TSCDY201017

systemic sclerosis 单细胞分析

**1.单细胞的质控**

我们从GEO 下载[GSE138669](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE138669)单细胞 RNA 测序数据，选取3个SSc 患者和 3 名健康样本进入后续分析。使用R 包Seurat进行单细胞分析。

如图1所示，我们展示质控6个样本的nFeature\_RNA、nCount\_RNA、percent.mt等参数的小提琴图（不放在文章图中，供分析参考）细胞的数目为：68249

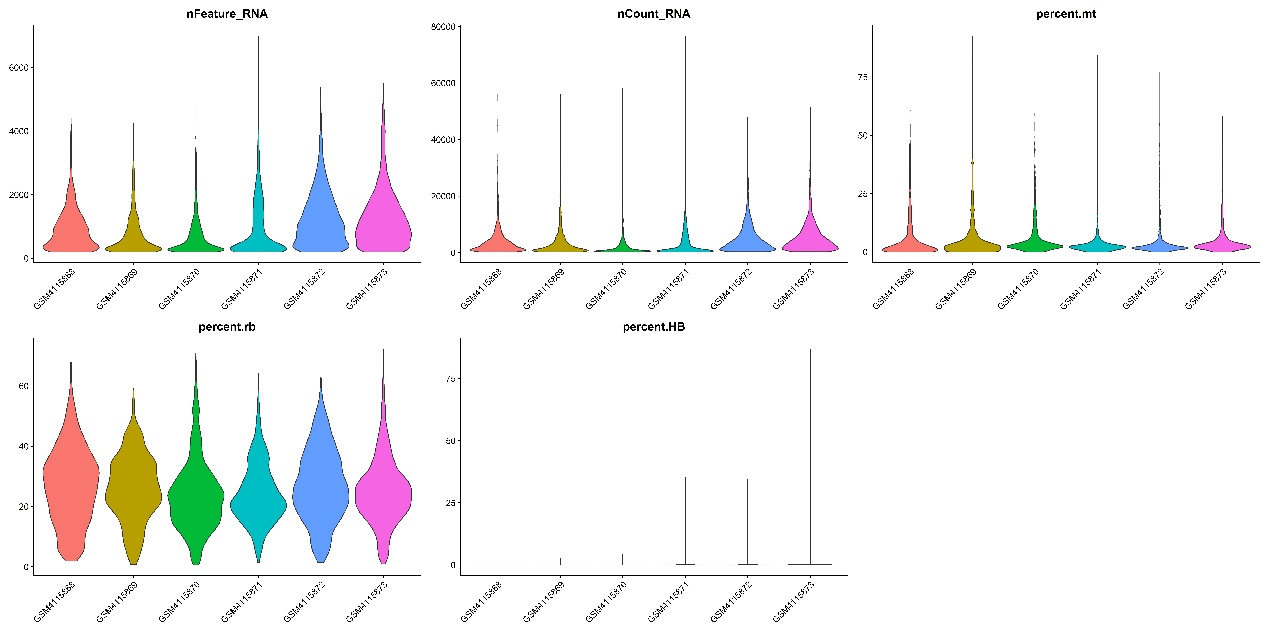


图1 样本的质控参考图

我们根据图1进行参数的设定，选取nFeature\_RNA 大于200，小于2000的细胞。以及线粒体数目小于20%。

如图2所示，我们展示质控后6个样本的UMI 、nCount\_RNA、percent.mt等参数的小提琴图（不放在文章图中，供分析参考）细胞的数目为：35063

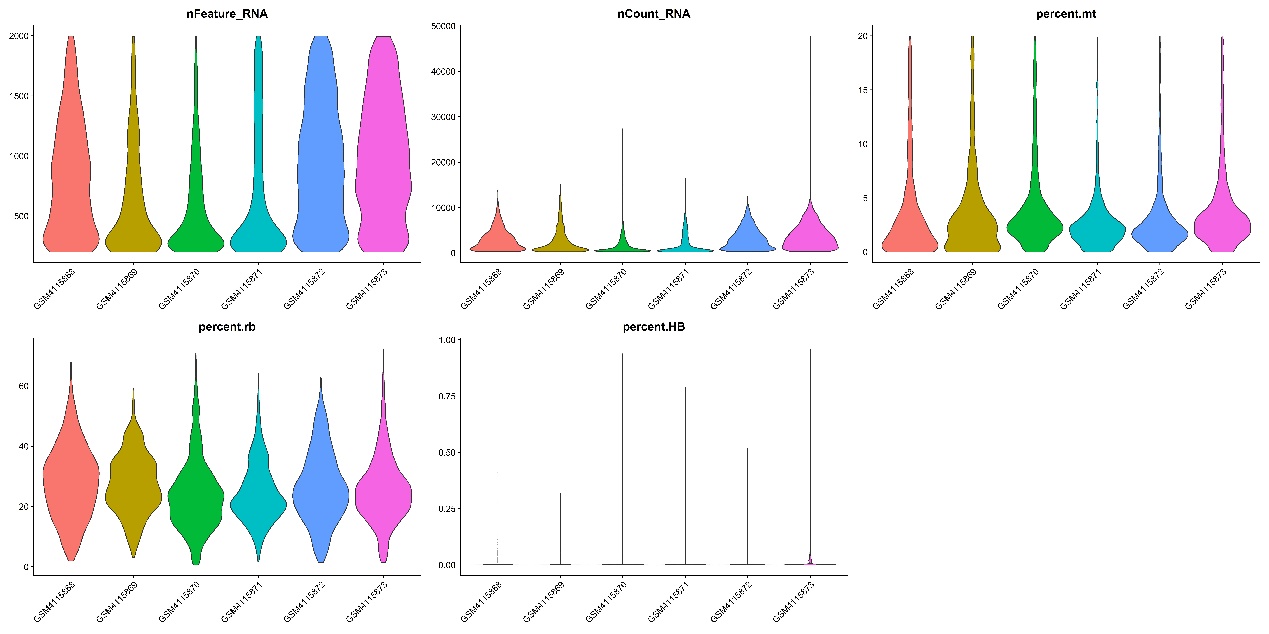
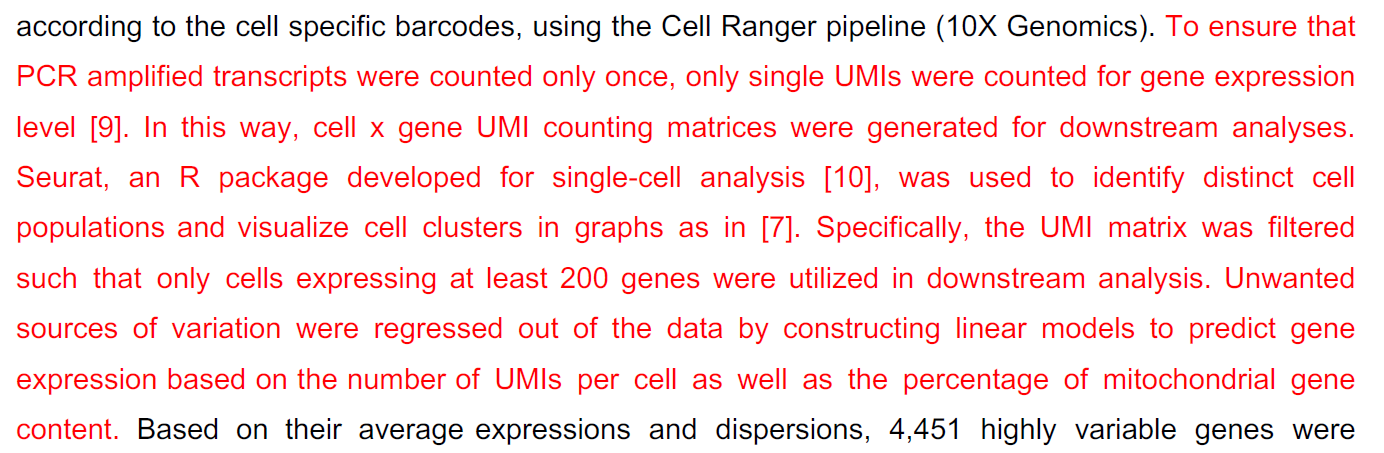


图1 样本的质控后参考图

可参考文献，来源于Single-cell transcriptome analysis identifies skin-specific T-cell responses in systemic sclerosis的附件，我已下载放在参考文件夹中



**2.样本的整合以及去批次效应**

在数据分析的时候，我们的目标是一般都是找到样本之间真实的生物学差异。但是这种真实的生物学因素往往会受到各种因素影响，这些因素之间有些是生物学真实的差异，有些是抽样时的随机波动。有些是系统性因素，统称为批次效应(batch effect)，顾名思义，不同批次带来的效应。如果效应比较小还可以接受，如果批次效应很严重，就可能会和真实的生物学差异相混淆。

我们使用Seurat v4自带的方式进行整合。首先分别对每个样本数据集进行 SCtransform 标准化。随后，运行 PrepSCTIntegration 来选择下游整合的特征，并运行 FindIntegrationAnchors 来识别锚基因。对综合数据进行缩放，并进行主成分分析（Subsequently, PrepSCTIntegration was run to select features for downstream integration and FindIntegrationAnchors to identify anchor genes. The integrated data were scaled, and principal component analysis was performed.）

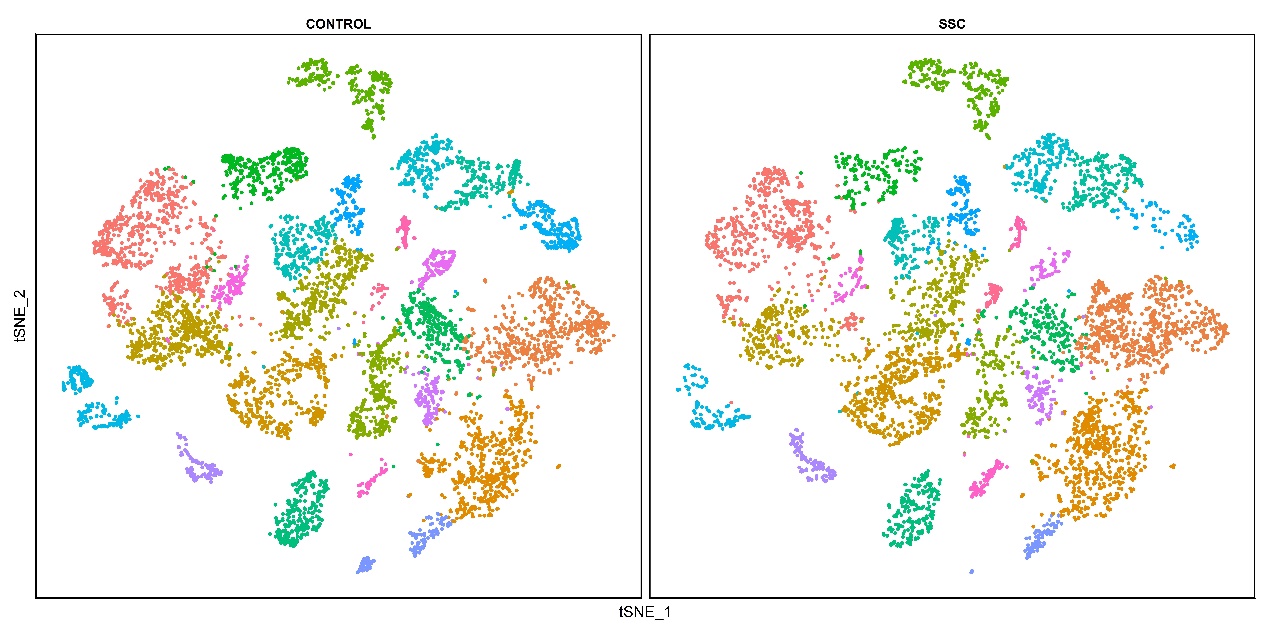


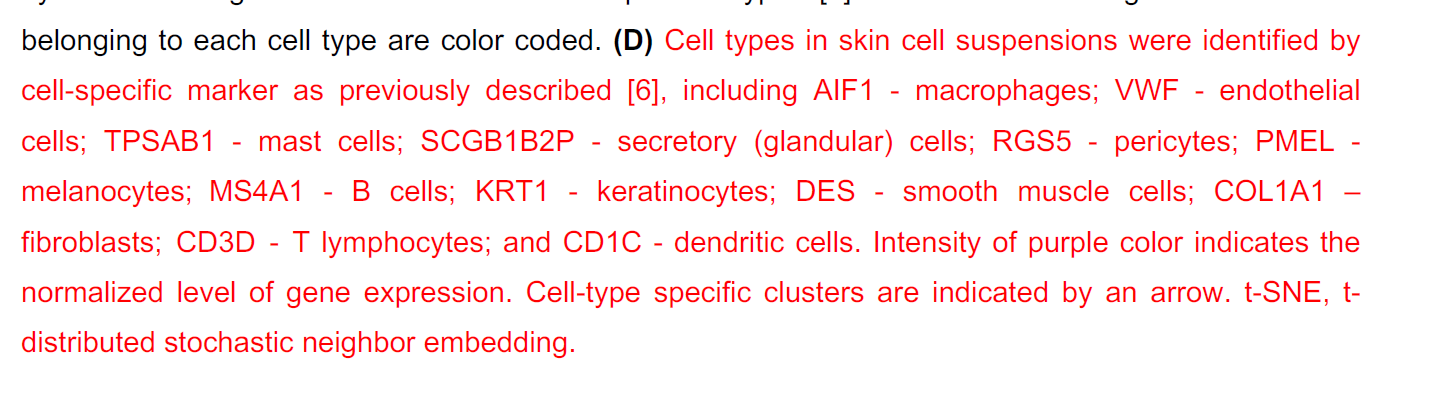
图3可以看到对照组于与ssc组细胞分布很均匀

**3 细胞类型的鉴定**

在对单细胞RNA—seq数据进行质控、降维、聚类以后，接着对数据进行可视化处理。尽管这一步在概念上与降维分析相同，但在仅仅两到三个维度中，可视地分离密切相关的细胞类型，同时确保细胞类型和轨迹之间的相对距离反映了这些细胞类型之间基因表达差异的大小是一项复杂的任务。目前许多线性降维方法，如PCA，无法在二维或三维中生成准确的数据视觉表示。因此，可视化方法倾向于使用非线性方法对数据进行转换。T.SNE算法定义了数据的局部和全局结构之间的界限，既可以使点在局部分散又能在全局聚集，同时照顾近距离和远距离的点。（这里仅是原理说明，写文章在可视化需要提一下使用了tsne的方法进行单细胞分布的2D可视化）

鉴定细胞类型，我们使用原文提供的marker基因进行鉴定，这里有些细胞是我通过cellmarker上找的marker基因，所以需要在文章说明细胞的marker基因

例如：



在附件中提供

图4所示，我们可视化单细胞的tsne图（请不要修改图片，NC原图）

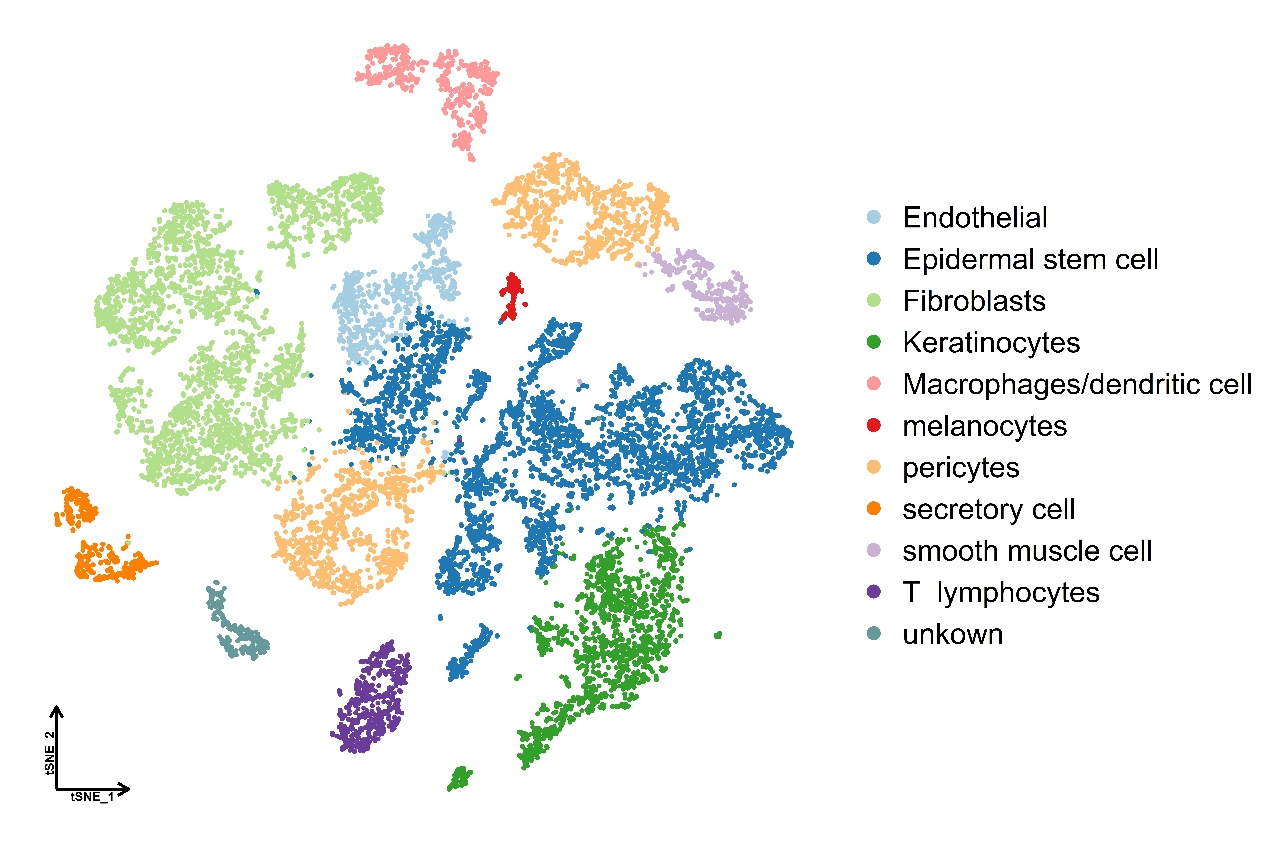
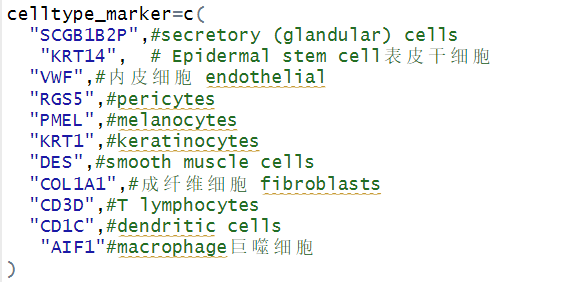


图4 细胞类型的TSNE图

我们观察marker基因在不同簇的分布，确定每个簇的类型。基本上与原文分析细胞类型没有差别（这个图要放）

以及我代码中细胞对应的marker基因不用放参考用的。



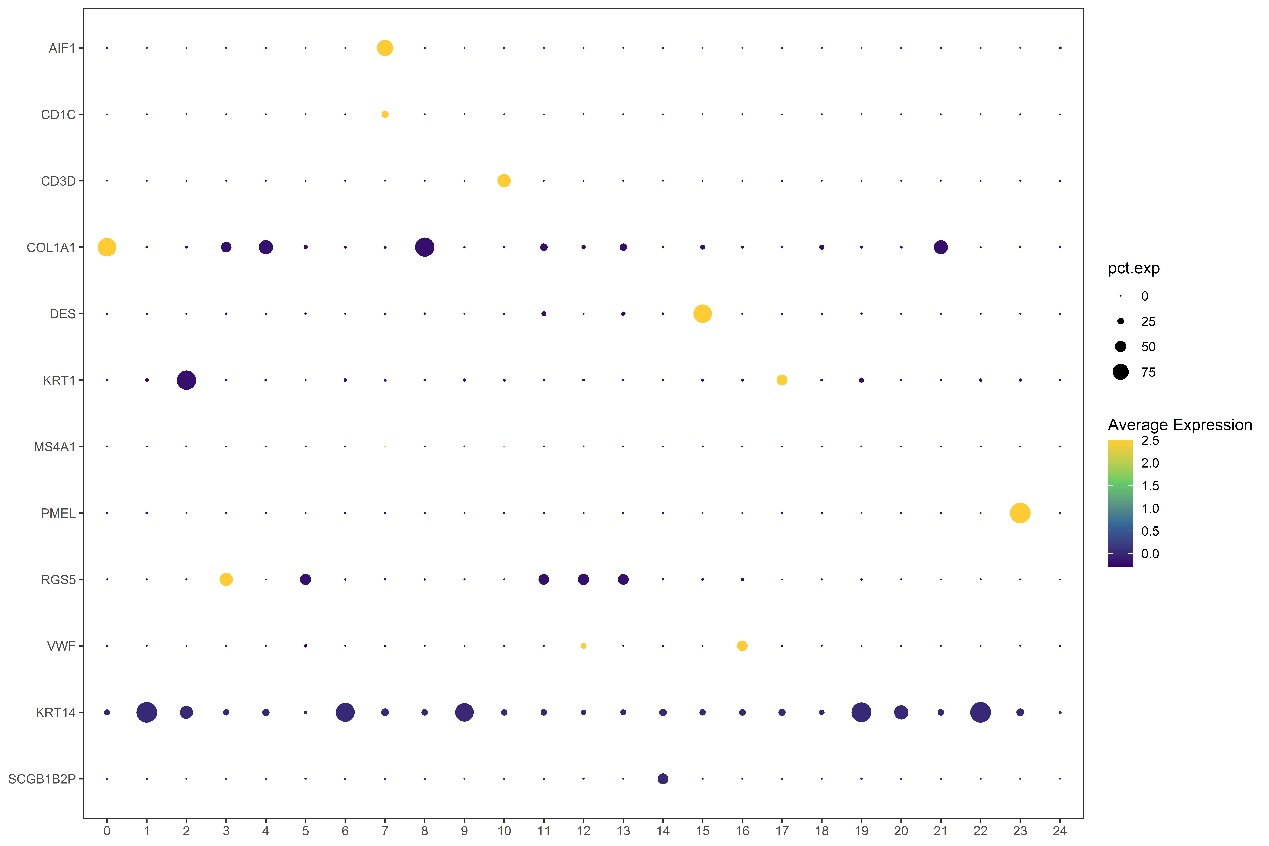


图5 Dotplots图显示细胞的比例独特表达的基因的平均基因表达比例

我们可视化这些marker基因分布

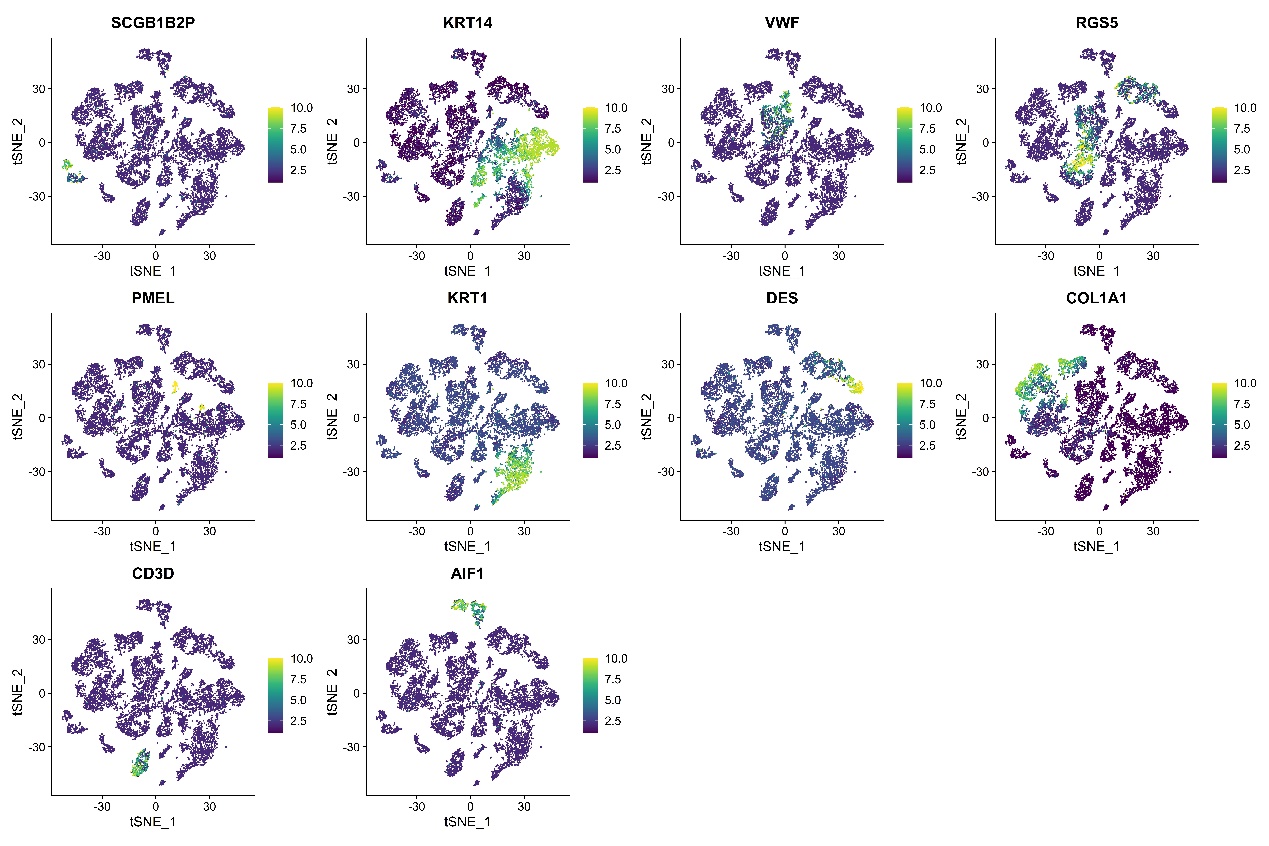


图6 marker基因的样本分布图

接着我们使用GSVA的方法分析对照样本和SSC的成纤维细胞，我们可以发现MECHANOREGULATION AND PATHOLOGY OF YAPTAZ VIA HIPPO AND NONHIPPO MECHANISMS这个通路在图中SSC组中显著富集

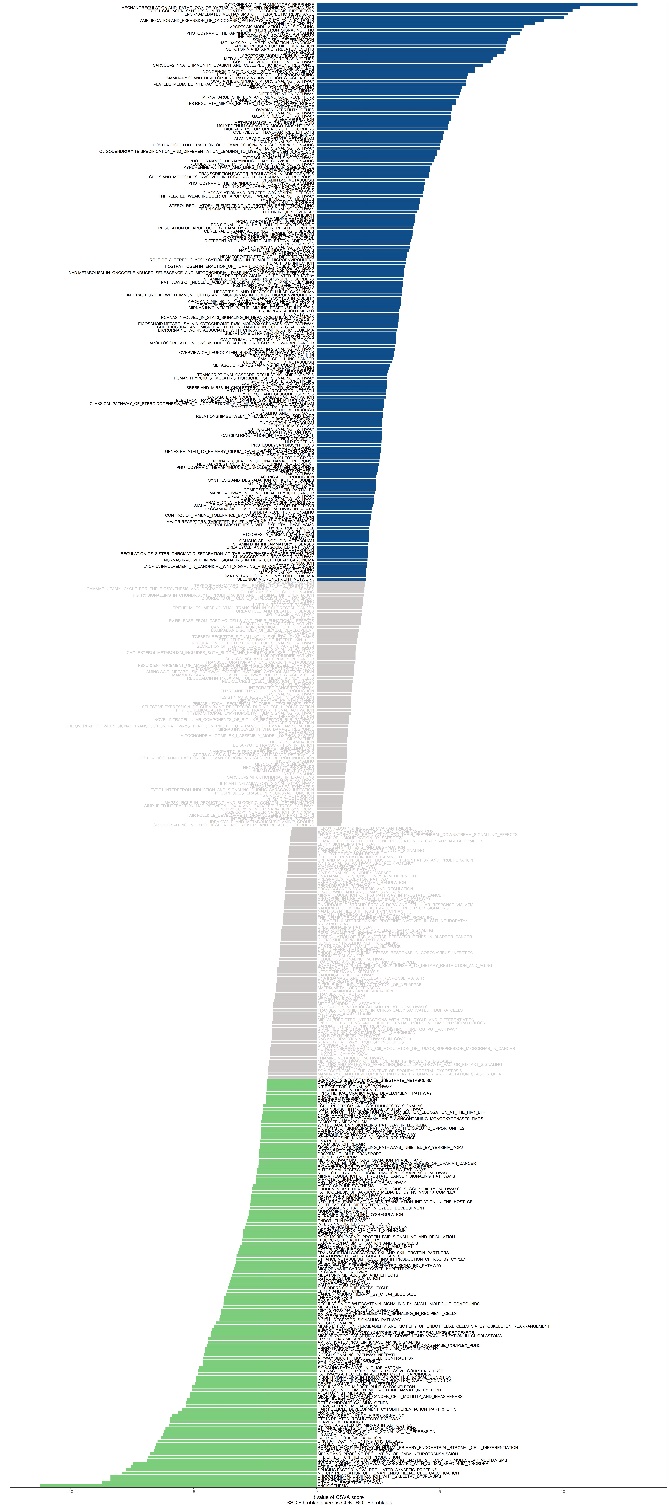
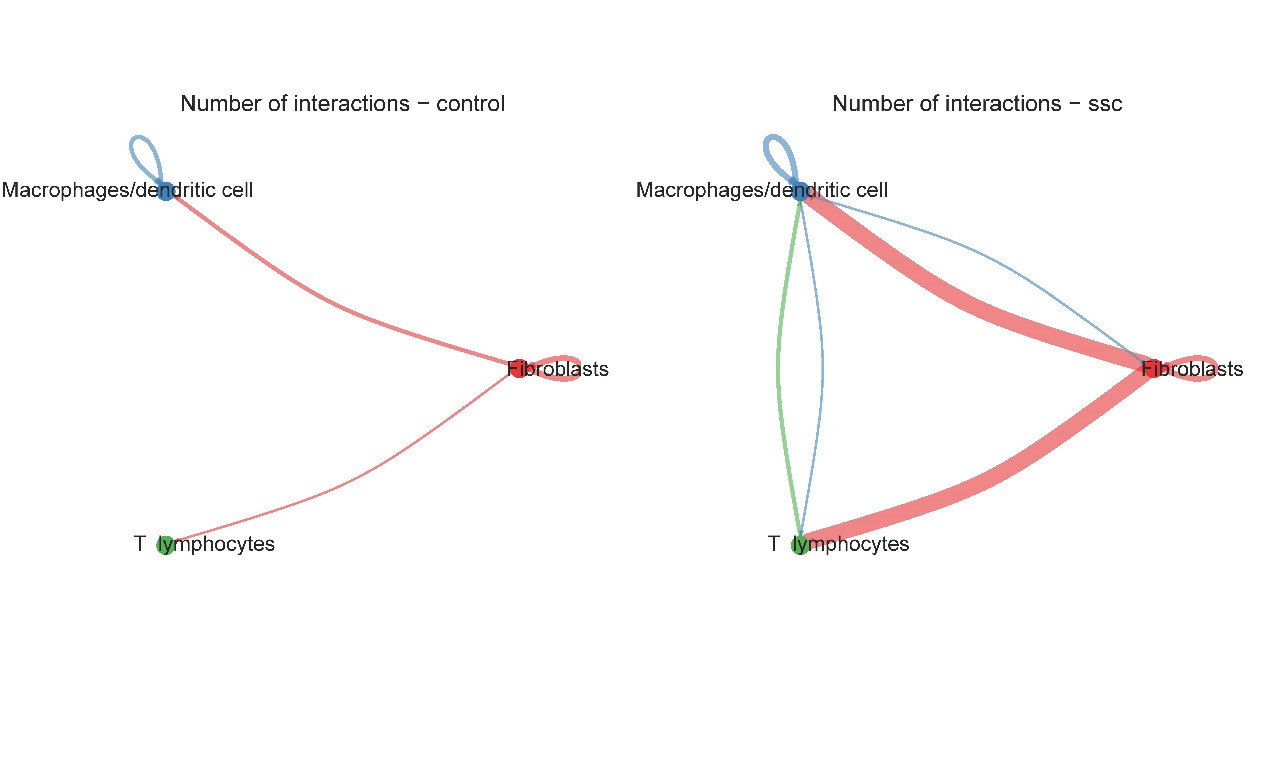


图7 蓝色为ssc显著富集，绿色为对照组富集，我提到的通路在蓝色的第二个

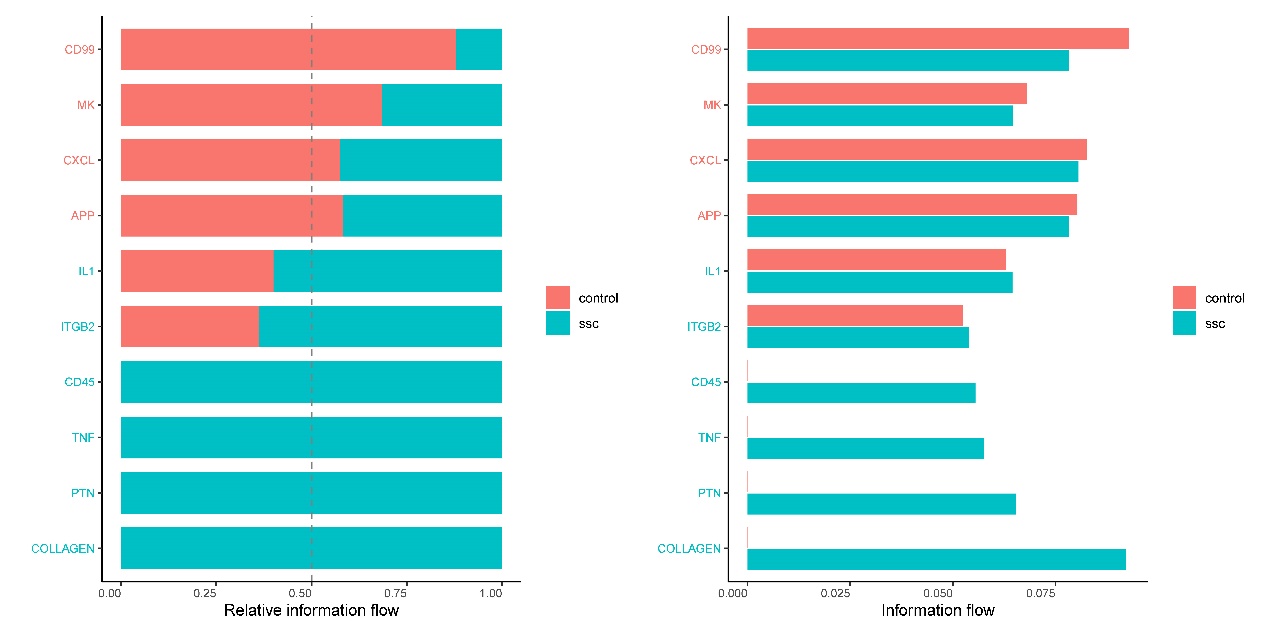
**细胞通讯分析**

CellChat能够从单细胞RNA测序（scRNA-seq）数据中定量推断和分析细胞间通讯网络的工具。CellChat利用网络分析和模式识别方法预测细胞的主要信号输入和输出以及这些细胞和信号如何协调功能。通过多种学习模式和定量对比，CellChat对信号通路进行分类，并在不同的数据集中划分出保守的和特定环境的通路。将CellChat应用于小鼠和人类皮肤数据集显示其有能力提取复杂的信号传导模式，将有助于发现新的细胞间通讯，并建立不同的细胞-细胞通讯图谱。我们使用R软件包‘CellChat’进行分析。我们对成纤维细胞、T lymphocytes细胞、Macrophages/dendritic cell细胞进行相互作用分析，我们可视化了对照组和SSC参与相互作用的配体。（通过检索文献了解配体在细胞通讯之间的作用，进而详细描述）

如下图所示，可以观察到对照组和SSC组中细胞互作数量的对比图，发现SSC组在免疫细胞中信号转导更加频繁，侧面说明巨噬细胞与T淋巴细胞对SSC的发展具有潜在的作用



同时可视化对照组和SSC组保守和特异性信号通路的识别与可视化，发现SSC独有的信号通路（绿色左下角4个通路在SSC组表达）



最后，我们对成纤维细胞、T lymphocytes细胞、Macrophages/dendritic cell细胞进行相互作用分析，我们可视化了对照组和SSC参与相互作用的配体。通过检索文献了解配体在细胞通讯之间的作用，进而详细描述）

