

# 精准单细胞聚类算法研究

法博涛，肖正涛

西安交通大学医学部，分子转化医学研究所，中国西部创新港，7100060



分子转化医学研究所  
Institute of Molecular and Translational Medicine

## 摘要

单细胞RNA测序技术可以检测每个细胞内的基因表达水平，已被广泛应用于研究器官或组织内的细胞类型和细胞状态。高维度的单细胞测序数据通常需要进行降维处理，并映射于二维平面。基于不同细胞之间基因表达谱的相关性，可以将高度相似的细胞进行聚类（分群）。对细胞进行准确聚类有利于进一步定义细胞类型和状态。目前大部分单细胞聚类算法均无法自动决定最优的聚类数目，并且常常会忽视一些数量稀有的关键细胞类型，因此全面、准确地刻画器官或组织内的细胞构成依然存在挑战。为此，我们正在开发一套智能的单细胞分类方法，该方法基于相邻细胞类群之间距离的一阶差分来寻找最佳的细胞分类，可以非常敏感地区分关键细胞类群并进一步将其划分为精细的细胞亚群。此外，该方法的聚类数量不受分辨率等参数的影响，而且对噪音不敏感，聚类结果稳定、鲁棒性强。总之，该方法可用于准确解析器官或组织内的细胞构成及状态。当前的结果表明，在计算机模拟和真实的单细胞转录组数据中，该方法均能准确、全面地鉴定出肿瘤组织中的单细胞类型和其亚类。

## 研究背景

单细胞转录组测序技术（scRNA-seq）被广泛应用于刻画单个细胞状态及类型，是研究器官或组织内的细胞组成及异质性的强有力工具，极大地推动了肿瘤等复杂疾病的研究。在scRNA-seq数据分析中，识别细胞类群是细胞功能鉴定、细胞分化轨迹推断等研究的基础。开发准确识别不同细胞类型的计算方法一直是单细胞数据应用研究领域的重点。目前，已发表的相关算法主要有SC3<sup>[1]</sup>，CIDR<sup>[2]</sup>以及Seurat<sup>[3]</sup>等方法。但是这些方法都需要预先设定细胞类群数目（K），对于组织中细胞类群数目未知的研究提出了挑战。此外，这些方法通常无法准确识别组织中数量较少的稀有细胞，例如在肿瘤发生发展过程中起重要作用的肿瘤干细胞、循环肿瘤细胞等。因此准确地从大量单细胞中识别出起重要作用的稀有细胞类群成为细胞类群鉴定的另一个挑战。针对上述问题，我们开发了一款全新的单细胞聚类分析算法，可以自动确定最佳的聚类数量，并且能够准确识别稀有细胞类型。我们目前的研究结果显示该方法运行效率高、对参数不敏感，相比现有方法的分析结果，可以发现新的细胞类群并描述细胞之间的层次关系，对于研究肿瘤等复杂疾病组织内的微环境和异质性具有重要意义。

## 研究方法

该方法首先在低维空间上对细胞之间的距离进行分析，通过一阶差分识别主要的细胞类群（类群之间的边界明显），然后利用距离密度分析寻找每种细胞类群中的亚型，最终实现细胞大类到亚类的鉴定。

1. **主要细胞类型鉴定**（图1）：从单细胞基因表达矩阵出发，首先利用高变异基因进行降维分析，然后选取高密度区域的中心点作为起始细胞，计算该细胞与其他细胞距离的一阶差分，选取一阶差分最大值所在的细胞位置作为分类边界，从而确定主要的细胞类型。

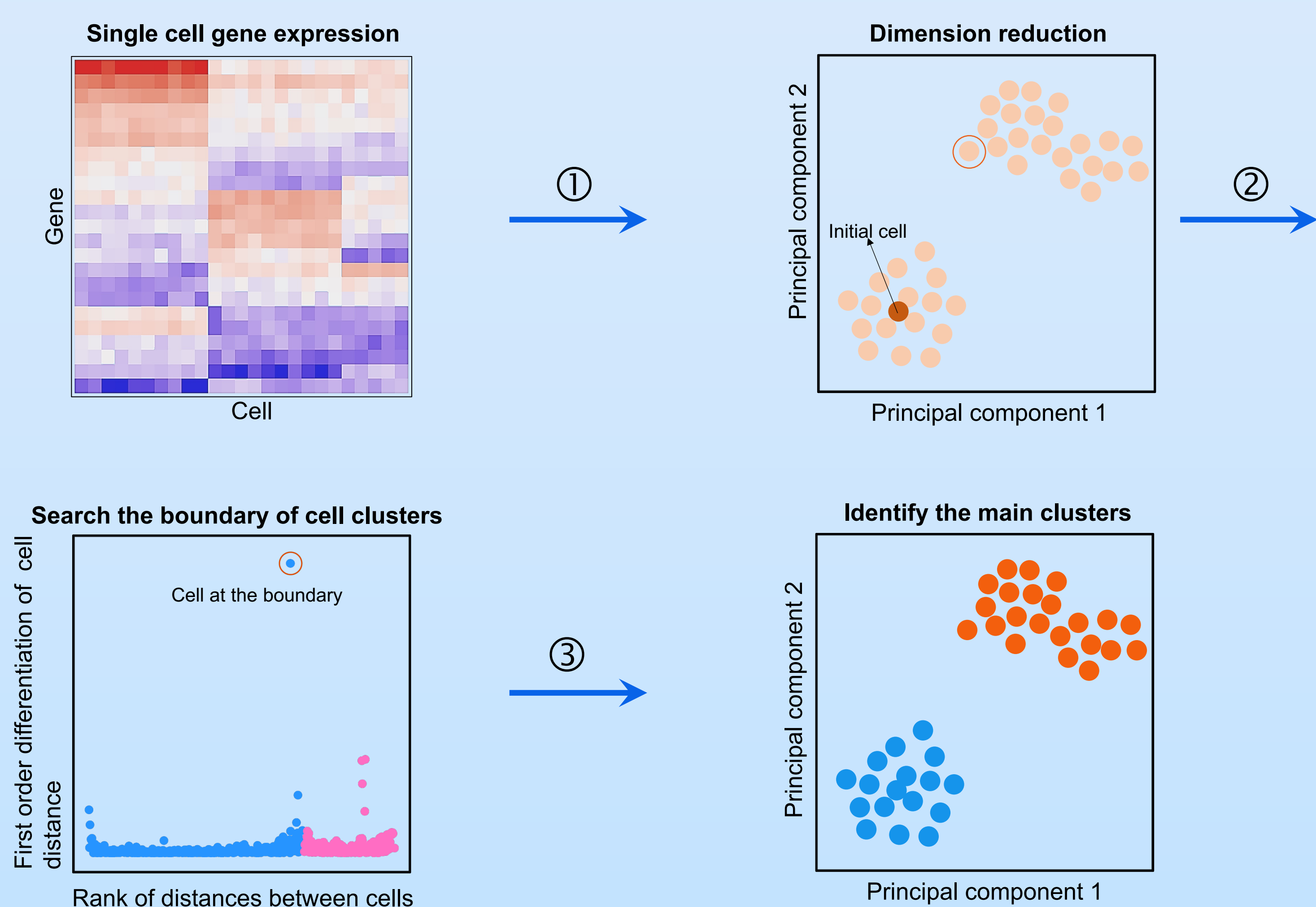


图1. 主要细胞类型的鉴定方法

2. **细胞亚类的鉴定**（图2）：对于每个已鉴定的主要细胞类型，选择关键维度上高密度的中心点作为起始细胞，计算该细胞与类内其他邻近细胞之间的距离，通过距离排序，并进行平均转换，寻找密度函数的全局最低点并作为细胞亚类的分界点。

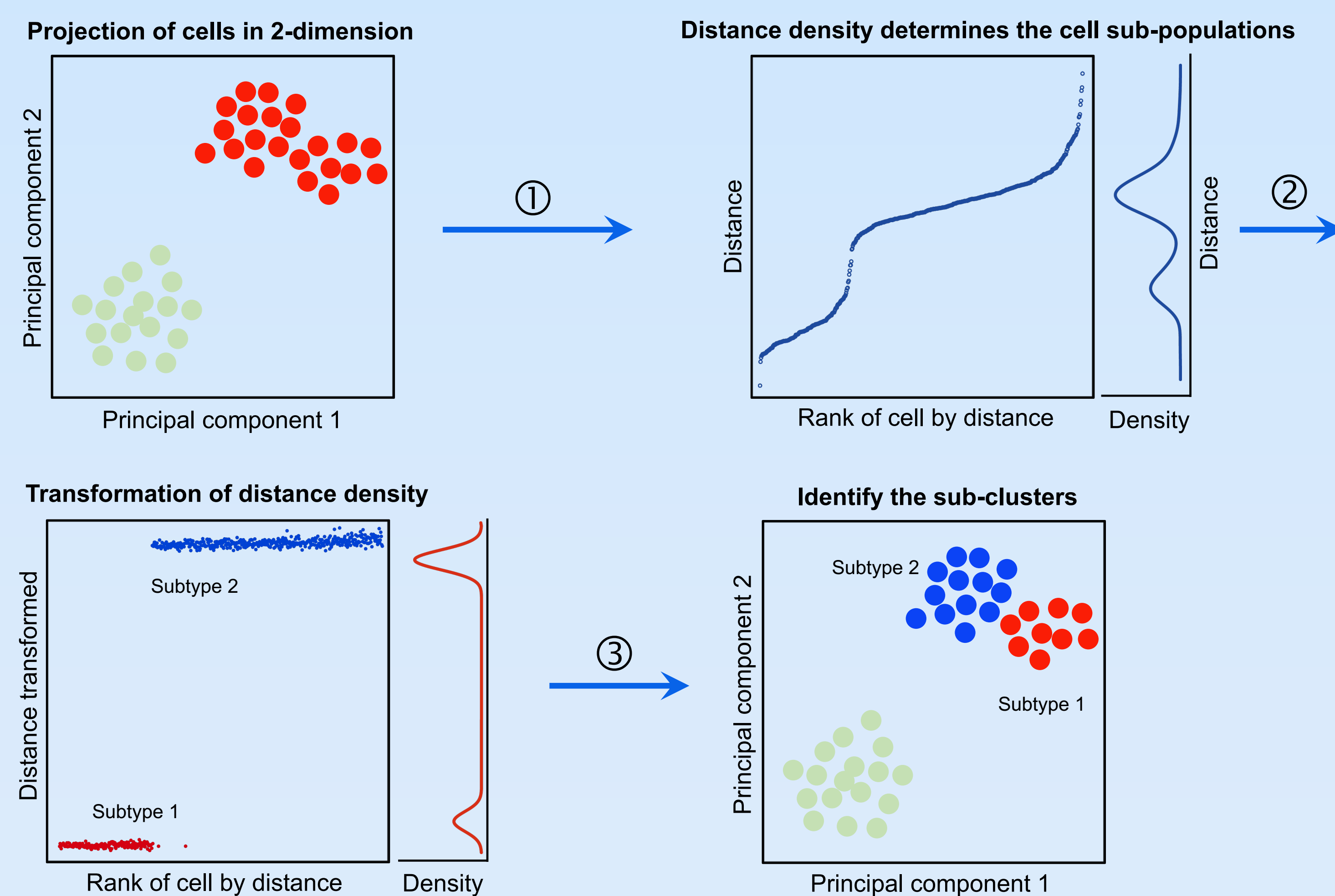


图2 细胞亚类的鉴定方法

## 研究结果

利用该算法，我们重新鉴定了小鼠视觉皮质组织的<sup>[4]</sup>细胞构成。其中主要的细胞类型有8种（图3左上所示），这些细胞类型有非常明显的标志基因（图3左下所示）。进一步可以将这些细胞聚为38类，我们发现了3种新的、未报道的细胞类群（图3右上所示），并且这些新鉴定的细胞类群有很明显的特征基因（图3右下所示）。

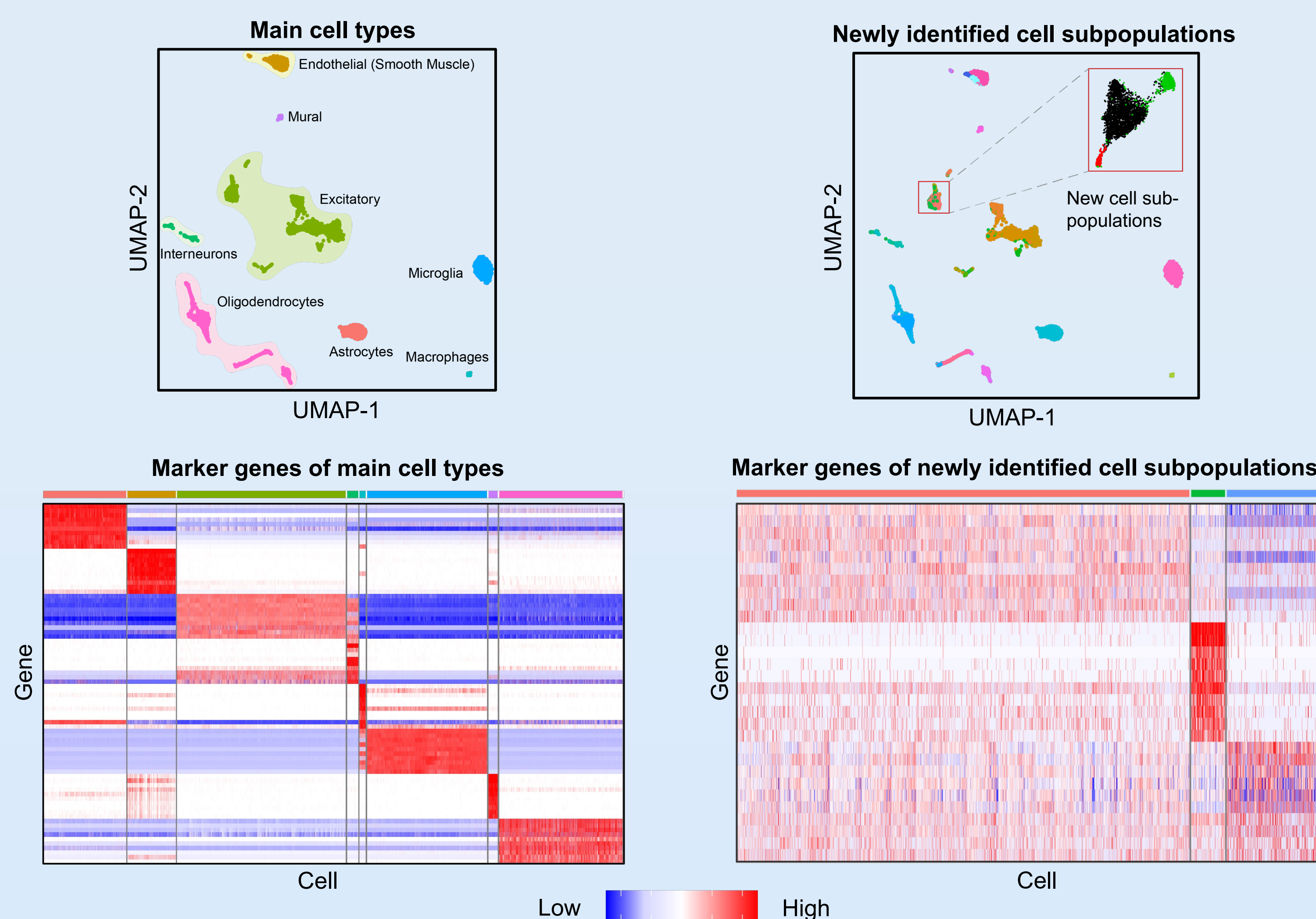


图3. 解析小鼠视觉皮质组织的细胞构成

## 结论

准确解析单细胞转录组数据，并对细胞进行分类和注释对于研究器官和组织的细胞构成及异质性非常关键。本研究，我们开发了一款智能的单细胞聚类方法，该方法可以逐步寻找关键的细胞类群，并进行更精细的亚群分类。我们利用该方法重新刻画了小鼠视觉皮质组织的单细胞构成，并发现了新的细胞亚类。我们将进一步优化该方法，并拓展其在单细胞研究中的应用范围。总之，该方法可以对器官、组织的细胞构成和异质性等实现更全面、准确的刻画。

## 参考文献

1. Kiselev VY, et al. 2017. Nat. Methods. 14(5):483-486
2. Lin P, et al. 2017. Genome Biol. 18(1):59
3. Satija R, et al. 2015. Nat. Biotechnol. 33(5):495-502
4. Hrvatin S, et al. 2018. Nat. Neurosci. 21(1):120-129

感谢Dr. Huang在研究中提供的帮助和支持！



PS. 欢迎大家关注Medicina专刊并投稿！