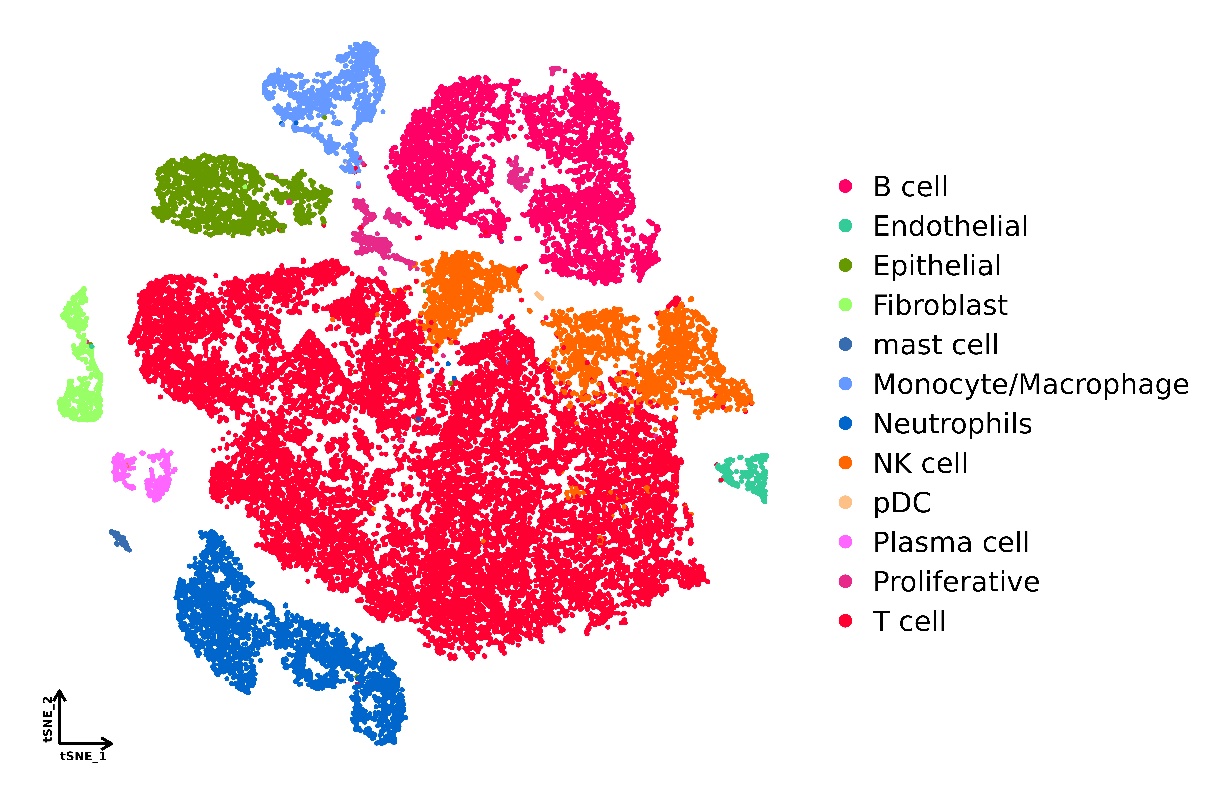


图5.胃癌原发与转移的单细胞tSNE图

同时，我们根据单细胞数据中各种类型细胞的marker基因，使用Seurat包中的FeaturePlot进行可视化marker基因在原发以及转移胃癌单细胞tSNE图的表达。使用Seurat包中的DotPlot进行可视化每种细胞类型的marker基因。

如图6和图7所示，我们可视化marker基因在原发以及转移胃癌单细胞的气泡图和tSNE图的表达，同时在02.celltype的feature中获取单独的marker基因表达图。

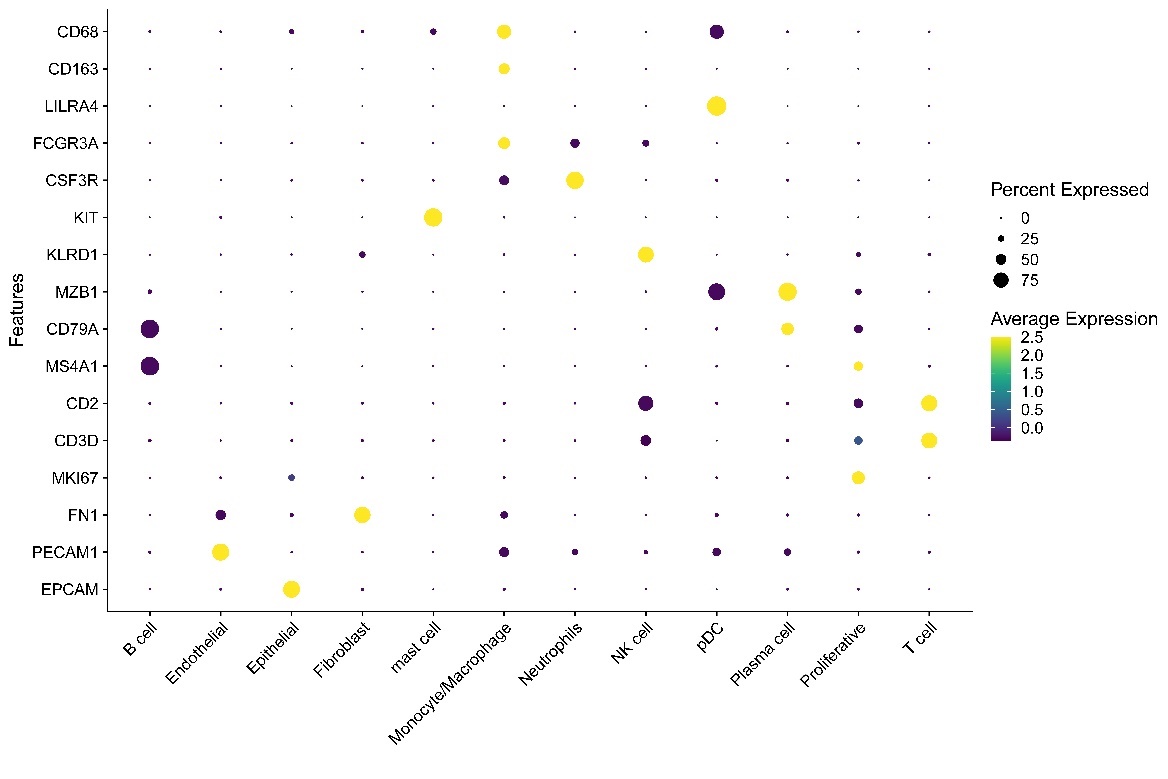


图6.不同细胞类型的marker基因的气泡图

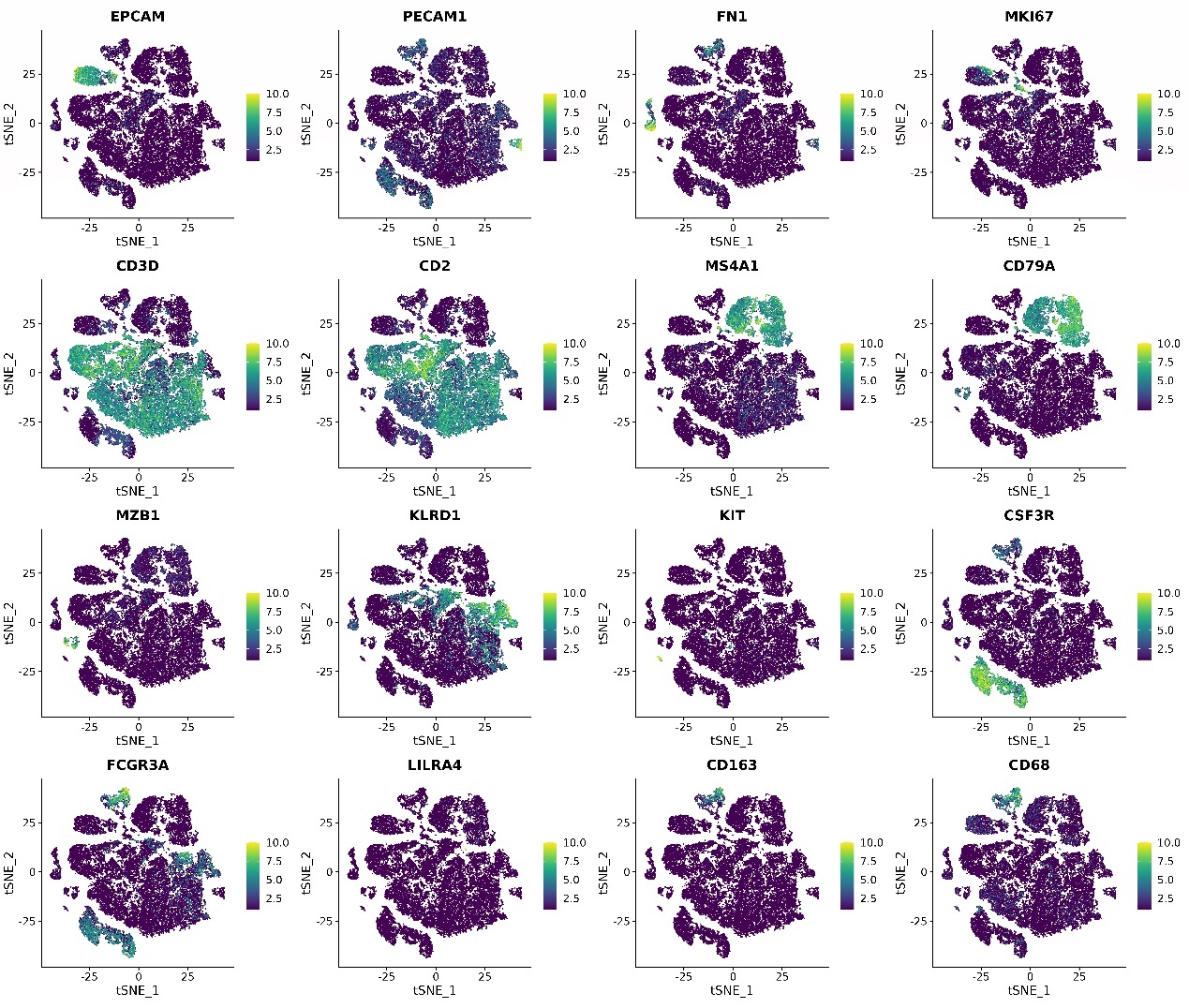


图7.不同细胞类型的marker基因分布图

## 3.4 胃癌内细胞的异质性分析

我们计算原发胃癌与3种转移胃癌的各种细胞占比比较胃癌细胞转移前后单细胞景观的变化。

如图6所示，我们展示原发与3种转移胃癌的各种细胞占比以及原发与转移的各种细胞占比。可以发现在转移的胃癌中T和B细胞明显增加，而Monocyte/Macrophage转移后明显降低，上皮细胞、内皮细胞和成纤维细胞相较转移前降低。通过比较具体的转移部位可以发现，肝转移、淋巴结转移和腹膜转移T细胞含量全部增加，Monocyte/Macrophage在腹膜转移变化不大。可在03.celltype中获取所有高dpi图。

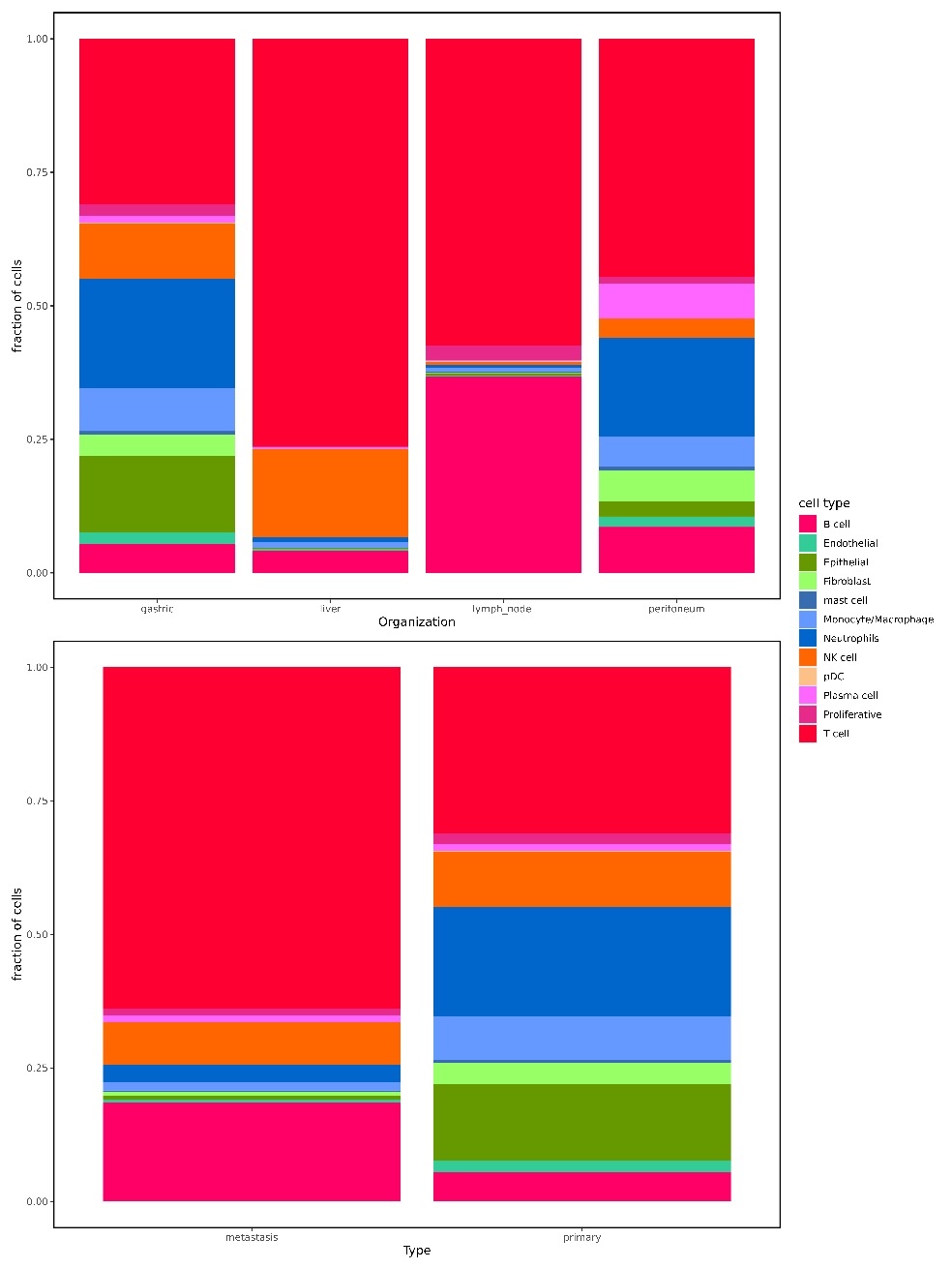


图8. 不同类型细胞在样本中的比例

## 3.5 胃癌内细胞EMT、血管新生、细胞干性之间的相关性

为了研究EMT、血管新生、细胞干性之间的相关性，我们从MSigDB中下载EMT、血管新生、细胞干性相关的基因集，使用ssGSEA量化EMT、血管新生、细胞干性的分数，接着利用R包Hmisc计算EMT、血管新生、细胞干性三者spearman相关性并绘制相关性散点图。

如图7所示，我们分析了EMT、血管新生、细胞干性之间的相关性。所有图在04.EMT\_COR中获取，通过比较发现三者都成正相关，其中细胞干性与EMT具有较高的相关系数为0.52。

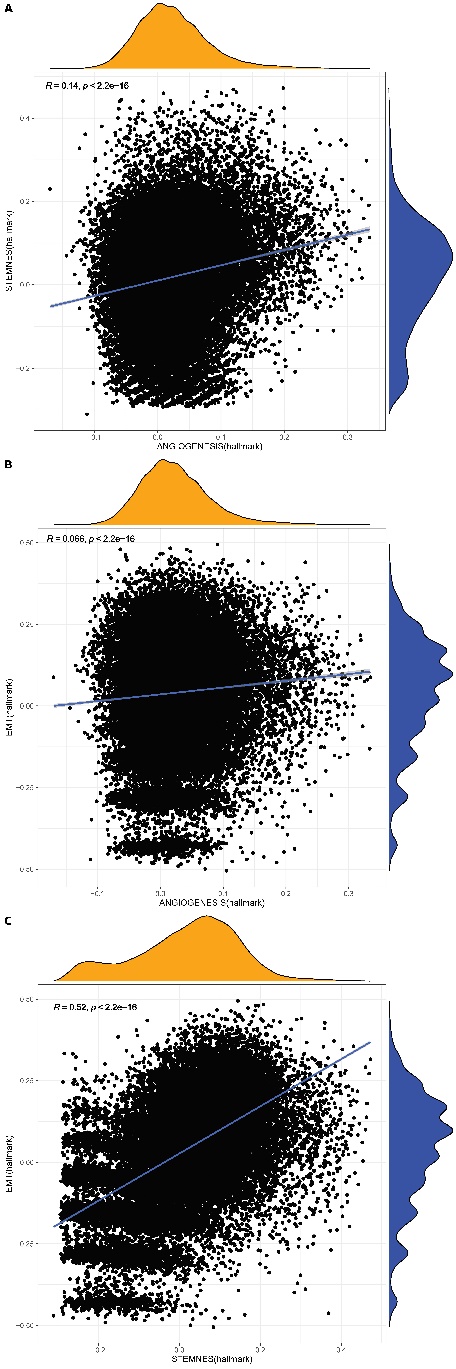


图9. EMT与血管生成的相关性；EMT与细胞干性的相关性；细胞干性与血管生成的相关性

## 3.6 胃癌内Monocyte/Macrophage谱系分析

选取单细胞Monocyte/Macrophage进行亚群分析，通过使用Seurat包中的SCTransform 函数对Monocyte/Macrophage数据集的RNA的表达标准化和缩放。随后进行PCA分析，选取前40个PC纳入后续分析。使用 Seurat 中的“FindClusters”功能进行细胞聚类，并通过表2标记基因的表达对细胞簇进行注释。使用scanpy进行tSNE可视化分析。

表2.胃癌单细胞数据Monocyte/Macrophage亚群中不同细胞类型的标记基因

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 细胞类型 | | 标记基因 | | 来源 | |
| Monocyte | | S100A9、S100A8 | | 测数据数据文献 | |
| Macrophage | | CD68、CD163 | | 胃癌单细胞文献报道 | |
| Macrophage M1 | | TNF | | 巨噬细胞泛癌cell期刊 | |
| Macrophage M2 | | MRC1 | | 巨噬细胞泛癌cell期刊 | |
| Dendritic cell | | CD1C、LAMP3 | | 巨噬细胞泛癌cell期刊 | |

如图10所示，我们可视化tSNE聚类分析得到的细胞簇和不同细胞簇来源的患者以及原发胃癌和转移胃癌细胞的分布，通过观察样本的混合状况，并不需要进行批次校正，来自不同样本的细胞聚类效果好不存在明显的批次效应。可在05. macrophage中获取原始的高DPI图。

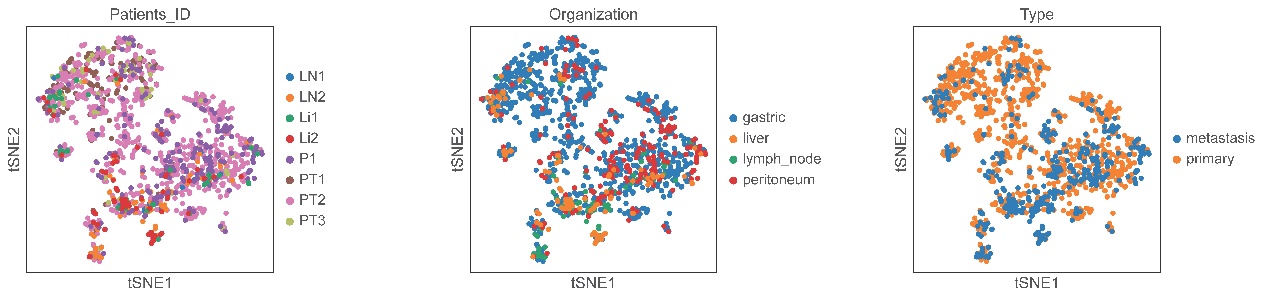


图10.胃癌原发与转移的单细胞细胞簇、患者来源、原发与转移的分类图

通过表2的标记基因，如图11所示，我们可视化原发以及转移胃癌单细胞Monocyte/Macrophage亚群的tSNE图共计1358个细胞。可在05. macrophage中获取原图。其中Macrophage M1有183个细胞、Macrophage M2有738个细胞、Macrophage Unknown有67个细胞、Monocyte有181个细胞、Dendritic cell有189个细胞。

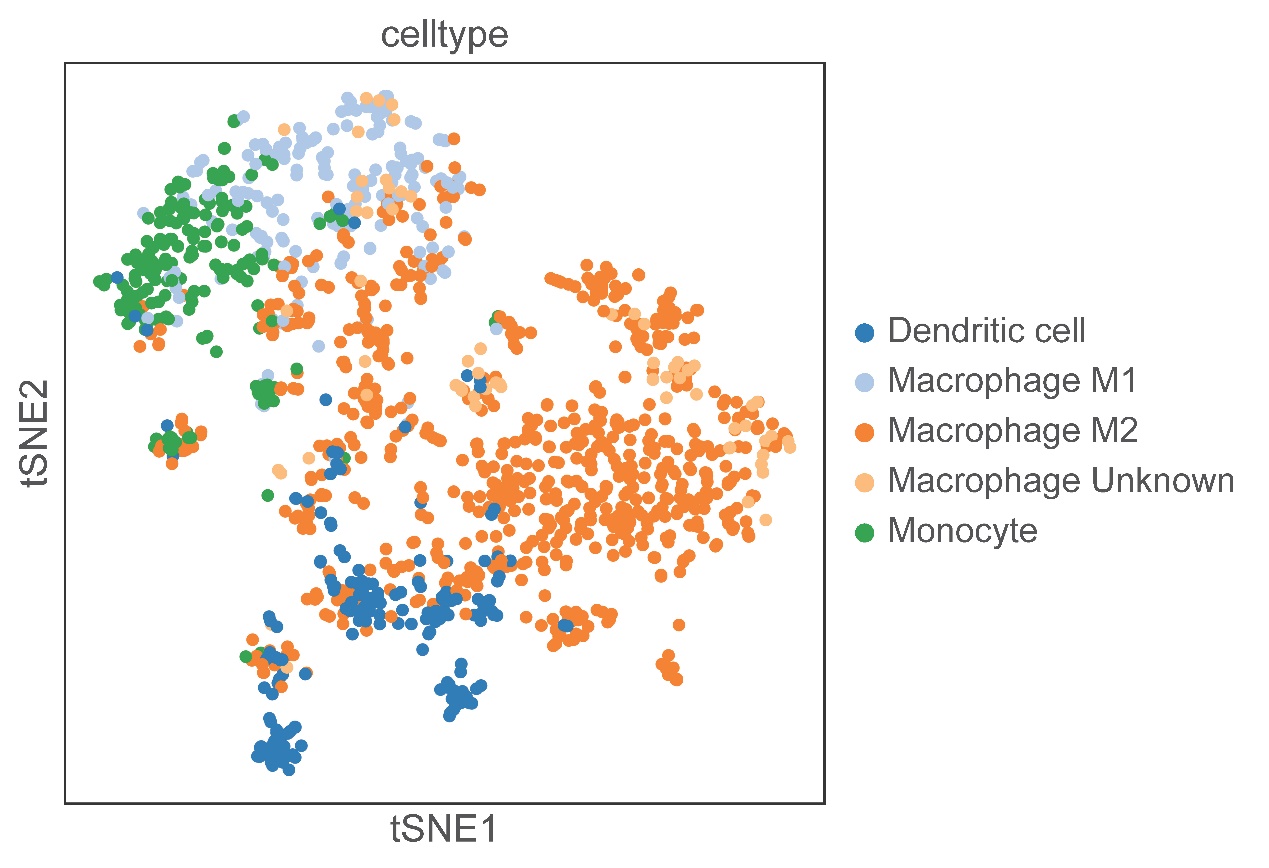


图11.胃癌原发与转移的单细胞Monocyte/Macrophage亚群的tSNE图

同时，我们根据单细胞数据中各种类型细胞的marker基因，使用Seurat包中的FeaturePlot进行可视化marker基因在原发以及转移胃癌单细胞tSNE图的表达。使用Seurat包中的DotPlot进行可视化每种细胞类型的marker基因。

如图12和图13所示，我们可视化marker基因在原发以及转移胃癌单细胞的气泡图和tSNE图的表达，同时在05. macrophage的feature中获取单独的marker基因表达图。

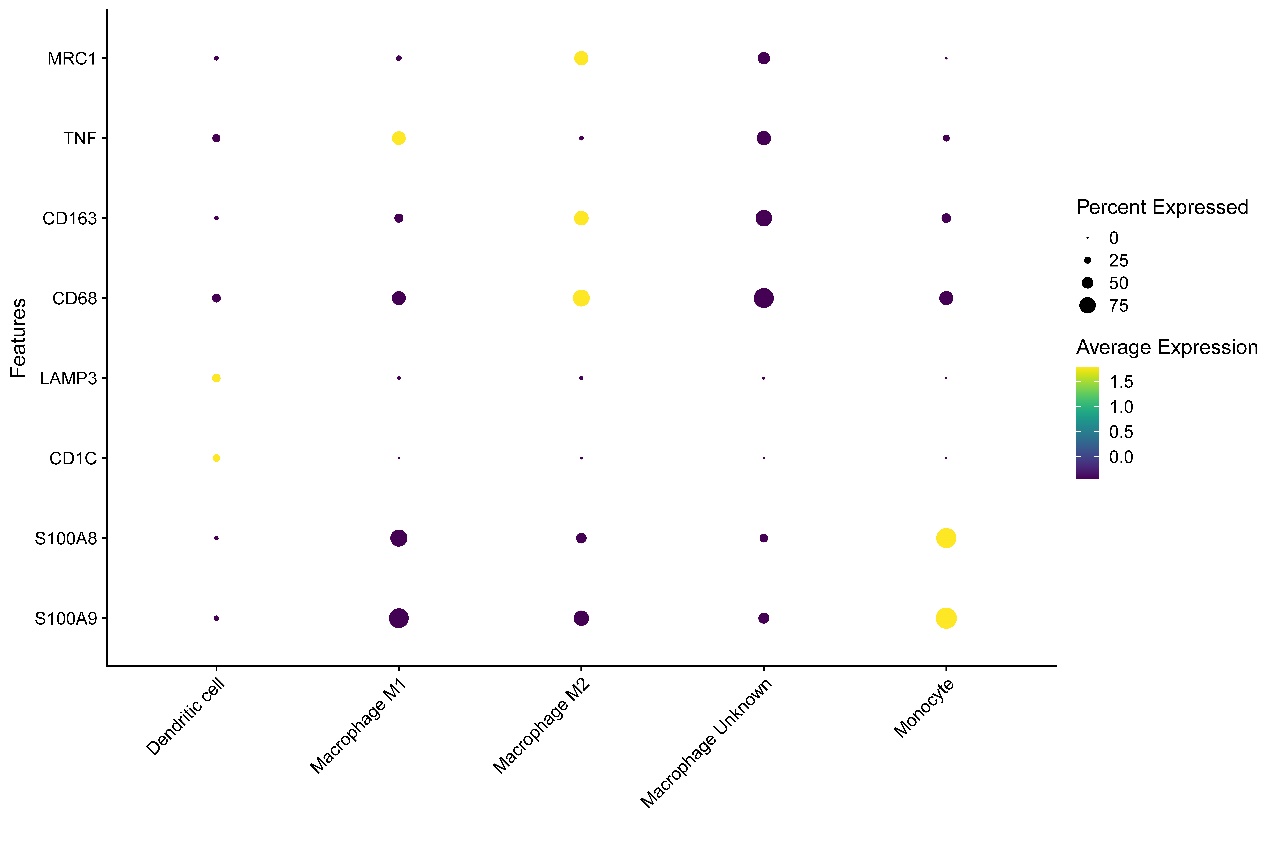


图12.不同细胞类型的marker基因的气泡图

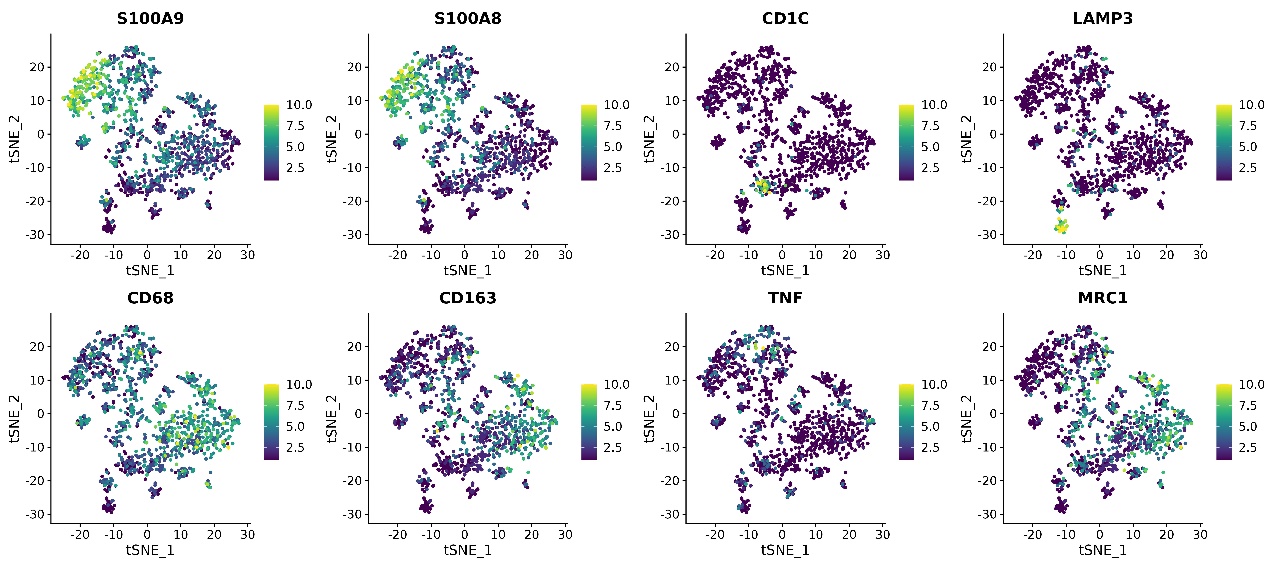


图13.不同细胞类型的marker基因分布图

我们计算原发胃癌与3种转移胃癌的Monocyte/Macrophage亚群中占比比较胃癌细胞转移前后单细胞景观的变化。

如图14所示，我们展示原发与3种转移胃癌的Monocyte/Macrophage亚群中占比。发现Macrophage M2在原发与腹膜转移中占比较多，而Macrophage M1在原发占比多。我们无法鉴定Macrophage类型全都包含在腹膜中。

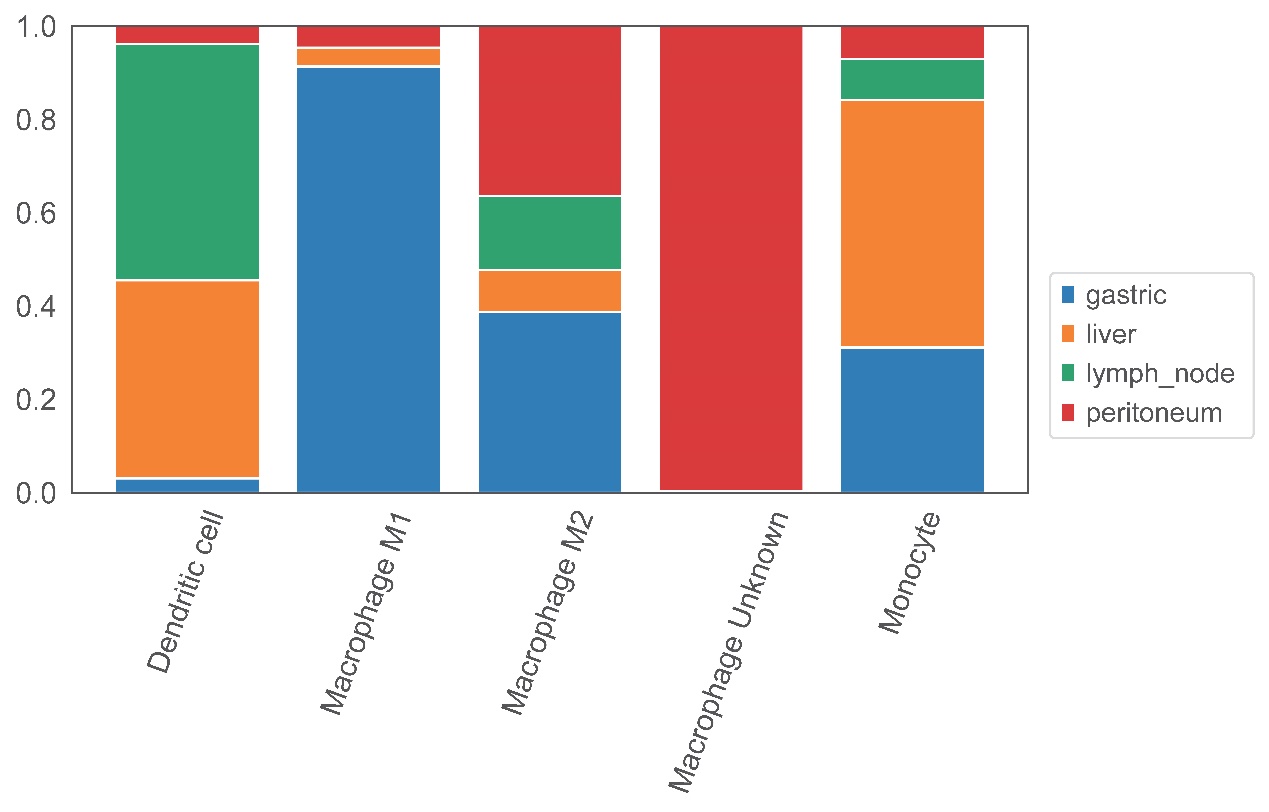


图14. 不同类型细胞在样本中的比例

## 3.7 胃癌内Monocyte/Macrophage细胞EcoTyper分析

EcoTyper 是一种机器学习框架，用于从基因表达数据中识别和表征细胞状态和生态系统，它可以深入了解人类癌症的细胞landscope和群落结构，这是癌症相关死亡率的主要原因。

EcoTyper 采用统计学习算法，包括无监督和有监督的非负矩阵分解 (NMF) 的变体，来识别特定细胞类型的转录程序（“细胞状态”）。肿瘤是由空间和时间相关的细胞状态组成的复杂生态系统，EcoTyper可以鉴定肿瘤生态型进一步了解肿瘤进展中不同细胞组成的生态群的作用。

我们使用EcoTyper对Monocyte/Macrophage 进行分析。如图15所示，可以发现Macrophage M1被分入了生态型E1，Macrophage M2被分入了生态型E3。

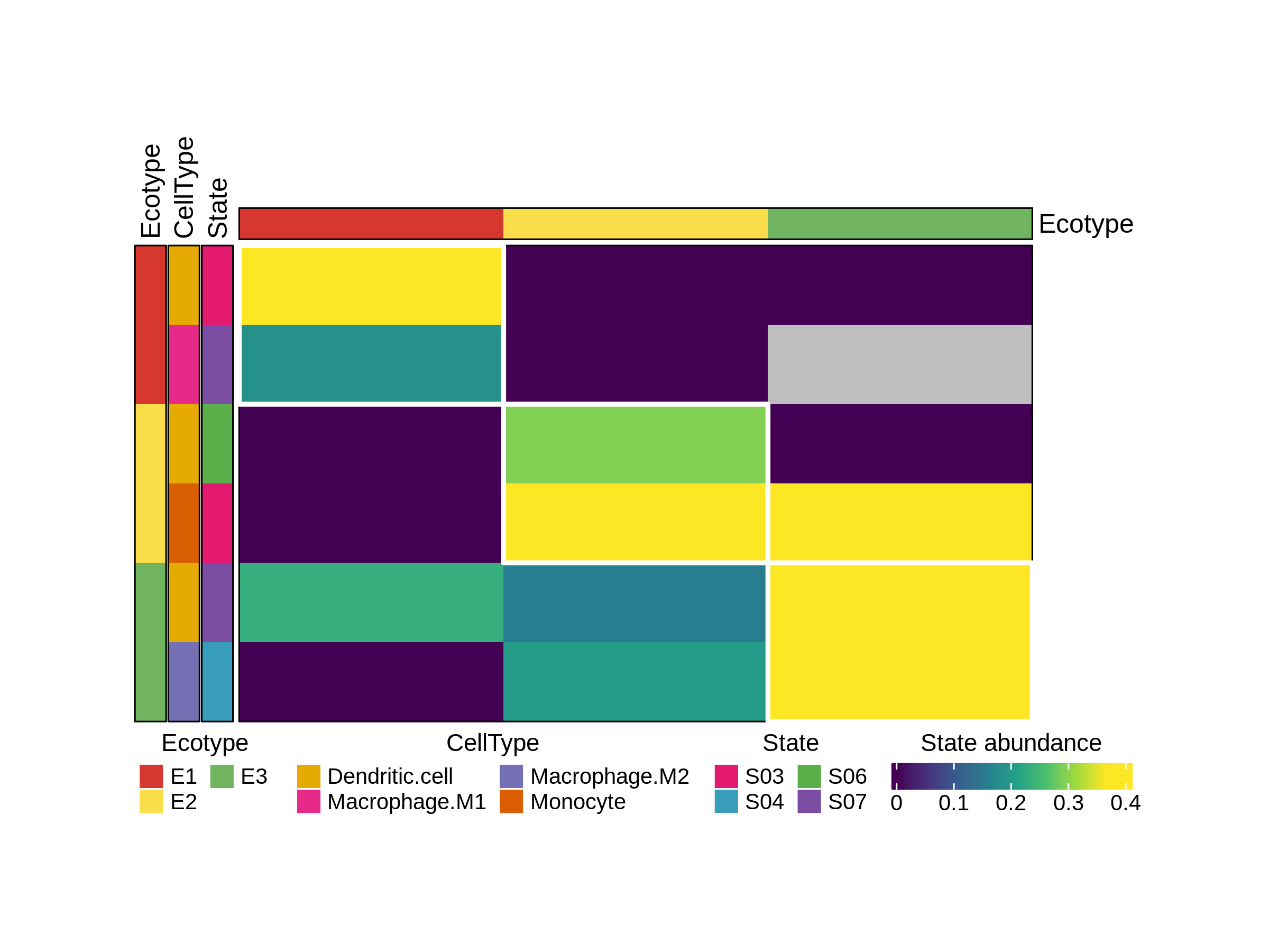


图15. EcoTyper分析热图

为了观察Macrophage M1和Macrophage M2那些细胞状态被分入生态型E1和生态型E3。如图16和17所示，分别展示了Macrophage M1和Macrophage M2细胞状态热图。

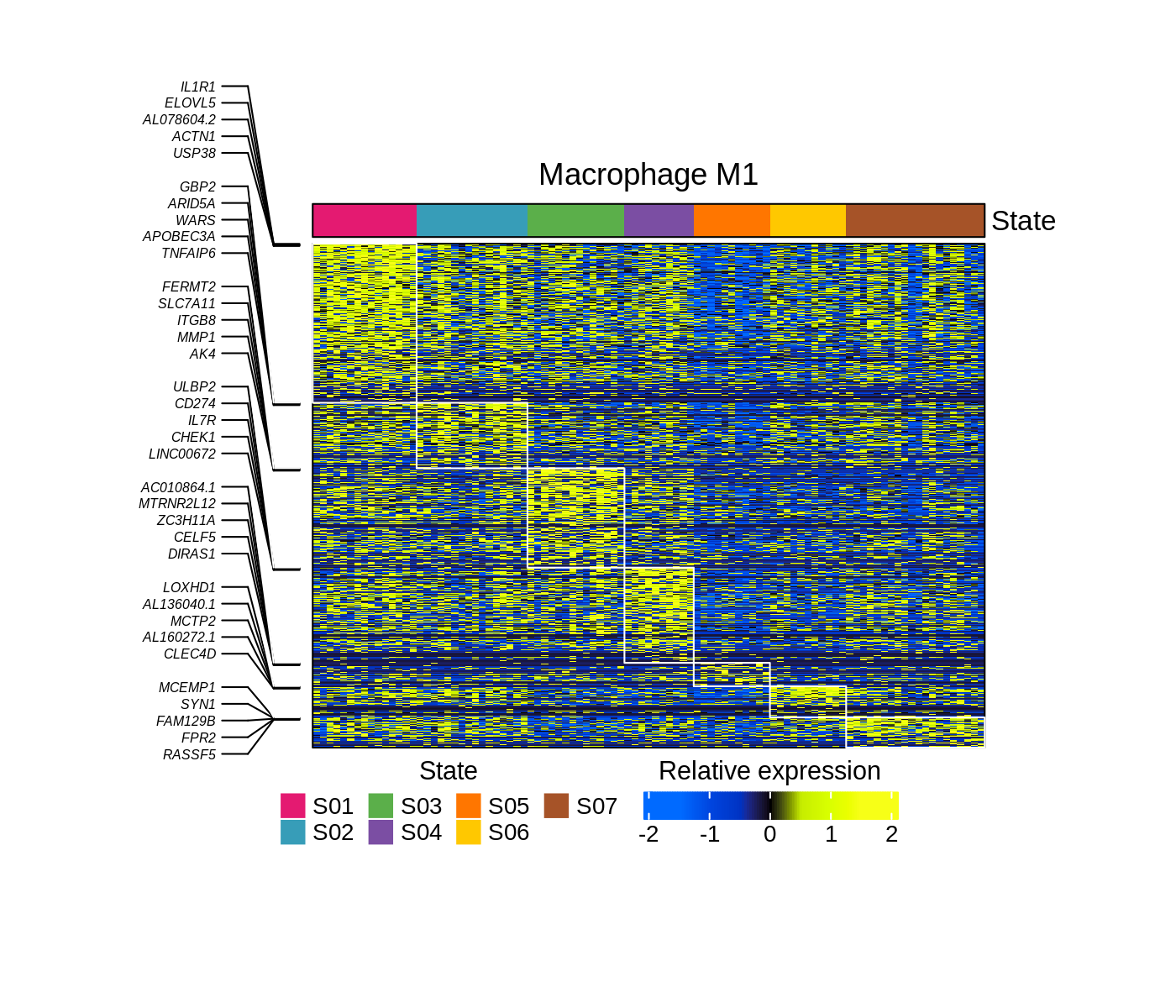


图16 EcoTyper中Macrophage M1细胞状态热图

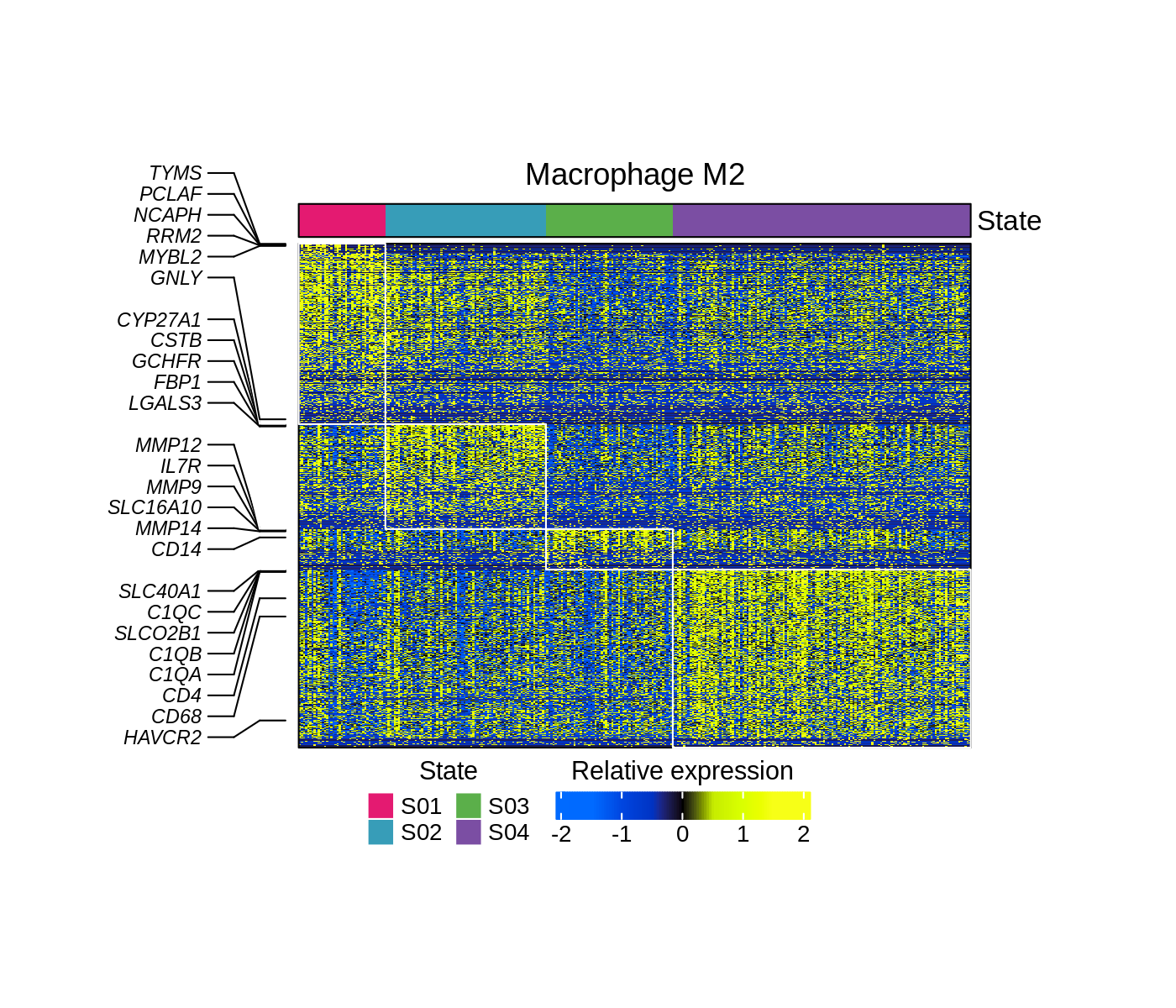


图17 EcoTyper中Macrophage M2细胞状态热图

通过观察我们发现Macrophage M1的S07状态被分入生态型E1, Macrophage M2的S04状态分入生态型E3。因此，我们选取S07状态Macrophage M1和S04状态Macrophage M2纳入后续分析。

## 3.8 不同生态型间M1和M2的差异分析

利用R包MAST在生态型E1肿瘤M1样巨噬细胞和生态型E3肿瘤M2样巨噬细胞间进行差异分析得到差异基因。利用R包Hmisc计算差异基因spearman相关性并绘制相关性散点图。GSEA分析比较差异基因富集到的通路间的差异。

我们选取生态型E1中S07 状态的Macrophage M1样本20个，选取生态型E3中S04 状态的Macrophage M2样本141个进行差异分析，得到了528个差异基因，可在07.Ecological\_diff中获取。随后我们将EcoTyper中获取S04和S07状态的409个特征基因与528个差异基因取交集。如图18所示，最终获取了213个基因，其中上调185个基因，下调23个基因。

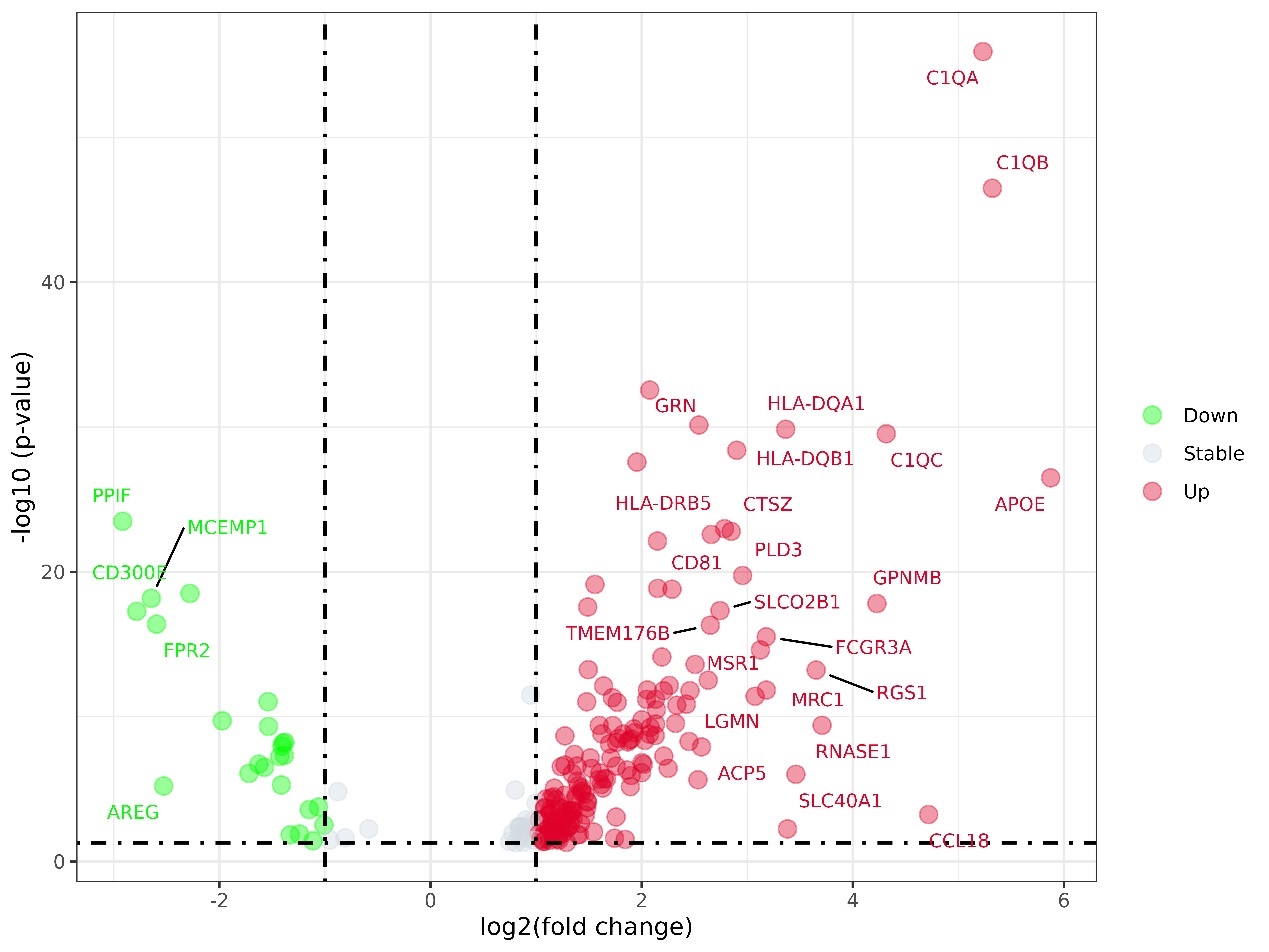


图18.不同生态型 Macrophage M2和Macrophage M1差异火山图

随后，我们将213个差异基因在生态型E1和E3共计161个样本中进行相关性分析，如图19所示，我们选取相关性绝对值大于0.8，p值小于0.01的基因认为具有高度相关，可在07.Ecological\_diff/cor中获取所有结果，共计30个符合结果，由于图过多我们仅展示一张。

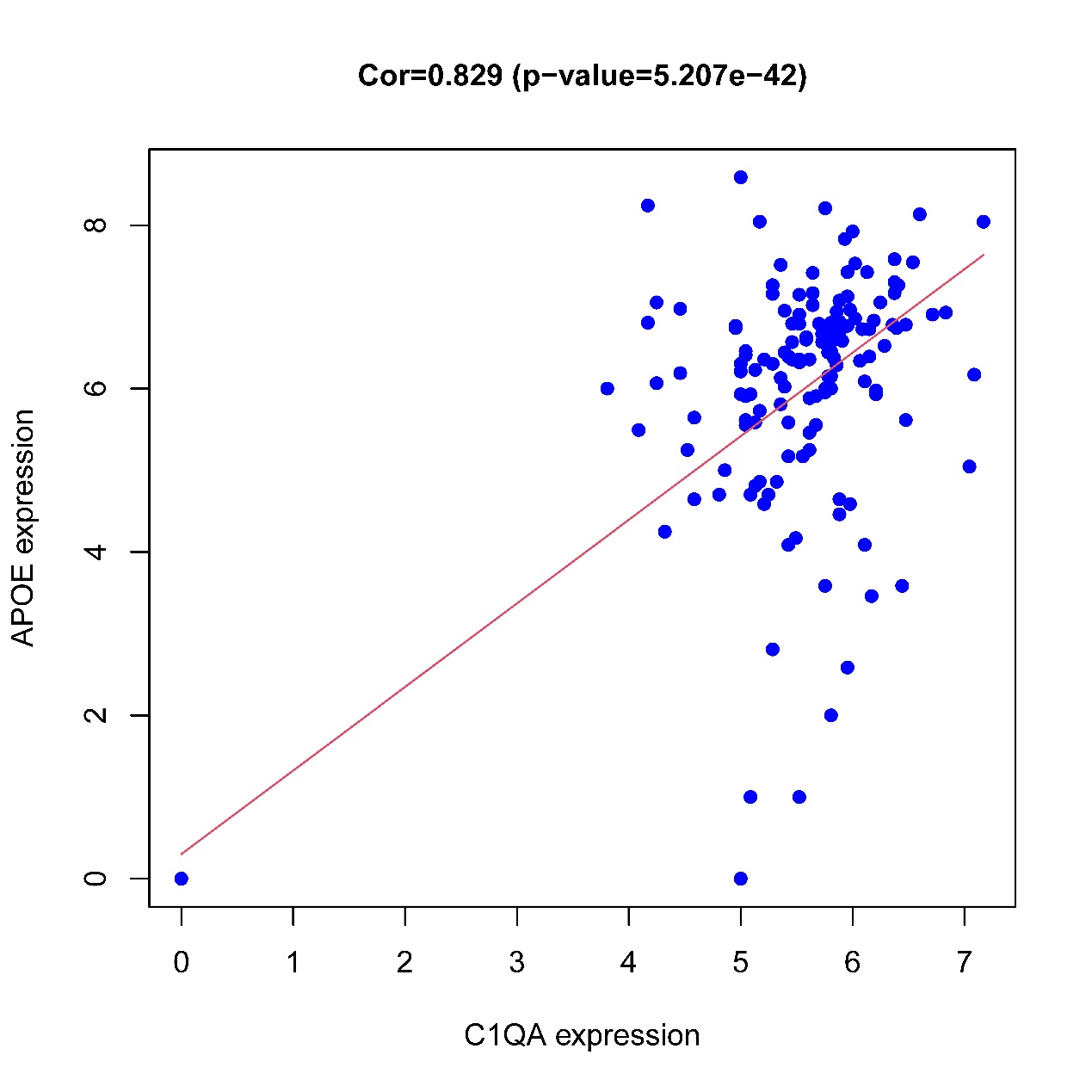


图19.差异基因相关性分析图

为了研究不同生态型间差异基因参与的通路，我们对生态型E3中Macrophage M2和生态型E1中Macrophage M1进行GSVA分析，使用癌症最常用的50组hallmark gene sets （癌症）特征基因集合。如图19所示，我们发现生态型E3中Macrophage M2 主要参与多种信号级联，先前研究表明IL2 STAT5 SIGNALING会促进regulatory T cell的分化抑制肿瘤免疫，从而阻碍人体抑制癌细胞生长。而生态型E1中Macrophage M1主要参与TNFA SIGNALING VIA NFKB在肿瘤中会促进T 细胞受体和 B 细胞受体的分化。

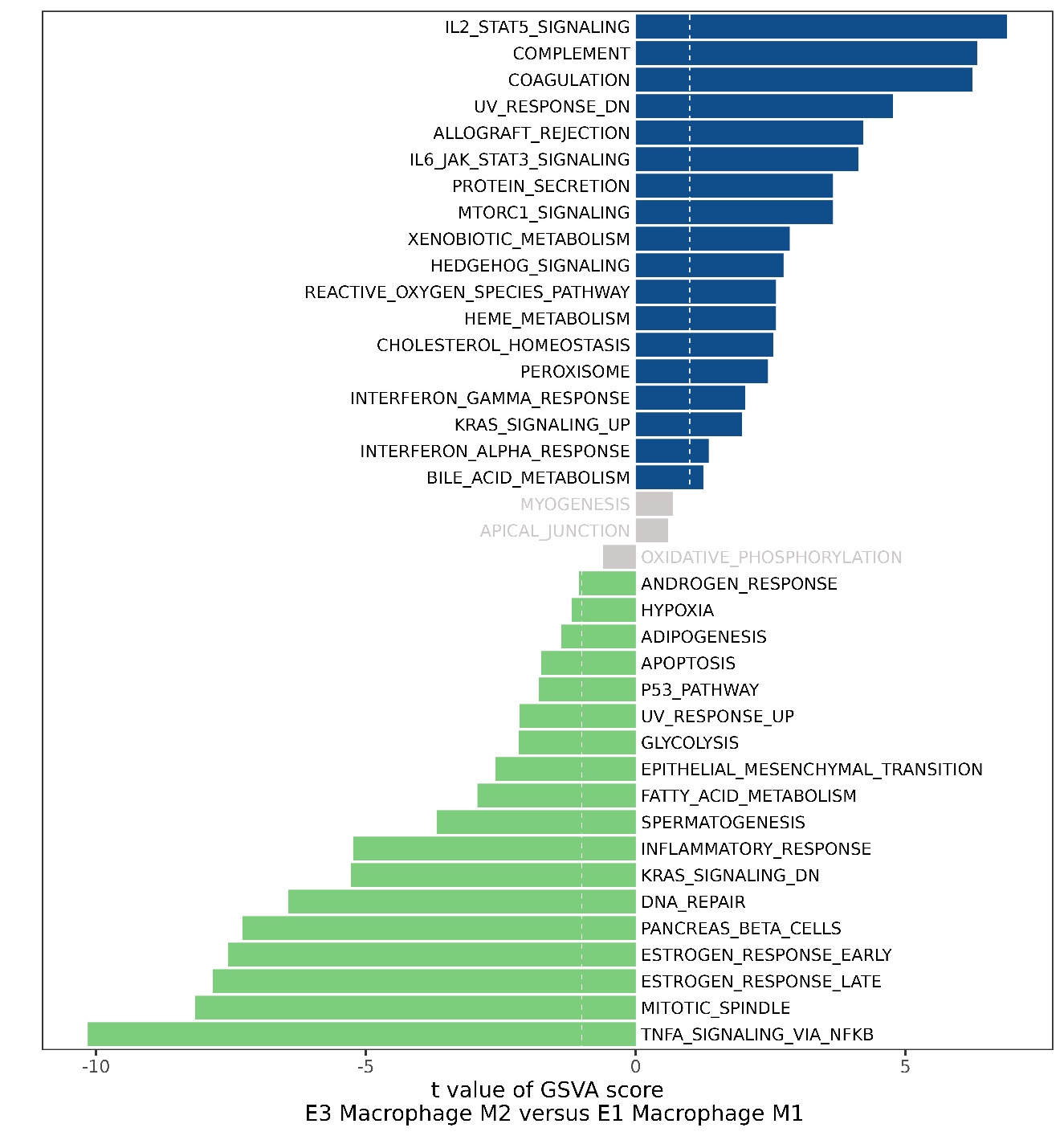


图19. 不同生态型差异基因的GSVA分析

## 3.9 WGCNA分析筛选TAM相关模块基因

用CIBERSORTx计算TCGA-BC样本中TAMs的丰度。使用WGCNA R包进行共表达网络构建，鉴定TAM相关标记基因。

首先使用CIBERSORTx计算TCGA-BC样本中TAMs的丰度，我们认为肿瘤中TAMs主要是Macrophage M2，因此后续分析使用Macrophage M2来阐述TAMs的功能。如图20所示，CIBERSORTx计算TCGA- STAD免疫浸润丰度，通过获取最佳的截止值来区分样本M2型TAM的高低表达，结果表明高表达的M2型TAM与患者预后不良显著相关。因此，得到47个高M2浸润样本和302个低浸润样本。

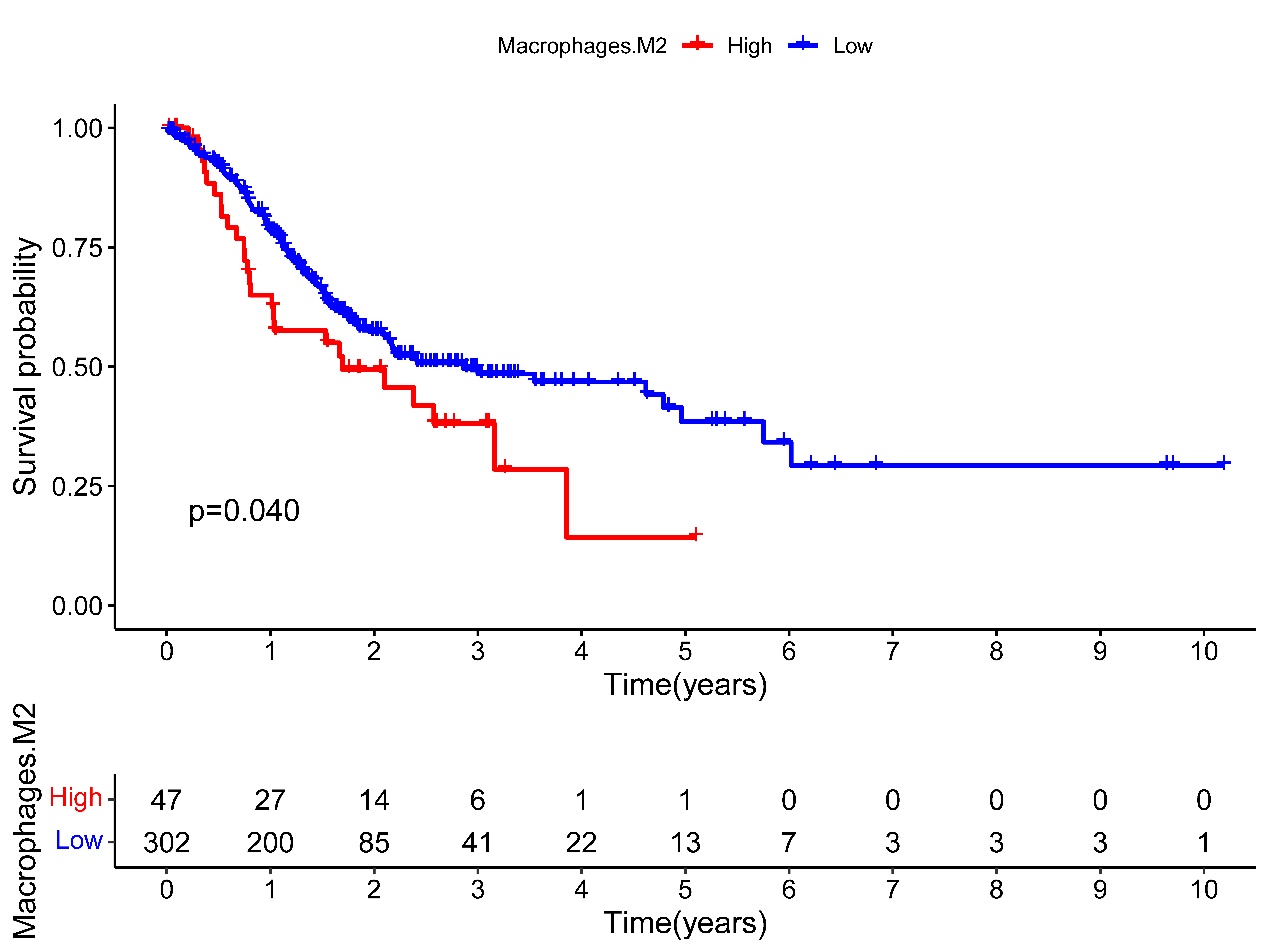


图19. M2巨噬细胞高低表达水平患者的生存分析

WGCNA是用R软件包"WGCNA"进行分析。加权基因共表达网络分析是一种分析多个样本基因表达模式的分析方法，可将表达模式相似的基因进行聚类，分析模块与特定性状或表型之间的关联关系，因此在疾病以及其他性状与基因关联分析等方面的研究中被广泛应用。本研究通过对进行WGCNA 分析，得到与M2型TAM关联的基因。

为了查看数据集里所有样本的整体相关性，我们先对样本进行聚类，查看是否有离群样本。由图20可知无离群样本。

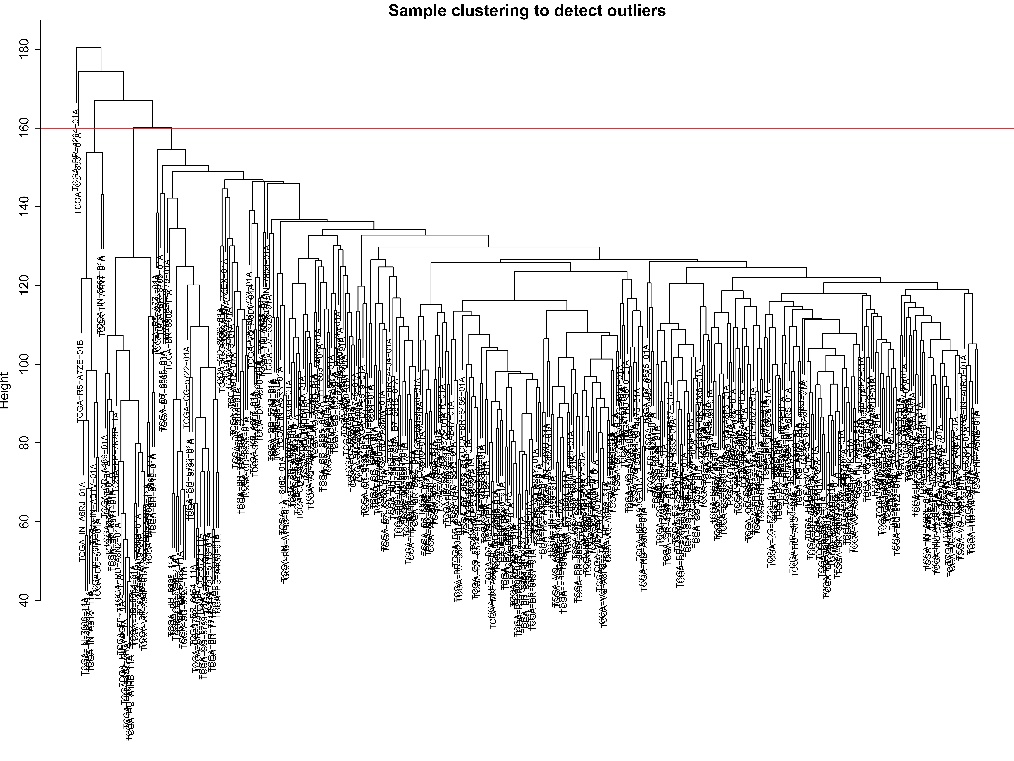


图20. STAD样本聚类分析

为保证基因间相互作用最大限度符合无尺度分布，我们首先对数据进行软阈值的确定，如下图所示：横轴均代表权重参数，左图纵轴scale-free fit index，即signed R2，是对应的网络中log(k)与log(p(k))相关系数的平方R2。相关系数的平方越高，说明该网络越逼近无网路尺度的分布。

如图21所示，右图的纵轴代表对应的基因模块中所有基因邻接函数的均值。根据WGCNA 包的说明，一般选择R2 在0.9 或者0.85 对应的power 作为软阈值。本研究筛选标准为R2=0.9，软阈值设定为3。

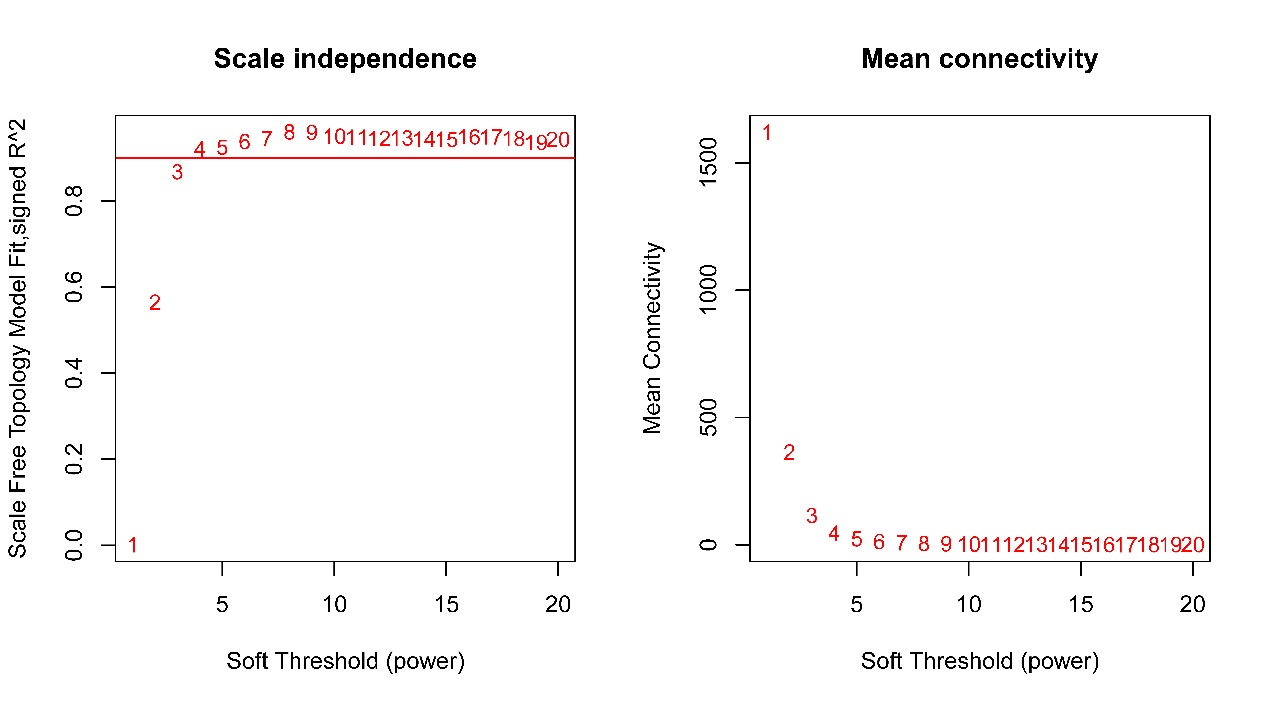


图21. 软阈值选择

构建共表达矩阵的核心是把输入的表达矩阵的基因归类成了多个模块。大体思路：计算基因间的邻接性，根据邻接性计算基因间的相似性，然后推出基因间的相异性系数，并据此得到基因间的系统聚类树。然后按照动态剪切树算法（dynamic tree cutting）的标准，设置每个基因模块最少的基因数目为50。如图22所示，最终生成5个模块。

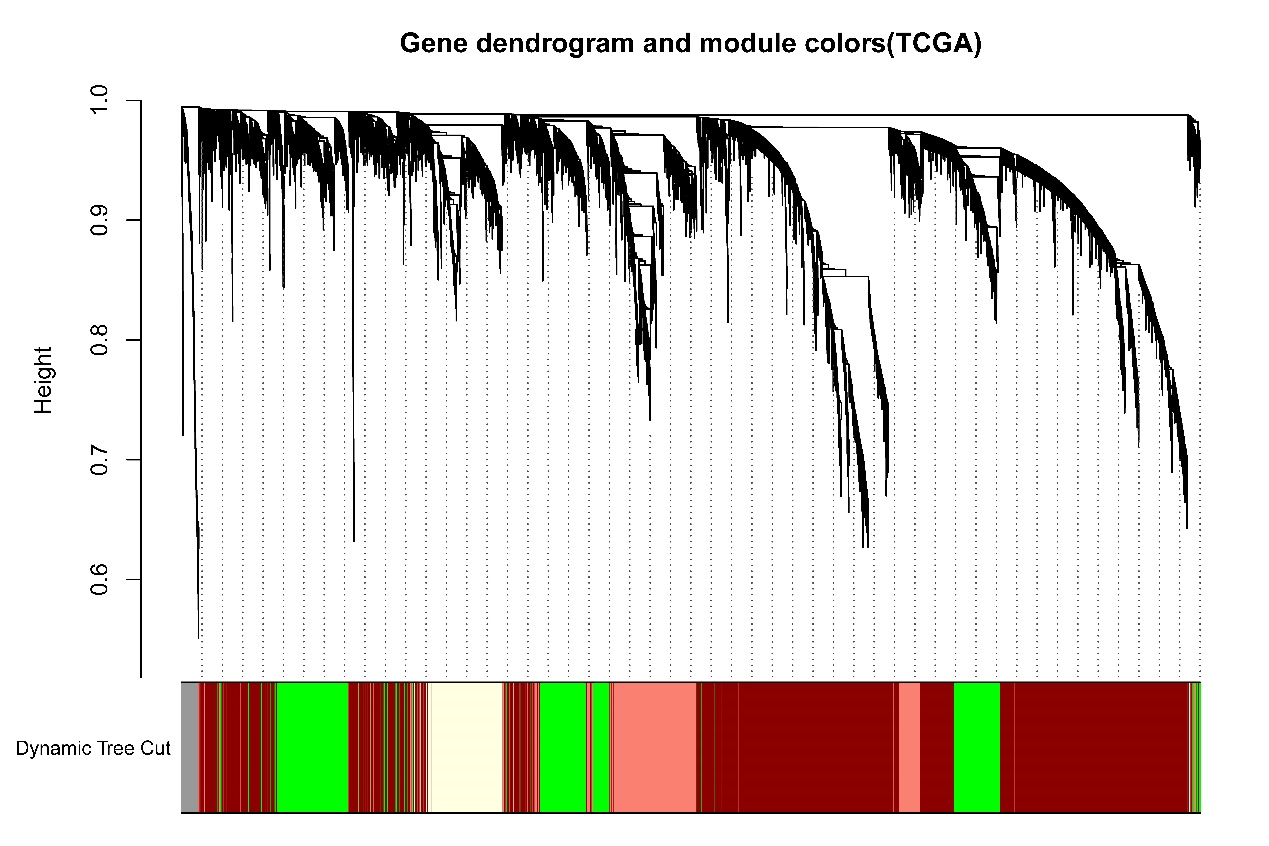


图22 模块共表达网络图

随后进行模块与性状的相关性分析，一共产生了共产生了4个非灰色模块。模块与性状的相关性图，纵坐标为不同模块，横坐标为性状，每一个方块表示某模块和某性状的相关性系数和显著性P 值。在这些模块中，深红色模块的相关度(r = 0.32, p < 0.0001) 被认为是与TAM最相关模块。

如图23所示：

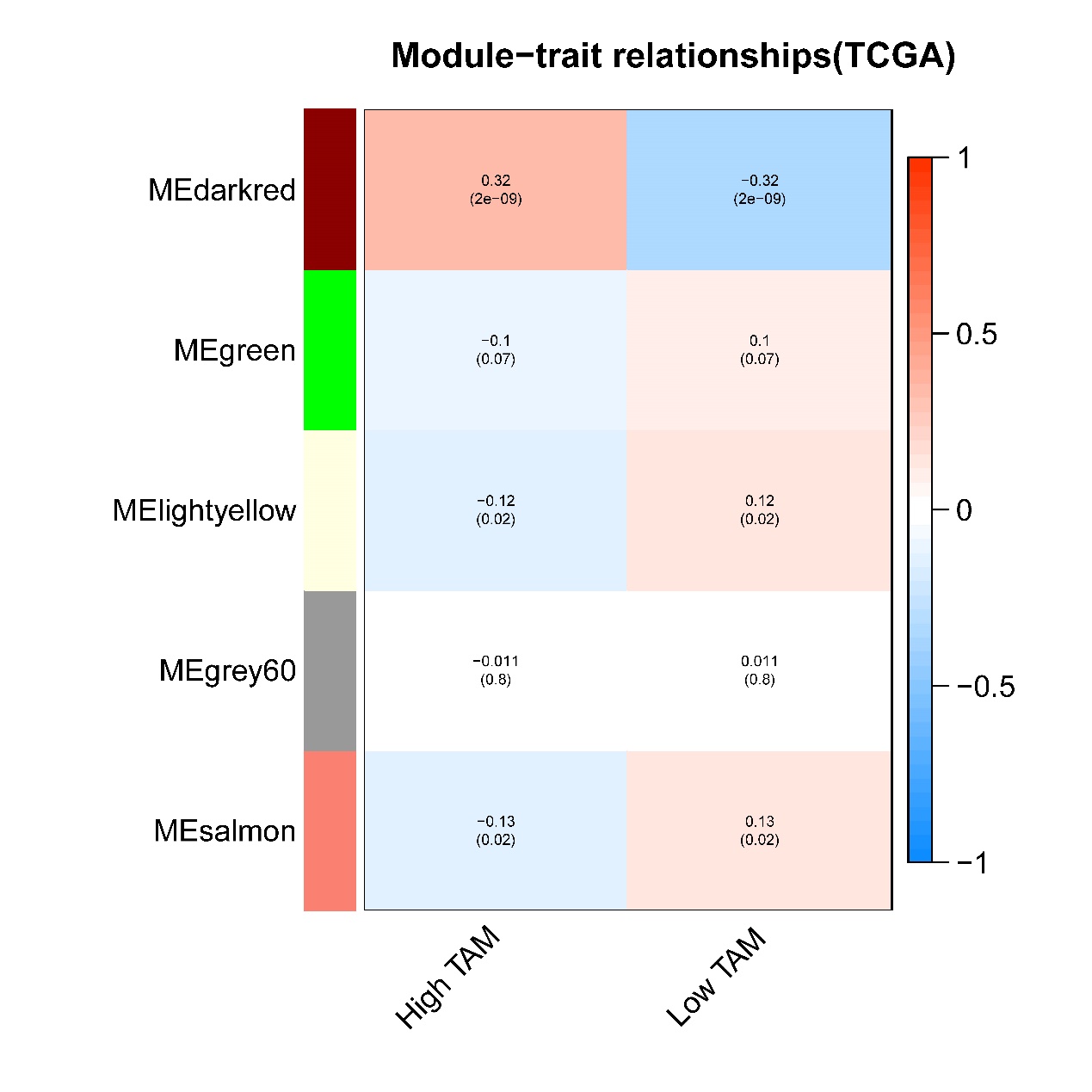


图23 性状与模块的相关性热图

基因模块意义代表了单个基因与TAM密度的关联，而模块量化了模块特征值（"中心点"）和基因表达谱之间的相关性。散点图24显示，在深红色模块中，基因显著性和模块之间存在正相关关系（r = 0.52，P <0.001），表明该模块与M2型TAM高度相关。

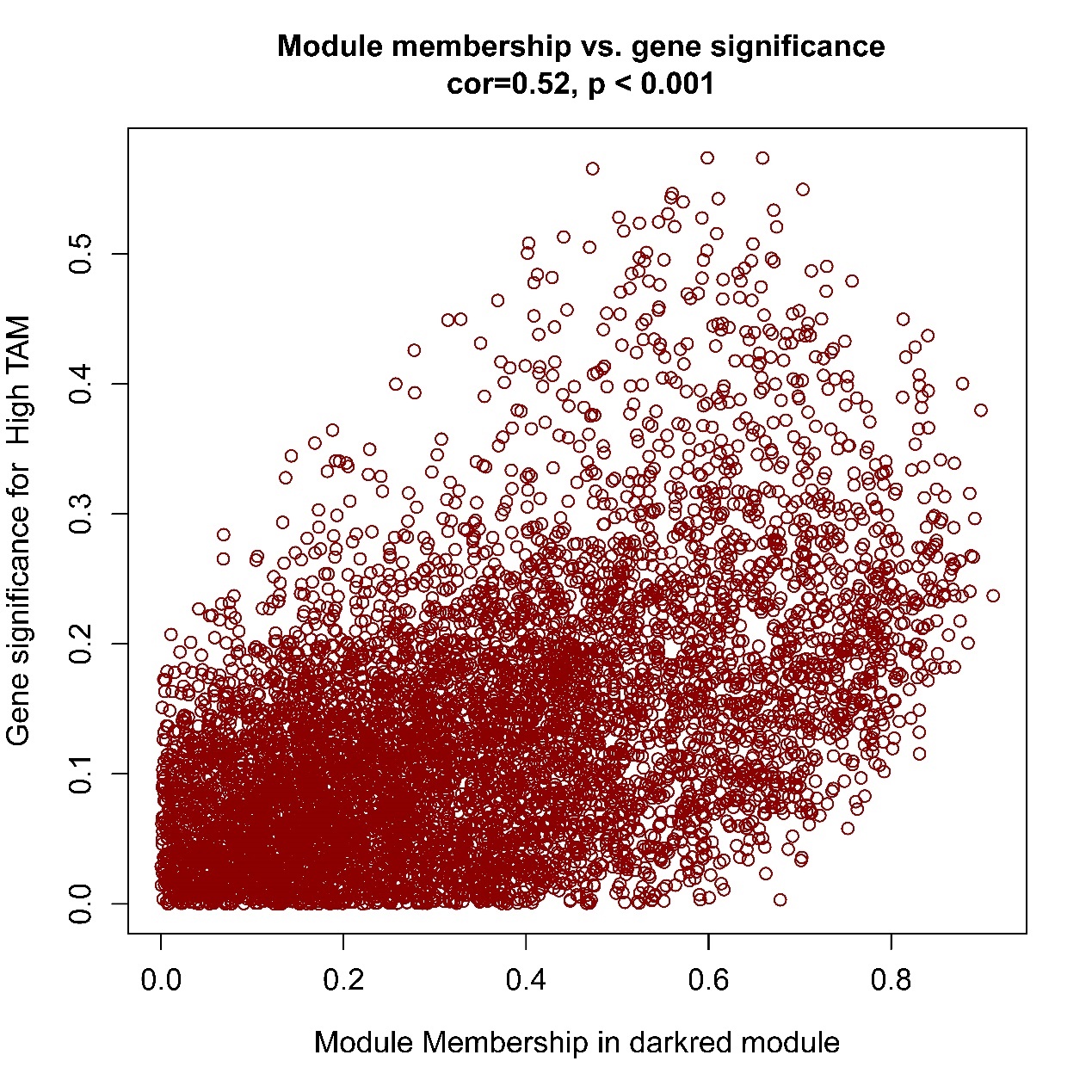


图24 GS 与MM 相关性散点图

如图25所示，来自红色模块的7213个基因与不同生态型得到的差异基因进行取交集获取168个基因。我们认为这168个基因与M2型TAM高度相关。

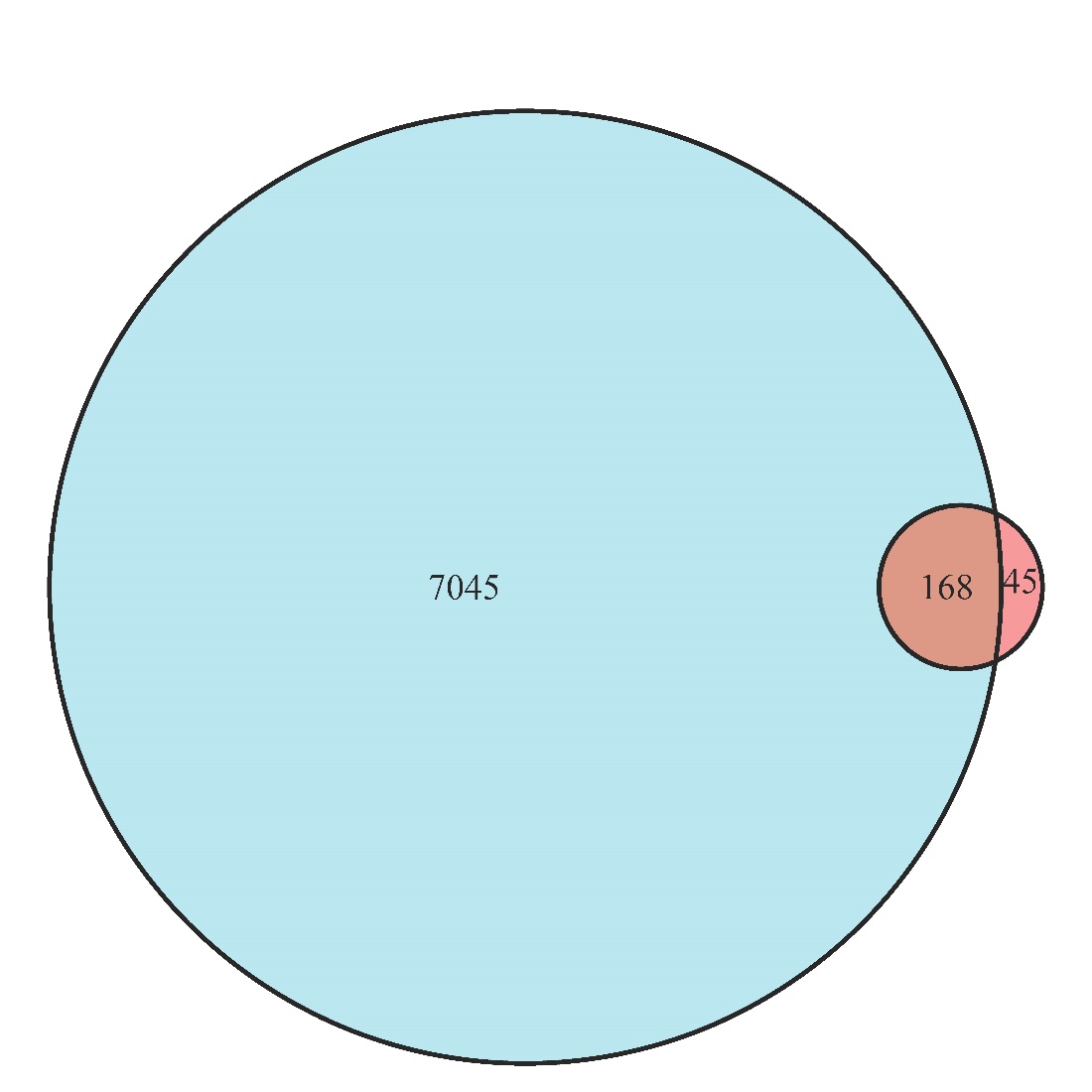


图25 wgcna模块与生态型差异基因交集韦恩图

## 3.10 通过TAM相关标记基因对样本进行分子分型

根据上一步得到的M2型TAM相关标记基因，将GC样本进行一致性聚类。R软件包 "ConsensusClusterPlus "被用来进行基于K-means的共识聚类，基于TAM基因的模型。使用默认参数的 "ConsensusClusterPlus "包获得Delta area图和共识矩阵。设置K最大为8，进行聚类分组。根据M2型TAM相关基因特征的表达模式，进行了基于K-means的共识聚类。

图26所示，将胃癌分为连个的亚组。落石图（Delta area）展示累积分布（CDF）曲线每个k从2到8的共识矩阵的CDF。

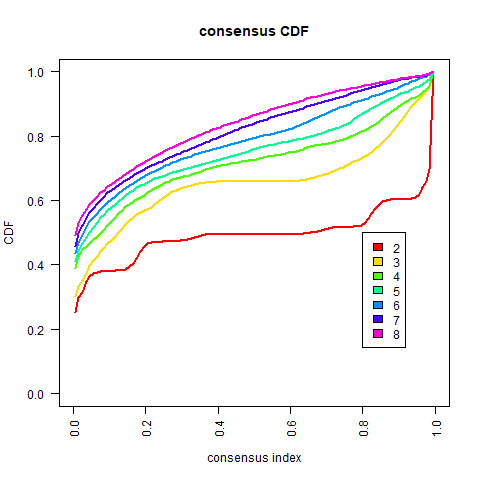


图26. 累积分布图

图27所示，选择k=2作为最佳参数，将TCGA- LIHC分为两个簇，其中A簇有185个样本，B簇有165个样本。

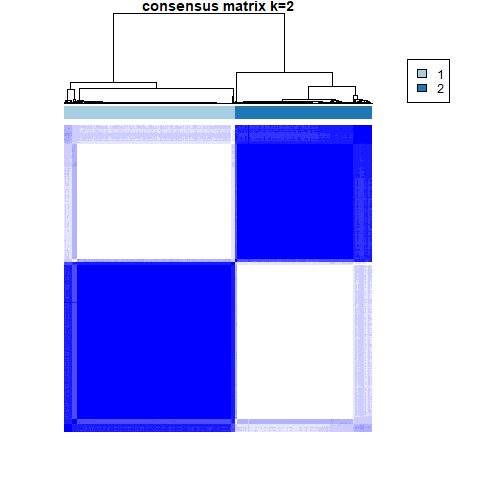


图27. K=2的矩阵热图

如图28所示，热图显示了两个群组之间的M2型TAM相关168个基因表达状况，可以发现A组中M2型TAM基因表达明显高于B组。

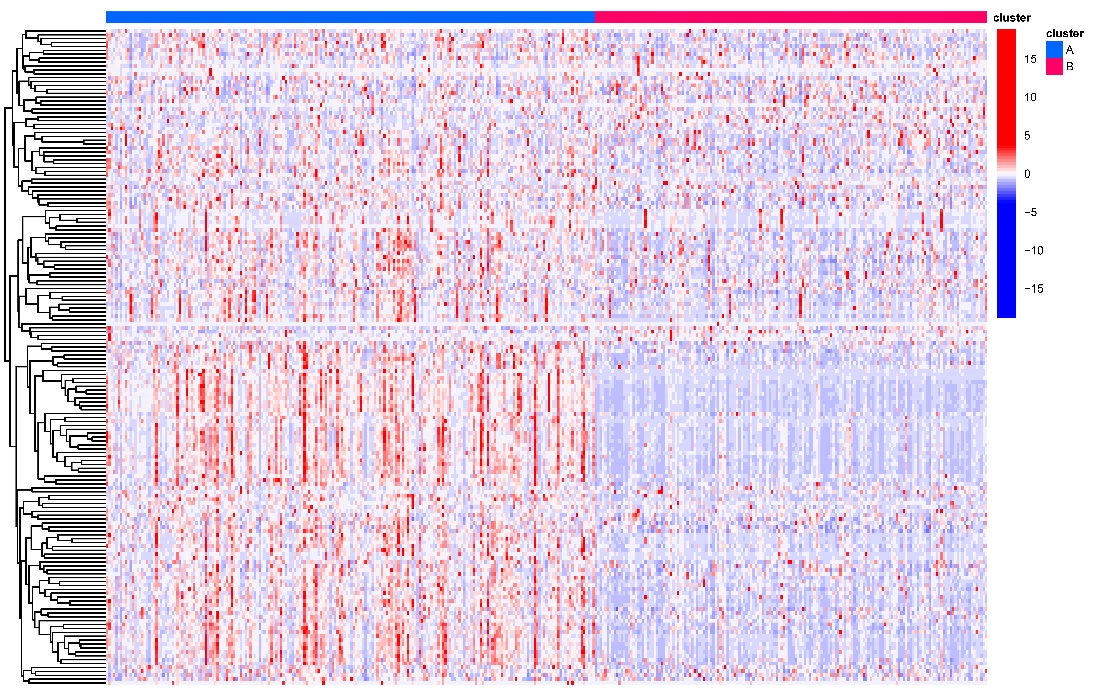


图28 M2型TAM相关基因在不同组的表达

如图29所示，我们使用Kaplan-Meier的方法比较2个不同簇的生存曲线，发现不同簇之间患者预后（p=0.003）。可以发现B组患者的预后明显高于A组。

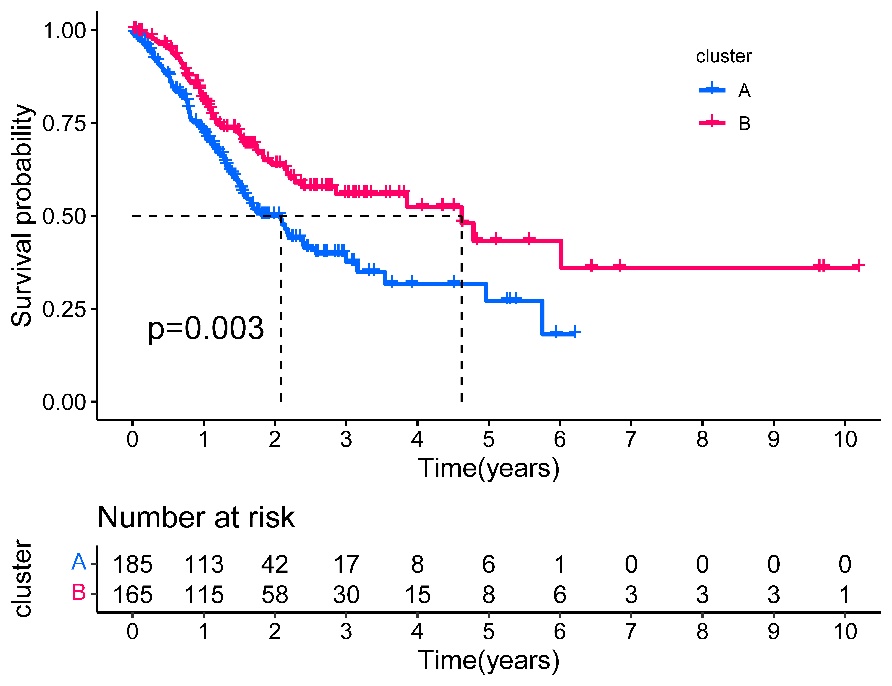


图29 两个簇间生存分析图

如图30所示，我们使用PCA的方法对TCGA-STAD进行将为分析，根据PC1和PC2的系数可视化两个簇的分布，结果发现A簇与B簇显著区分。

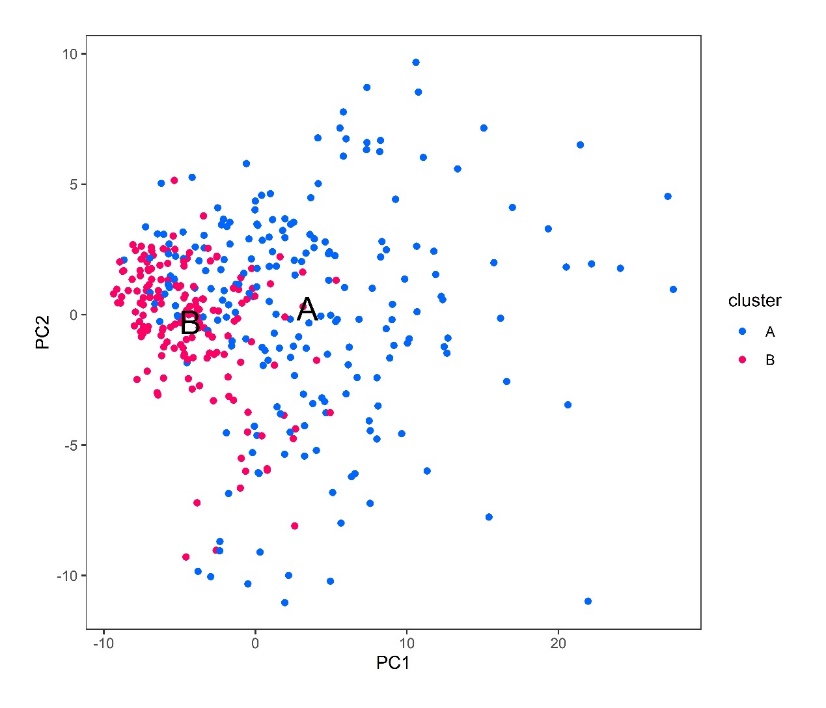


图30 两个簇的PCA分类图

如图31所示，我们比较CIBERSORTx文件中M2型TAM在不同簇的浸润分数，我们发现A组中的M2型TAM浸润分数显著高于B组的M2型TAM浸润分数（p<0.001）。与上述结果呈一致。

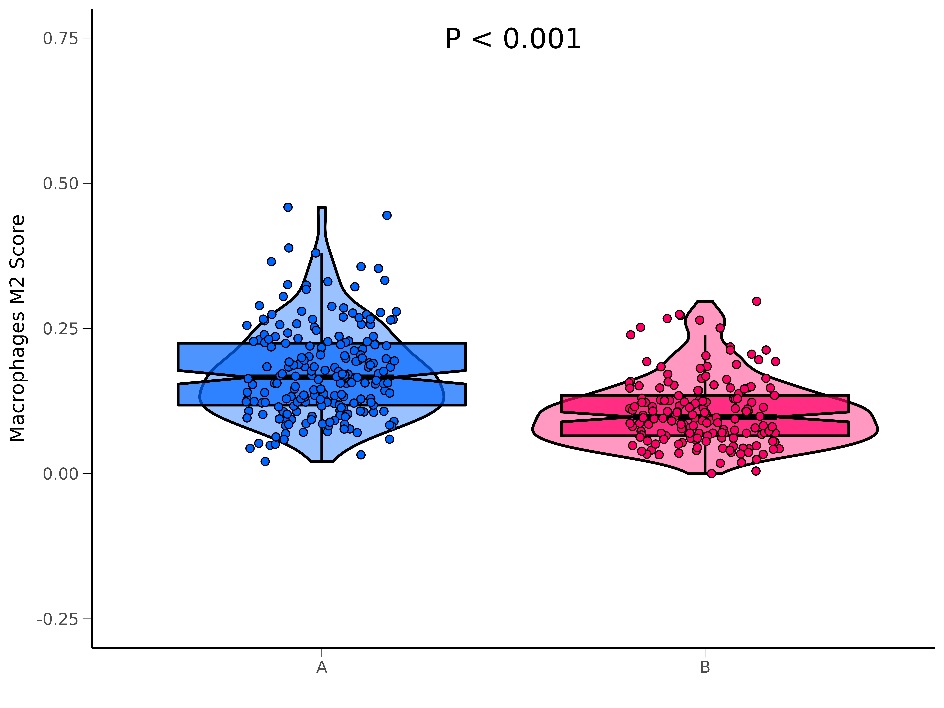


图31 不同簇样本的M2型TAM浸润丰度

为了深入研究两个簇之间的关系，我们使用GSEA分析两个簇之间潜在的生物学关系。

如图32所示，我们筛选A和B簇中GSEA结果FDR<0.05 的结果，我们发现A簇主要参与细胞的迁移例如JAK STAT SIGNALING PATHWAY和MAPK SIGNALING PATHWAY这些都是细胞分化迁移的通路，而A簇主要富集到各种代谢通路如BIOSYNTHESIS OF UNSATURATED FATTY ACIDS这种脂肪酸合成通路。我们发现A簇中更强的增殖分化以及迁移，可能是由于巨噬细胞M2型较多造成的

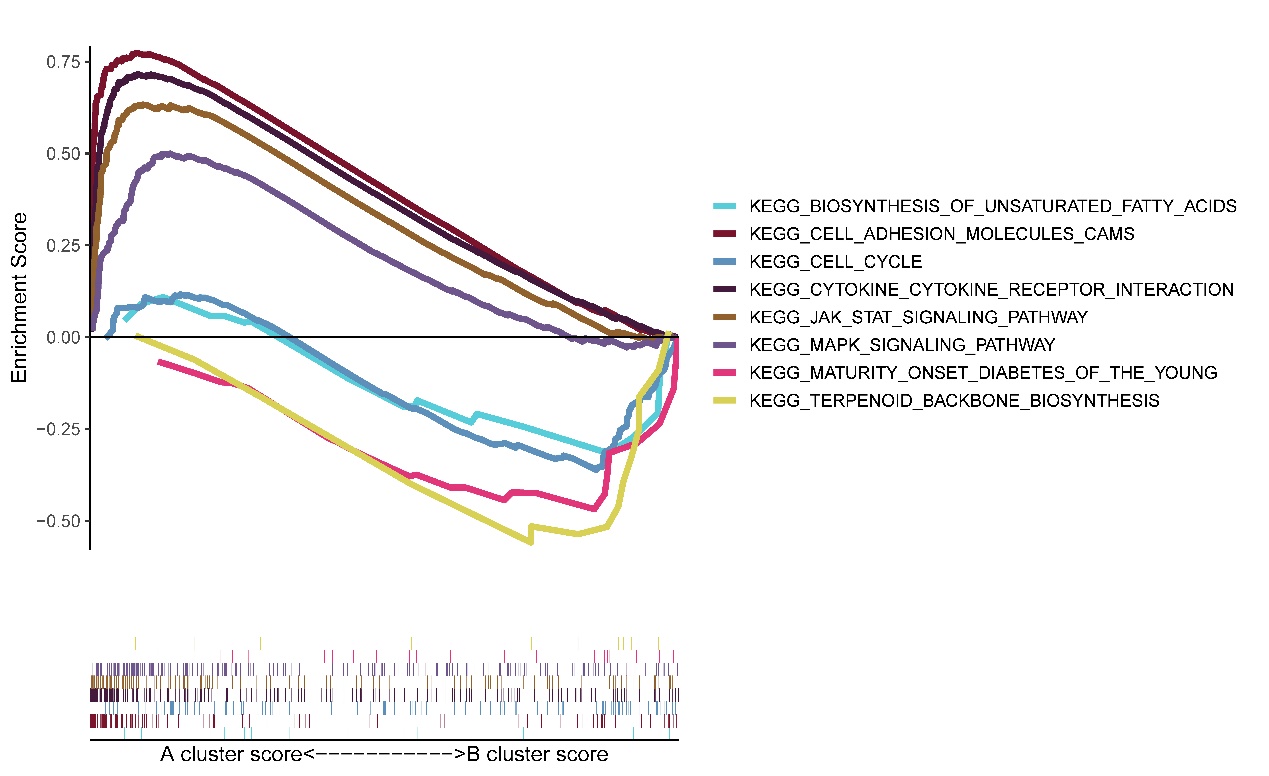


图32 GSEA分析不同簇中上调和下调的通路

# 4 总结

本研究利用GEO公共数据库和UCSC Xena 进行分析，通过探究胃癌单细胞数据构建胃癌单细胞图谱。利用R包Seurat对单细胞测序数据质控，过滤除掉表达基因数异常或线粒体基因表达占比过高的细胞，随后，进行标准化和归一化之后利用PCA和tsne对8个样品的单细胞测序数据进行降维聚类。利用R包CellMarker数据库和文献报道的marker基因对细胞进行注释，以识别不同的细胞类型。之后分别基于来源分组、样品分组、各类细胞marker基因表达情况分别对细胞分群进行可视化；我们分析Monocyte/Macrophage亚群鉴定亚群的细胞类型，随后使用EcoTyper识别Monocyte/Macrophage亚群中不同生态型和细胞状态，接着根据分析不同生态型Macrophage M1和Macrophage M2差异。最后在bulk数据中接着使用CIBERSORTx分析TCGA-STAD数据进行分出高低浸润的Macrophage M2 的TAMs进行WGCNA和一致性聚类分析。

# 5 软件列表

本研究中使用的软件如下：

|  |  |  |
| --- | --- | --- |
| **软件** | **版本** | **用途** |
| EcoTyper |  | 机器学习 |
| clusterProfiler | 4.0.2 | 富集分析 |
| GEOquery | 2.60.0 | GEO数据下载 |
| ggplot2 | 3.3.5 | 统计图绘制 |
| limma | 3.48.3 | 普通转录组差异分析 |
| CIBERSORTx |  | 免疫浸润分析 |
| Org.Hs.eg.db | 3.13.0 | 富集分析 |
| pheatmap | 1.0.12 | 热图绘制 |
| Seurat | 4.0.6 | 单细胞数据分析 |
| VennDiagram | 1.6.20 | Venn图绘制 |
| Cytoscape | 3.8.2 | 网络图绘制 |
| ConsensusClusterPlus | 1.58.0 | 聚类 |
| Scanpy |  | 单细胞数据分析 |

# 6 结果文件列表

BJTC-277/

├── 00.data

│ ├── 01.single\_data

│ │ ├── scRNA\_bacth\_celltype.Rdata

│ │ ├── scRNA\_bath.Rdata

│ │ ├── scRNA\_harmony\_celltype.Rdata

│ │ ├── scRNA\_harmony.Rdata

│ │ ├── scRNAlist.Rdata

│ │ ├── scRNA\_macrophageM2.Rdata

│ │ ├── scRNA\_macrophage.Rdata

│ │ ├── scRNA\_orig.Rdata

│ │ ├── scRNA\_qc.Rdata

│ │ ├── seurat\_EcoTyper.h5ad

│ │ ├── seurat\_EcoTyper.h5seurat

│ │ ├── seurat\_macrophage.h5ad

│ │ ├── seurat\_macrophage.h5seurat

│ │ ├── seurat\_raw.h5ad

│ │ └── seurat\_raw.h5seurat

│ ├── metadata.txt

│ ├── seurat\_10x

│ │ ├── barcodes.tsv

│ │ ├── genes.tsv

│ │ └── matrix.mtx

│ └── single\_data

│ ├── GSM5004180\_PT1

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004181\_PT2

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004182\_PT3

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004184\_LN1

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004185\_LN2

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004187\_P1

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004188\_Li1

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ └── GSM5004189\_Li2

│ ├── barcodes.tsv.gz

│ ├── features.tsv.gz

│ └── matrix.mtx.gz

├── 01.QC

│ ├── 1.QC

│ │ ├── Elbow.jpg

│ │ └── Elbow.pdf

│ ├── 2.UMAP

│ │ ├── all1.jpg

│ │ ├── all1.pdf

│ │ ├── cluster.jpg

│ │ ├── cluster.pdf

│ │ ├── Patients.jpg

│ │ ├── Patients.pdf

│ │ ├── Type.jpg

│ │ └── Type.pdf

│ ├── 3.TSNE

│ │ ├── all2.jpg

│ │ ├── all2.pdf

│ │ ├── cluster.jpg

│ │ ├── cluster.pdf

│ │ ├── Patients.jpg

│ │ ├── Patients.pdf

│ │ ├── Type.jpg

│ │ └── Type.pdf

│ ├── all1.RData

│ ├── all2.RData

│ ├── vlnplot\_after\_qc.jpg

│ ├── vlnplot\_after\_qc.pdf

│ ├── vlnplot\_before\_qc.jpg

│ └── vlnplot\_before\_qc.pdf

├── 02.harmony

│ ├── 1.UMAP

│ │ ├── all1.jpg

│ │ ├── all1.pdf

│ │ ├── all.jpg

│ │ ├── all.pdf

│ │ ├── cluster.jpg

│ │ ├── cluster.pdf

│ │ ├── Patients.jpg

│ │ ├── Patients.pdf

│ │ ├── Type.jpg

│ │ └── Type.pdf

│ └── 2.TSNE

│ ├── all2.jpg

│ ├── all2.pdf

│ ├── all.jpg

│ ├── all.pdf

│ ├── cluster.jpg

│ ├── cluster.pdf

│ ├── Patients.jpg

│ ├── Patients.pdf

│ ├── Type.jpg

│ └── Type.pdf

├── 03.celltype

│ ├── 01.harmony

│ │ ├── 02.UMAP

│ │ │ ├── cell\_DotPlot.jpg

│ │ │ ├── cell\_DotPlot.pdf

│ │ │ ├── cell\_type.jpg

│ │ │ └── cell\_type.pdf

│ │ ├── cell\_barplot1.jpg

│ │ ├── cell\_barplot1.pdf

│ │ ├── cell\_barplot2.jpg

│ │ ├── cell\_barplot2.pdf

│ │ ├── cell\_barplot.jpg

│ │ ├── cell\_barplot.pdf

│ │ ├── cell\_DotPlot.jpg

│ │ ├── cell\_DotPlot.pdf

│ │ ├── cell\_Feature.jpg

│ │ ├── cell\_Feature.pdf

│ │ ├── cell\_type.jpg

│ │ ├── cell\_type.pdf

│ │ ├── celltype.R

│ │ ├── diff\_genes\_mast.csv

│ │ ├── feature

│ │ │ ├── CD2.jpg

│ │ │ ├── CD2.pdf

│ │ │ ├── CD3D.jpg

│ │ │ ├── CD3D.pdf

│ │ │ ├── CD3E.jpg

│ │ │ ├── CD3E.pdf

│ │ │ ├── CD68.jpg

│ │ │ ├── CD68.pdf

│ │ │ ├── CD79A.jpg

│ │ │ ├── CD79A.pdf

│ │ │ ├── CLDN4.jpg

│ │ │ ├── CLDN4.pdf

│ │ │ ├── CSF1R.jpg

│ │ │ ├── CSF1R.pdf

│ │ │ ├── CSF3R.jpg

│ │ │ ├── CSF3R.pdf

│ │ │ ├── EPCAM.jpg

│ │ │ ├── EPCAM.pdf

│ │ │ ├── FN1.jpg

│ │ │ ├── FN1.pdf

│ │ │ ├── GNLY.jpg

│ │ │ ├── GNLY.pdf

│ │ │ ├── IGHG1.jpg

│ │ │ ├── IGHG1.pdf

│ │ │ ├── KDR.jpg

│ │ │ ├── KDR.pdf

│ │ │ ├── KIT.jpg

│ │ │ ├── KIT.pdf

│ │ │ ├── KLRD1.jpg

│ │ │ ├── KLRD1.pdf

│ │ │ ├── KLRF1.jpg

│ │ │ ├── KLRF1.pdf

│ │ │ ├── KRT19.jpg

│ │ │ ├── KRT19.pdf

│ │ │ ├── MKI67.jpg

│ │ │ ├── MKI67.pdf

│ │ │ ├── MMP2.jpg

│ │ │ ├── MMP2.pdf

│ │ │ ├── MS4A1.jpg

│ │ │ ├── MS4A1.pdf

│ │ │ ├── MZB1.jpg

│ │ │ ├── MZB1.pdf

│ │ │ ├── PCNA.jpg

│ │ │ ├── PCNA.pdf

│ │ │ ├── PDGFRA.jpg

│ │ │ ├── PDGFRA.pdf

│ │ │ ├── PECAM1.jpg

│ │ │ ├── PECAM1.pdf

│ │ │ ├── STMN1.jpg

│ │ │ ├── STMN1.pdf

│ │ │ ├── VWF.jpg

│ │ │ └── VWF.pdf

│ │ ├── top10\_diff\_genes\_mast.csv

│ │ └── type.csv

│ └── 02.bath

│ ├── cell\_barplot1.jpg

│ ├── cell\_barplot1.pdf

│ ├── cell\_barplot2.jpg

│ ├── cell\_barplot2.pdf

│ ├── cell\_barplot3.pdf

│ ├── cell\_barplot4.jpg

│ ├── cell\_barplot4.pdf

│ ├── cell\_barplot.jpg

│ ├── cell\_barplot.pdf

│ ├── cell\_DotPlot.jpg

│ ├── cell\_DotPlot.pdf

│ ├── cell\_Feature.jpg

│ ├── cell\_Feature.pdf

│ ├── cell\_type.jpg

│ ├── cell\_type.pdf

│ ├── celltype.R

│ ├── diff\_genes\_mast.csv

│ ├── feature

│ │ ├── CD163.jpg

│ │ ├── CD163.pdf

│ │ ├── CD2.jpg

│ │ ├── CD2.pdf

│ │ ├── CD3D.jpg

│ │ ├── CD3D.pdf

│ │ ├── CD68.jpg

│ │ ├── CD68.pdf

│ │ ├── CD79A.jpg

│ │ ├── CD79A.pdf

│ │ ├── CSF3R.jpg

│ │ ├── CSF3R.pdf

│ │ ├── EPCAM.jpg

│ │ ├── EPCAM.pdf

│ │ ├── FCGR3A.jpg

│ │ ├── FCGR3A.pdf

│ │ ├── FN1.jpg

│ │ ├── FN1.pdf

│ │ ├── KIT.jpg

│ │ ├── KIT.pdf

│ │ ├── KLRD1.jpg

│ │ ├── KLRD1.pdf

│ │ ├── LILRA4.jpg

│ │ ├── LILRA4.pdf

│ │ ├── MKI67.jpg

│ │ ├── MKI67.pdf

│ │ ├── MS4A1.jpg

│ │ ├── MS4A1.pdf

│ │ ├── MZB1.jpg

│ │ ├── MZB1.pdf

│ │ ├── PECAM1.jpg

│ │ └── PECAM1.pdf

│ ├── top10\_diff\_genes\_mast.csv

│ └── type.csv

├── 04.EMT\_cor

│ ├── all1.jpg

│ ├── all1.pdf

│ ├── all.jpg

│ ├── all.pdf

│ ├── angiogenesis.txt

│ ├── cor1.jpg

│ ├── cor1.pdf

│ ├── cor2.jpg

│ ├── cor2.pdf

│ ├── cor3.jpg

│ ├── cor3.pdf

│ └── gsva\_result.csv

├── 05.macrophage

│ ├── barplot.pdf

│ ├── cell\_DotPlot.jpg

│ ├── cell\_DotPlot.pdf

│ ├── cell\_Feature.jpg

│ ├── cell\_Feature.pdf

│ ├── cell\_type.jpg

│ ├── cell\_type.pdf

│ ├── cluster.pdf

│ ├── diff\_genes\_mast.csv

│ ├── feature

│ │ ├── CD163.jpg

│ │ ├── CD163.pdf

│ │ ├── CD1C.jpg

│ │ ├── CD1C.pdf

│ │ ├── CD68.jpg

│ │ ├── CD68.pdf

│ │ ├── LAMP3.jpg

│ │ ├── LAMP3.pdf

│ │ ├── MRC1.jpg

│ │ ├── MRC1.pdf

│ │ ├── S100A8.jpg

│ │ ├── S100A8.pdf

│ │ ├── S100A9.jpg

│ │ ├── S100A9.pdf

│ │ ├── TNF.jpg

│ │ └── TNF.pdf

│ ├── figures

│ │ ├── barplot.pdf

│ │ ├── cluster.pdf

│ │ ├── tsnecelltype.pdf

│ │ └── umapcelltype\_umap.pdf

│ ├── top10\_diff\_genes\_mast.csv

│ ├── tsnecelltype.pdf

│ ├── type.csv

│ └── umapcelltype\_umap.pdf

├── 06.EcoTyper

│ └── DiscoveryOutput3\_scRNA

│ ├── Ecotypes

│ │ ├── ecotype\_abundance.txt

│ │ ├── ecotype\_assignment.txt

│ │ ├── ecotypes.txt

│ │ ├── heatmap\_assigned\_samples\_viridis.pdf

│ │ ├── heatmap\_assigned\_samples\_viridis.png

│ │ ├── jaccard\_matrix.pdf

│ │ ├── jaccard\_matrix.png

│ │ ├── nclusters\_jaccard.pdf

│ │ └── nclusters\_jaccard.png

│ ├── Macrophage.M1

│ │ ├── gene\_info.txt

│ │ ├── heatmap\_data.txt

│ │ ├── heatmap\_top\_ann.txt

│ │ ├── state\_abundances.txt

│ │ ├── state\_assignment\_heatmap.pdf

│ │ ├── state\_assignment\_heatmap.png

│ │ └── state\_assignment.txt

│ ├── Macrophage.M2

│ │ ├── gene\_info.txt

│ │ ├── heatmap\_data.txt

│ │ ├── heatmap\_top\_ann.txt

│ │ ├── state\_abundances.txt

│ │ ├── state\_assignment\_heatmap.pdf

│ │ ├── state\_assignment\_heatmap.png

│ │ └── state\_assignment.txt

│ ├── rank\_data.txt

│ ├── rank\_plot.pdf

│ └── rank\_plot.png

├── 07.Ecological\_diff

│ ├── alldiff.csv

│ ├── cor

│ │ ├── all\_cor.csv

│ │ ├── C1QA

│ │ │ ├── C1QA\_APOE.cor.pdf

│ │ │ ├── C1QA\_C1QB.cor.pdf

│ │ │ ├── C1QA\_C1QC.cor.pdf

│ │ │ ├── C1QA\_FPR2.cor.pdf

│ │ │ ├── C1QA\_HLA-DQA1.cor.pdf

│ │ │ └── C1QA\_MCEMP1.cor.pdf

│ │ ├── C1QB

│ │ │ ├── C1QB\_APOE.cor.pdf

│ │ │ ├── C1QB\_HLA-DQA1.cor.pdf

│ │ │ └── C1QB\_MCEMP1.cor.pdf

│ │ ├── C1QC

│ │ │ ├── C1QC\_C1QB.cor.pdf

│ │ │ └── C1QC\_MCEMP1.cor.pdf

│ │ ├── CD74

│ │ │ ├── CD74\_HLA-DPB1.cor.pdf

│ │ │ ├── CD74\_HLA-DQA1.cor.pdf

│ │ │ ├── CD74\_HLA-DQB1.cor.pdf

│ │ │ └── CD74\_HLA-DRB1.cor.pdf

│ │ ├── CD81

│ │ │ └── CD81\_APOE.cor.pdf

│ │ ├── GPNMB

│ │ │ ├── GPNMB\_APOE.cor.pdf

│ │ │ └── GPNMB\_GRN.cor.pdf

│ │ ├── GRN

│ │ │ └── GRN\_CTSZ.cor.pdf

│ │ ├── HLA-A

│ │ │ └── HLA-A\_B2M.cor.pdf

│ │ ├── HLA-DPA1

│ │ │ └── HLA-DPA1\_HLA-DPB1.cor.pdf

│ │ ├── HLA-DQA1

│ │ │ ├── HLA-DQA1\_HLA-DPB1.cor.pdf

│ │ │ └── HLA-DQA1\_HLA-DQB1.cor.pdf

│ │ ├── HLA-DRB1

│ │ │ ├── HLA-DRB1\_HLA-DPB1.cor.pdf

│ │ │ ├── HLA-DRB1\_HLA-DQA1.cor.pdf

│ │ │ └── HLA-DRB1\_HLA-DQB1.cor.pdf

│ │ ├── HLA-DRB5

│ │ │ ├── HLA-DRB5\_HLA-DQA1.cor.pdf

│ │ │ ├── HLA-DRB5\_HLA-DQB1.cor.pdf

│ │ │ └── HLA-DRB5\_HLA-DRB1.cor.pdf

│ │ └── MCEMP1

│ │ └── MCEMP1\_FPR2.cor.pdf

│ ├── cor\_input.RData

│ ├── eco.RData

│ ├── Ecotyper\_diff.jpg

│ ├── Ecotyper\_diff.pdf

│ ├── Ecotyper\_geni\_nfo.txt

│ ├── Ecotyper.txt

│ ├── gsva

│ │ ├── easy\_input2\_for39bar.csv

│ │ ├── gsva.jpg

│ │ ├── gsva\_limma.csv

│ │ ├── gsva\_output.csv

│ │ └── gsva.pdf

│ └── result\_diff.csv

├── 08.WGCNA

│ ├── CIBERSOFT.txt

│ ├── clean.R

│ ├── data.txt

│ ├── gencode.v22.annotation.gene.probeMap

│ ├── geneLength.txt

│ ├── M2.pdf

│ ├── meta.csv

│ ├── result

│ │ ├── 1\_sample\_cluster.pdf

│ │ ├── 2\_sample\_heatmap.pdf

│ │ ├── 3\_scale\_independence.pdf

│ │ ├── 4\_gene\_clustering.pdf

│ │ ├── 5\_Dynamic\_Tree.pdf

│ │ ├── 6\_Clustering\_module.pdf

(base) -bash-4.2$ tree BJTC-277/

BJTC-277/

├── 00.data

│ ├── 01.single\_data

│ │ ├── scRNA\_bacth\_celltype.Rdata

│ │ ├── scRNA\_bath.Rdata

│ │ ├── scRNA\_harmony\_celltype.Rdata

│ │ ├── scRNA\_harmony.Rdata

│ │ ├── scRNAlist.Rdata

│ │ ├── scRNA\_macrophageM2.Rdata

│ │ ├── scRNA\_macrophage.Rdata

│ │ ├── scRNA\_orig.Rdata

│ │ ├── scRNA\_qc.Rdata

│ │ ├── seurat\_EcoTyper.h5ad

│ │ ├── seurat\_EcoTyper.h5seurat

│ │ ├── seurat\_macrophage.h5ad

│ │ ├── seurat\_macrophage.h5seurat

│ │ ├── seurat\_raw.h5ad

│ │ └── seurat\_raw.h5seurat

│ ├── metadata.txt

│ ├── seurat\_10x

│ │ ├── barcodes.tsv

│ │ ├── genes.tsv

│ │ └── matrix.mtx

│ └── single\_data

│ ├── GSM5004180\_PT1

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004181\_PT2

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004182\_PT3

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004184\_LN1

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004185\_LN2

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004187\_P1

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004188\_Li1

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ └── GSM5004189\_Li2

│ ├── barcodes.tsv.gz

│ ├── features.tsv.gz

│ └── matrix.mtx.gz

├── 01.QC

│ ├── 1.QC

│ │ ├── Elbow.jpg

│ │ └── Elbow.pdf

│ ├── 2.UMAP

│ │ ├── all1.jpg

│ │ ├── all1.pdf

│ │ ├── cluster.jpg

│ │ ├── cluster.pdf

│ │ ├── Patients.jpg

│ │ ├── Patients.pdf

│ │ ├── Type.jpg

│ │ └── Type.pdf

│ ├── 3.TSNE

│ │ ├── all2.jpg

│ │ ├── all2.pdf

│ │ ├── cluster.jpg

│ │ ├── cluster.pdf

│ │ ├── Patients.jpg

│ │ ├── Patients.pdf

│ │ ├── Type.jpg

│ │ └── Type.pdf

│ ├── all1.RData

│ ├── all2.RData

│ ├── vlnplot\_after\_qc.jpg

│ ├── vlnplot\_after\_qc.pdf

│ ├── vlnplot\_before\_qc.jpg

│ └── vlnplot\_before\_qc.pdf

├── 02.harmony

│ ├── 1.UMAP

│ │ ├── all1.jpg

│ │ ├── all1.pdf

│ │ ├── all.jpg

│ │ ├── all.pdf

│ │ ├── cluster.jpg

│ │ ├── cluster.pdf

│ │ ├── Patients.jpg

│ │ ├── Patients.pdf

│ │ ├── Type.jpg

│ │ └── Type.pdf

│ └── 2.TSNE

│ ├── all2.jpg

│ ├── all2.pdf

│ ├── all.jpg

│ ├── all.pdf

│ ├── cluster.jpg

│ ├── cluster.pdf

│ ├── Patients.jpg

│ ├── Patients.pdf

│ ├── Type.jpg

│ └── Type.pdf

├── 03.celltype

│ ├── 01.harmony

│ │ ├── 02.UMAP

│ │ │ ├── cell\_DotPlot.jpg

│ │ │ ├── cell\_DotPlot.pdf

│ │ │ ├── cell\_type.jpg

│ │ │ └── cell\_type.pdf

│ │ ├── cell\_barplot1.jpg

│ │ ├── cell\_barplot1.pdf

│ │ ├── cell\_barplot2.jpg

│ │ ├── cell\_barplot2.pdf

│ │ ├── cell\_barplot.jpg

│ │ ├── cell\_barplot.pdf

│ │ ├── cell\_DotPlot.jpg

│ │ ├── cell\_DotPlot.pdf

│ │ ├── cell\_Feature.jpg

│ │ ├── cell\_Feature.pdf

│ │ ├── cell\_type.jpg

│ │ ├── cell\_type.pdf

│ │ ├── celltype.R

│ │ ├── diff\_genes\_mast.csv

│ │ ├── feature

│ │ │ ├── CD2.jpg

│ │ │ ├── CD2.pdf

│ │ │ ├── CD3D.jpg

│ │ │ ├── CD3D.pdf

│ │ │ ├── CD3E.jpg

│ │ │ ├── CD3E.pdf

│ │ │ ├── CD68.jpg

│ │ │ ├── CD68.pdf

│ │ │ ├── CD79A.jpg

│ │ │ ├── CD79A.pdf

│ │ │ ├── CLDN4.jpg

│ │ │ ├── CLDN4.pdf

│ │ │ ├── CSF1R.jpg

│ │ │ ├── CSF1R.pdf

│ │ │ ├── CSF3R.jpg

│ │ │ ├── CSF3R.pdf

│ │ │ ├── EPCAM.jpg

│ │ │ ├── EPCAM.pdf

│ │ │ ├── FN1.jpg

│ │ │ ├── FN1.pdf

│ │ │ ├── GNLY.jpg

│ │ │ ├── GNLY.pdf

│ │ │ ├── IGHG1.jpg

│ │ │ ├── IGHG1.pdf

│ │ │ ├── KDR.jpg

│ │ │ ├── KDR.pdf

│ │ │ ├── KIT.jpg

│ │ │ ├── KIT.pdf

│ │ │ ├── KLRD1.jpg

│ │ │ ├── KLRD1.pdf

│ │ │ ├── KLRF1.jpg

│ │ │ ├── KLRF1.pdf

│ │ │ ├── KRT19.jpg

│ │ │ ├── KRT19.pdf

│ │ │ ├── MKI67.jpg

│ │ │ ├── MKI67.pdf

│ │ │ ├── MMP2.jpg

│ │ │ ├── MMP2.pdf

│ │ │ ├── MS4A1.jpg

│ │ │ ├── MS4A1.pdf

│ │ │ ├── MZB1.jpg

│ │ │ ├── MZB1.pdf

│ │ │ ├── PCNA.jpg

│ │ │ ├── PCNA.pdf

│ │ │ ├── PDGFRA.jpg

│ │ │ ├── PDGFRA.pdf

│ │ │ ├── PECAM1.jpg

│ │ │ ├── PECAM1.pdf

│ │ │ ├── STMN1.jpg

│ │ │ ├── STMN1.pdf

│ │ │ ├── VWF.jpg

│ │ │ └── VWF.pdf

│ │ ├── top10\_diff\_genes\_mast.csv

│ │ └── type.csv

│ └── 02.bath

│ ├── cell\_barplot1.jpg

│ ├── cell\_barplot1.pdf

│ ├── cell\_barplot2.jpg

│ ├── cell\_barplot2.pdf

│ ├── cell\_barplot3.pdf

│ ├── cell\_barplot4.jpg

│ ├── cell\_barplot4.pdf

│ ├── cell\_barplot.jpg

│ ├── cell\_barplot.pdf

│ ├── cell\_DotPlot.jpg

│ ├── cell\_DotPlot.pdf

│ ├── cell\_Feature.jpg

│ ├── cell\_Feature.pdf

│ ├── cell\_type.jpg

│ ├── cell\_type.pdf

│ ├── celltype.R

│ ├── diff\_genes\_mast.csv

│ ├── feature

│ │ ├── CD163.jpg

│ │ ├── CD163.pdf

│ │ ├── CD2.jpg

│ │ ├── CD2.pdf

│ │ ├── CD3D.jpg

│ │ ├── CD3D.pdf

│ │ ├── CD68.jpg

│ │ ├── CD68.pdf

│ │ ├── CD79A.jpg

│ │ ├── CD79A.pdf

│ │ ├── CSF3R.jpg

│ │ ├── CSF3R.pdf

│ │ ├── EPCAM.jpg

│ │ ├── EPCAM.pdf

│ │ ├── FCGR3A.jpg

│ │ ├── FCGR3A.pdf

│ │ ├── FN1.jpg

│ │ ├── FN1.pdf

│ │ ├── KIT.jpg

│ │ ├── KIT.pdf

│ │ ├── KLRD1.jpg

│ │ ├── KLRD1.pdf

│ │ ├── LILRA4.jpg

│ │ ├── LILRA4.pdf

│ │ ├── MKI67.jpg

│ │ ├── MKI67.pdf

│ │ ├── MS4A1.jpg

│ │ ├── MS4A1.pdf

│ │ ├── MZB1.jpg

│ │ ├── MZB1.pdf

│ │ ├── PECAM1.jpg

│ │ └── PECAM1.pdf

│ ├── top10\_diff\_genes\_mast.csv

│ └── type.csv

├── 04.EMT\_cor

│ ├── all1.jpg

│ ├── all1.pdf

│ ├── all.jpg

│ ├── all.pdf

│ ├── angiogenesis.txt

│ ├── cor1.jpg

│ ├── cor1.pdf

│ ├── cor2.jpg

│ ├── cor2.pdf

│ ├── cor3.jpg

│ ├── cor3.pdf

│ └── gsva\_result.csv

├── 05.macrophage

│ ├── barplot.pdf

│ ├── cell\_DotPlot.jpg

│ ├── cell\_DotPlot.pdf

│ ├── cell\_Feature.jpg

│ ├── cell\_Feature.pdf

│ ├── cell\_type.jpg

│ ├── cell\_type.pdf

│ ├── cluster.pdf

│ ├── diff\_genes\_mast.csv

│ ├── feature

│ │ ├── CD163.jpg

│ │ ├── CD163.pdf

│ │ ├── CD1C.jpg

│ │ ├── CD1C.pdf

│ │ ├── CD68.jpg

│ │ ├── CD68.pdf

│ │ ├── LAMP3.jpg

│ │ ├── LAMP3.pdf

│ │ ├── MRC1.jpg

│ │ ├── MRC1.pdf

│ │ ├── S100A8.jpg

│ │ ├── S100A8.pdf

│ │ ├── S100A9.jpg

│ │ ├── S100A9.pdf

│ │ ├── TNF.jpg

│ │ └── TNF.pdf

│ ├── figures

│ │ ├── barplot.pdf

│ │ ├── cluster.pdf

│ │ ├── tsnecelltype.pdf

│ │ └── umapcelltype\_umap.pdf

│ ├── top10\_diff\_genes\_mast.csv

│ ├── tsnecelltype.pdf

│ ├── type.csv

│ └── umapcelltype\_umap.pdf

├── 06.EcoTyper

│ └── DiscoveryOutput3\_scRNA

│ ├── Ecotypes

│ │ ├── ecotype\_abundance.txt

│ │ ├── ecotype\_assignment.txt

│ │ ├── ecotypes.txt

│ │ ├── heatmap\_assigned\_samples\_viridis.pdf

│ │ ├── heatmap\_assigned\_samples\_viridis.png

│ │ ├── jaccard\_matrix.pdf

│ │ ├── jaccard\_matrix.png

│ │ ├── nclusters\_jaccard.pdf

│ │ └── nclusters\_jaccard.png

│ ├── Macrophage.M1

│ │ ├── gene\_info.txt

│ │ ├── heatmap\_data.txt

│ │ ├── heatmap\_top\_ann.txt

│ │ ├── state\_abundances.txt

│ │ ├── state\_assignment\_heatmap.pdf

│ │ ├── state\_assignment\_heatmap.png

│ │ └── state\_assignment.txt

│ ├── Macrophage.M2

│ │ ├── gene\_info.txt

│ │ ├── heatmap\_data.txt

│ │ ├── heatmap\_top\_ann.txt

│ │ ├── state\_abundances.txt

│ │ ├── state\_assignment\_heatmap.pdf

│ │ ├── state\_assignment\_heatmap.png

│ │ └── state\_assignment.txt

│ ├── rank\_data.txt

│ ├── rank\_plot.pdf

│ └── rank\_plot.png

├── 07.Ecological\_diff

│ ├── alldiff.csv

│ ├── cor\_input.RData

│ ├── eco.RData

│ ├── Ecotyper\_diff.jpg

│ ├── Ecotyper\_diff.pdf

│ ├── Ecotyper\_geni\_nfo.txt

│ ├── Ecotyper.txt

│ ├── gsva

│ │ ├── easy\_input2\_for39bar.csv

│ │ ├── gsva.jpg

│ │ ├── gsva\_limma.csv

│ │ ├── gsva\_output.csv

│ │ └── gsva.pdf

│ └── result\_diff.csv

├── 08.WGCNA

│ ├── CIBERSOFT.txt

│ ├── clean.R

│ ├── data.txt

│ ├── gencode.v22.annotation.gene.probeMap

│ ├── geneLength.txt

│ ├── M2.pdf

│ ├── meta.csv

│ ├── result

│ │ ├── 1\_sample\_cluster.pdf

│ │ ├── 2\_sample\_heatmap.pdf

│ │ ├── 3\_scale\_independence.pdf

│ │ ├── 4\_gene\_clustering.pdf

│ │ ├── 5\_Dynamic\_Tree.pdf

│ │ ├── 6\_Clustering\_module.pdf

(base) -bash-4.2$ tree BJTC-277/

BJTC-277/

├── 00.data

│ ├── 01.single\_data

│ │ ├── scRNA\_bacth\_celltype.Rdata

│ │ ├── scRNA\_bath.Rdata

│ │ ├── scRNA\_harmony\_celltype.Rdata

│ │ ├── scRNA\_harmony.Rdata

│ │ ├── scRNAlist.Rdata

│ │ ├── scRNA\_macrophageM2.Rdata

│ │ ├── scRNA\_macrophage.Rdata

│ │ ├── scRNA\_orig.Rdata

│ │ ├── scRNA\_qc.Rdata

│ │ ├── seurat\_EcoTyper.h5ad

│ │ ├── seurat\_EcoTyper.h5seurat

│ │ ├── seurat\_macrophage.h5ad

│ │ ├── seurat\_macrophage.h5seurat

│ │ ├── seurat\_raw.h5ad

│ │ └── seurat\_raw.h5seurat

│ ├── metadata.txt

│ ├── seurat\_10x

│ │ ├── barcodes.tsv

│ │ ├── genes.tsv

│ │ └── matrix.mtx

│ └── single\_data

│ ├── GSM5004180\_PT1

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004181\_PT2

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004182\_PT3

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004184\_LN1

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004185\_LN2

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004187\_P1

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ ├── GSM5004188\_Li1

│ │ ├── barcodes.tsv.gz

│ │ ├── features.tsv.gz

│ │ └── matrix.mtx.gz

│ └── GSM5004189\_Li2

│ ├── barcodes.tsv.gz

│ ├── features.tsv.gz

│ └── matrix.mtx.gz

├── 01.QC

│ ├── 1.QC

│ │ ├── Elbow.jpg

│ │ └── Elbow.pdf

│ ├── 3.TSNE

│ │ ├── all2.jpg

│ │ ├── all2.pdf

│ │ ├── cluster.jpg

│ │ ├── cluster.pdf

│ │ ├── Patients.jpg

│ │ ├── Patients.pdf

│ │ ├── Type.jpg

│ │ └── Type.pdf

│ ├── all1.RData

│ ├── all2.RData

│ ├── vlnplot\_after\_qc.jpg

│ ├── vlnplot\_after\_qc.pdf

│ ├── vlnplot\_before\_qc.jpg

│ └── vlnplot\_before\_qc.pdf

├── 02.harmony

│ └── 2.TSNE

│ ├── all2.jpg

│ ├── all2.pdf

│ ├── all.jpg

│ ├── all.pdf

│ ├── cluster.jpg

│ ├── cluster.pdf

│ ├── Patients.jpg

│ ├── Patients.pdf

│ ├── Type.jpg

│ └── Type.pdf

├── 03.celltype

│ └── 02.bath

│ ├── cell\_barplot1.jpg

│ ├── cell\_barplot1.pdf

│ ├── cell\_barplot2.jpg

│ ├── cell\_barplot2.pdf

│ ├── cell\_barplot3.pdf

│ ├── cell\_barplot4.jpg

│ ├── cell\_barplot4.pdf

│ ├── cell\_barplot.jpg

│ ├── cell\_barplot.pdf

│ ├── cell\_DotPlot.jpg

│ ├── cell\_DotPlot.pdf

│ ├── cell\_Feature.jpg

│ ├── cell\_Feature.pdf

│ ├── cell\_type.jpg

│ ├── cell\_type.pdf

│ ├── celltype.R

│ ├── diff\_genes\_mast.csv

│ ├── feature

│ │ ├── CD163.jpg

│ │ ├── CD163.pdf

│ │ ├── CD2.jpg

│ │ ├── CD2.pdf

│ │ ├── CD3D.jpg

│ │ ├── CD3D.pdf

│ │ ├── CD68.jpg

│ │ ├── CD68.pdf

│ │ ├── CD79A.jpg

│ │ ├── CD79A.pdf

│ │ ├── CSF3R.jpg

│ │ ├── CSF3R.pdf

│ │ ├── EPCAM.jpg

│ │ ├── EPCAM.pdf

│ │ ├── FCGR3A.jpg

│ │ ├── FCGR3A.pdf

│ │ ├── FN1.jpg

│ │ ├── FN1.pdf

│ │ ├── KIT.jpg

│ │ ├── KIT.pdf

│ │ ├── KLRD1.jpg

│ │ ├── KLRD1.pdf

│ │ ├── LILRA4.jpg

│ │ ├── LILRA4.pdf

│ │ ├── MKI67.jpg

│ │ ├── MKI67.pdf

│ │ ├── MS4A1.jpg

│ │ ├── MS4A1.pdf

│ │ ├── MZB1.jpg

│ │ ├── MZB1.pdf

│ │ ├── PECAM1.jpg

│ │ └── PECAM1.pdf

│ ├── top10\_diff\_genes\_mast.csv

│ └── type.csv

├── 04.EMT\_cor

│ ├── all1.jpg

│ ├── all1.pdf

│ ├── all.jpg

│ ├── all.pdf

│ ├── angiogenesis.txt

│ ├── cor1.jpg

│ ├── cor1.pdf

│ ├── cor2.jpg

│ ├── cor2.pdf

│ ├── cor3.jpg

│ ├── cor3.pdf

│ └── gsva\_result.csv

├── 05.macrophage

│ ├── barplot.pdf

│ ├── cell\_DotPlot.jpg

│ ├── cell\_DotPlot.pdf

│ ├── cell\_Feature.jpg

│ ├── cell\_Feature.pdf

│ ├── cell\_type.jpg

│ ├── cell\_type.pdf

│ ├── cluster.pdf

│ ├── diff\_genes\_mast.csv

│ ├── feature

│ │ ├── CD163.jpg

│ │ ├── CD163.pdf

│ │ ├── CD1C.jpg

│ │ ├── CD1C.pdf

│ │ ├── CD68.jpg

│ │ ├── CD68.pdf

│ │ ├── LAMP3.jpg

│ │ ├── LAMP3.pdf

│ │ ├── MRC1.jpg

│ │ ├── MRC1.pdf

│ │ ├── S100A8.jpg

│ │ ├── S100A8.pdf

│ │ ├── S100A9.jpg

│ │ ├── S100A9.pdf

│ │ ├── TNF.jpg

│ │ └── TNF.pdf

│ ├── figures

│ │ ├── barplot.pdf

│ │ ├── cluster.pdf

│ │ ├── tsnecelltype.pdf

│ │ └── umapcelltype\_umap.pdf

│ ├── top10\_diff\_genes\_mast.csv

│ ├── tsnecelltype.pdf

│ ├── type.csv

│ └── umapcelltype\_umap.pdf

├── 06.EcoTyper

│ └── DiscoveryOutput3\_scRNA

│ ├── Ecotypes

│ │ ├── ecotype\_abundance.txt

│ │ ├── ecotype\_assignment.txt

│ │ ├── ecotypes.txt

│ │ ├── heatmap\_assigned\_samples\_viridis.pdf

│ │ ├── heatmap\_assigned\_samples\_viridis.png

│ │ ├── jaccard\_matrix.pdf

│ │ ├── jaccard\_matrix.png

│ │ ├── nclusters\_jaccard.pdf

│ │ └── nclusters\_jaccard.png

│ ├── Macrophage.M1

│ │ ├── gene\_info.txt

│ │ ├── heatmap\_data.txt

│ │ ├── heatmap\_top\_ann.txt

│ │ ├── state\_abundances.txt

│ │ ├── state\_assignment\_heatmap.pdf

│ │ ├── state\_assignment\_heatmap.png

│ │ └── state\_assignment.txt

│ ├── Macrophage.M2

│ │ ├── gene\_info.txt

│ │ ├── heatmap\_data.txt

│ │ ├── heatmap\_top\_ann.txt

│ │ ├── state\_abundances.txt

│ │ ├── state\_assignment\_heatmap.pdf

│ │ ├── state\_assignment\_heatmap.png

│ │ └── state\_assignment.txt

│ ├── rank\_data.txt

│ ├── rank\_plot.pdf

│ └── rank\_plot.png

├── 07.Ecological\_diff

│ ├── alldiff.csv

│ ├── cor\_input.RData

│ ├── eco.RData

│ ├── Ecotyper\_diff.jpg

│ ├── Ecotyper\_diff.pdf

│ ├── Ecotyper\_geni\_nfo.txt

│ ├── Ecotyper.txt

│ ├── gsva

│ │ ├── easy\_input2\_for39bar.csv

│ │ ├── gsva.jpg

│ │ ├── gsva\_limma.csv

│ │ ├── gsva\_output.csv

│ │ └── gsva.pdf

│ └── result\_diff.csv

├── 08.WGCNA

│ ├── CIBERSOFT.txt

│ ├── clean.R

│ ├── data.txt

│ ├── gencode.v22.annotation.gene.probeMap

│ ├── geneLength.txt

│ ├── M2.pdf

│ ├── meta.csv

│ ├── result

│ │ ├── 1\_sample\_cluster.pdf

│ │ ├── 2\_sample\_heatmap.pdf

│ │ ├── 3\_scale\_independence.pdf

│ │ ├── 4\_gene\_clustering.pdf

│ │ ├── 5\_Dynamic\_Tree.pdf

│ │ ├── 6\_Clustering\_module.pdf

│ │ ├── 7\_merged\_dynamic.pdf

│ │ ├── 8\_Module\_trait.pdf

│ │ ├── 9\_High\ TAM\_darkred.pdf

│ │ ├── 9\_High\ TAM\_green.pdf

│ │ ├── 9\_High\ TAM\_grey60.pdf

│ │ ├── 9\_High\ TAM\_lightyellow.pdf

│ │ ├── 9\_High\ TAM\_salmon.pdf

│ │ ├── 9\_Low\ TAM\_darkred.pdf

│ │ ├── 9\_Low\ TAM\_green.pdf

│ │ ├── 9\_Low\ TAM\_grey60.pdf

│ │ ├── 9\_Low\ TAM\_lightyellow.pdf

│ │ ├── 9\_Low\ TAM\_salmon.pdf

│ │ ├── GS\_MM.xls

│ │ ├── TCGA\_darkred.txt

│ │ ├── TCGA\_green.txt

│ │ ├── TCGA\_grey60.txt

│ │ ├── TCGA\_lightyellow.txt

│ │ └── TCGA\_salmon.txt

│ ├── symbol.txt

│ ├── TCGA-STAD.htseq\_counts.tsv

│ ├── TCGA-STAD.htseq\_fpkm.tsv

│ ├── TCGA-STAD.survival.tsv

│ ├── venn

│ │ ├── Ecotyper\_diffgene.txt

│ │ ├── intersect.txt

│ │ ├── TCGA\_darkred.txt

│ │ ├── venn.pdf

│ │ └── venn.R

│ ├── wgcna\_ann.RData

│ └── wgcna\_input.RData

├── 09.ConsensusClusterPlus

│ ├── 2.csv

│ ├── Cluster1.txt

│ ├── consensus001.png

│ ├── consensus002.png

│ ├── consensus003.png

│ ├── consensus004.png

│ ├── consensus005.png

│ ├── consensus006.png

│ ├── consensus007.png

│ ├── consensus008.png

│ ├── consensus009.png

│ ├── consensus010.png

│ └── consensus011.png

├── BJTC-277.Rproj

├── setp1\_meger.R

├── setp2\_Normalization&clustering.R

├── setp3\_harmony.R

├── setp5\_EMTcor.R

├── setp6\_Macrophage.R

├── setp7\_diffM1\_M2.R

├── setp8\_wgcna.R

└── setp9\_ConsensusClusterPlus.R

# References

1. Hong KS, Kim H, Kim SH, Kim M, Yoo J. Calponin 3 Regulates Cell Invasion and Doxorubicin Resistance in Gastric Cancer. Gastroenterol Res Pract. 2019 Feb 18;2019:3024970.
2. Li Y, Wang D, Li Y, Liu X, Chen D, Yuan C, Zhou Y. Clinical Significance of Lymph Node Micrometastasis in pN0 Gastric Cancer Patients. Gastroenterol Res Pract. 2021 Mar 3;2021:6854646.
3. Li J. Gastric Cancer in Young Adults: A Different Clinical Entity from Carcinogenesis to Prognosis. Gastroenterol Res Pract. 2020 Mar 2;2020:9512707.
4. Xia W, Zhang Q, Li Q, Liang X. Relationship between long non-coding RNA TUG1 and prognosis of patients with gastric carcinoma: A protocol for systematic review and meta-analysis. Medicine (Baltimore). 2020 Dec 4;99(49):e23522.
5. Noy R, Pollard JW. Tumor-associated macrophages: from mechanisms to therapy. Immunity. 2014 Jul 17;41(1):49-61.
6. Biswas SK, Allavena P, Mantovani A. Tumor-associated macrophages: functional diversity, clinical significance, and open questions. Semin Immunopathol. 2013 Sep;35(5):585-600.
7. Galdiero MR, Garlanda C, Jaillon S, Marone G, Mantovani A. Tumor associated macrophages and neutrophils in tumor progression. J Cell Physiol. 2013 Jul;228(7):1404-1412.
8. Vinogradov S, Warren G, Wei X. Macrophages associated with tumors as potential targets and therapeutic intermediates. Nanomedicine (Lond). 2014 Apr;9(5):695-707.
9. Hu W, Li X, Zhang C, Yang Y, Jiang J, Wu C. Tumor-associated macrophages in cancers. Clin Transl Oncol. 2016 Mar;18(3):251-258.
10. Chanmee T, Ontong P, Konno K, Itano N. Tumor-associated macrophages as major players in the tumor microenvironment. Cancers (Basel). 2014 Aug 13;6(3):1670-1690.
11. Sawa-Wejksza K, Kandefer-Szerszeń M. Tumor-Associated Macrophages as Target for Antitumor Therapy. Arch Immunol Ther Exp (Warsz). 2018 Apr;66(2):97-111.
12. Mantovani A, Allavena P. The interaction of anticancer therapies with tumor-associated macrophages. J Exp Med. 2015 Apr 6;212(4):435-445.
13. Zhang WJ, Wang XH, Gao ST, Chen C, Xu XY, Sun Q, Zhou ZH, Wu GZ, Yu Q, Xu G, Yao YZ, Guan WX. Tumor-associated macrophages correlate with phenomenon of epithelial-mesenchymal transition and contribute to poor prognosis in triple-negative breast cancer patients. J Surg Res. 2018 Feb;222:93-101.
14. Yuan ZY, Luo RZ, Peng RJ, Wang SS, Xue C. High infiltration of tumor-associated macrophages in triple-negative breast cancer is associated with a higher risk of distant metastasis. Onco Targets Ther. 2014 Aug 21;7:1475-1480.
15. Eum HH, Kwon M, Ryu D, Jo A, Chung W, Kim N, Hong Y, Son DS, Kim ST, Lee J, Lee HO, Park WY. Tumor-promoting macrophages prevail in malignant ascites of advanced gastric cancer. Exp Mol Med. 2020 Dec;52(12):1976-1988.
16. Zhao R, Wan Q, Wang Y, Wu Y, Xiao S, Li Q, Shen X, Zhuang W, Zhou Y, Xia L, Song Y, Chen Y, Yang H, Wu X. M1-like TAMs are required for the efficacy of PD-L1/PD-1 blockades in gastric cancer. Oncoimmunology. 2020 Dec 30;10(1):1862520.
17. Vitale I, Manic G, Coussens LM, Kroemer G, Galluzzi L. Macrophages and Metabolism in the Tumor Microenvironment. Cell Metab. 2019 Jul 2;30(1):36-50.
18. Yamaguchi T, Fushida S, Yamamoto Y, Tsukada T, Kinoshita J, Oyama K, Miyashita T, Tajima H, Ninomiya I, Munesue S, Harashima A, Harada S, Yamamoto H, Ohta T. Tumor-associated macrophages of the M2 phenotype contribute to progression in gastric cancer with peritoneal dissemination. Gastric Cancer. 2016 Oct;19(4):1052-1065.
19. Zhang J, Yan Y, Yang Y, Wang L, Li M, Wang J, Liu X, Duan X, Wang J. High Infiltration of Tumor-Associated Macrophages Influences Poor Prognosis in Human Gastric Cancer Patients, Associates With the Phenomenon of EMT. Medicine (Baltimore). 2016 Feb;95(6):e2636.

