Automated identification of stratifying signatures in cellular subpopulations

（细胞亚群分层特征的自动识别）

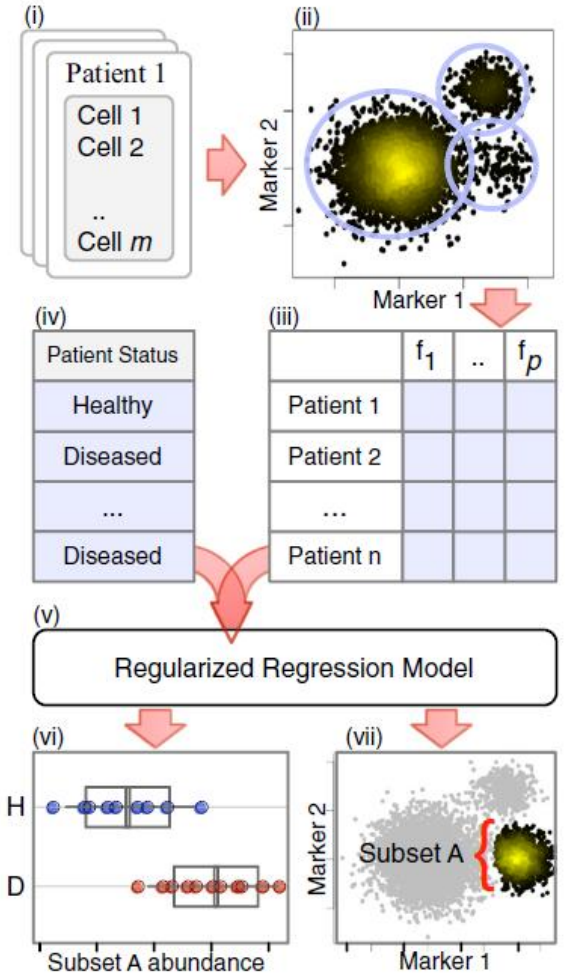
李健英

2014.10.13

Background

- 在特定的条件下识别细胞亚群在理解疾病分子机制有着很重要的作用，但是识别细胞亚群还是用手动来实现。随着细胞计数多参数化的发展，手动的分析会减缓实验进程和拓展性。本文开发了Citrus识别系统（ cluster identification, characterization,regression ），采用数据驱动的方法来识别细胞亚群分层。实验采用了Citrus系统鉴定出了已知和未知的代谢路径在应对外周血细胞受刺激时，除此之外在公共数据集中Citrus和现有的几种方法做了比较。随着流式细胞计数数据集的增加，Citrus需要更好的维护，尤其是对高维数的细胞集，对多种实验假设都可以进行测试。
- 流式细胞术只能测定特定条件下的12~17个参数，高通量大规模细胞计数，next generation of mass cytometry platforms(CyTOF)，测40个参数，并且产生表型丰富的数据集，可能传达相关生物信号。

Citrus analysis pipeline



所有样品所有的细胞层级聚类

每个样品细胞亚群被识别

回归模型

预测细胞亚群

In this example, the abundance of cells in subset A was found to differ between healthy and diseased samples (vi; H, subset A abundance in healthy patients; D, subset A abundance in diseased patients). Scatter plots show that cells in subset A have high expression of marker 1 and low expression of marker 2 relative to all measured cells (shown in gray).

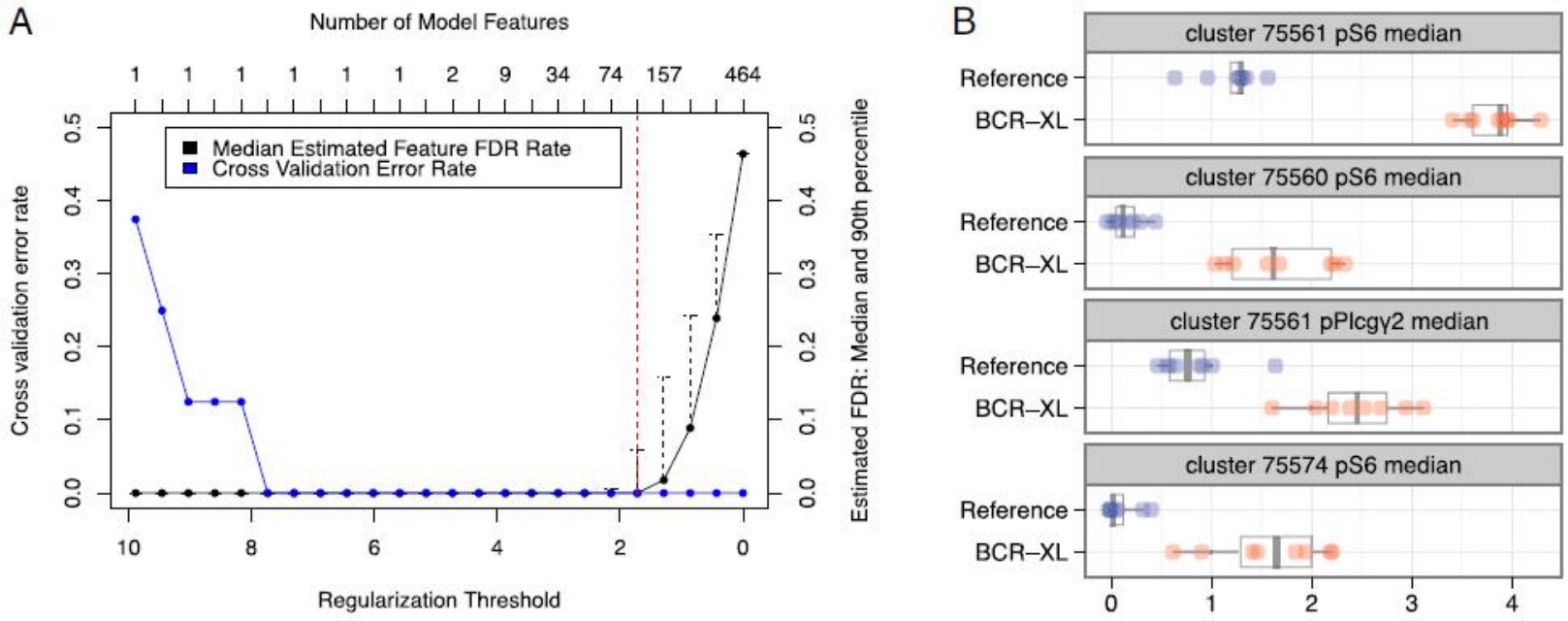
Validation of Citrus by Analysis of BCR/FCR Cross-Linked PBMCs

数据收集在8个人中16份PBMCs细胞中，8个受刺激的（应对B细胞和Fc细胞受体），8个未受刺激。

Table 1. Summary of markers measured in healthy PBMCs

Marker type	Measured marker
Lineage	CD45, CD4, CD20, CD33, CD123, CD14, IgM, HLA-DR, CD7, CD3
Functional	pNF-κB (pS529), pp38 (pT180/pY182), pSTAT5 (pY694), pAKT (pT308), pSTAT1 (pY701), pSHP2 (pY580), pZAP70/pSYK (pY319/pY352), pSTAT3 (pY705), pSLP76/pBLNK (pY128/pY72), pBTK/pITK (pY551/pY511), pPLCγ2 (pY759), pERK1/2 (pT202/pY204), pLAT (pY226), pS6 (pS235/pS236)

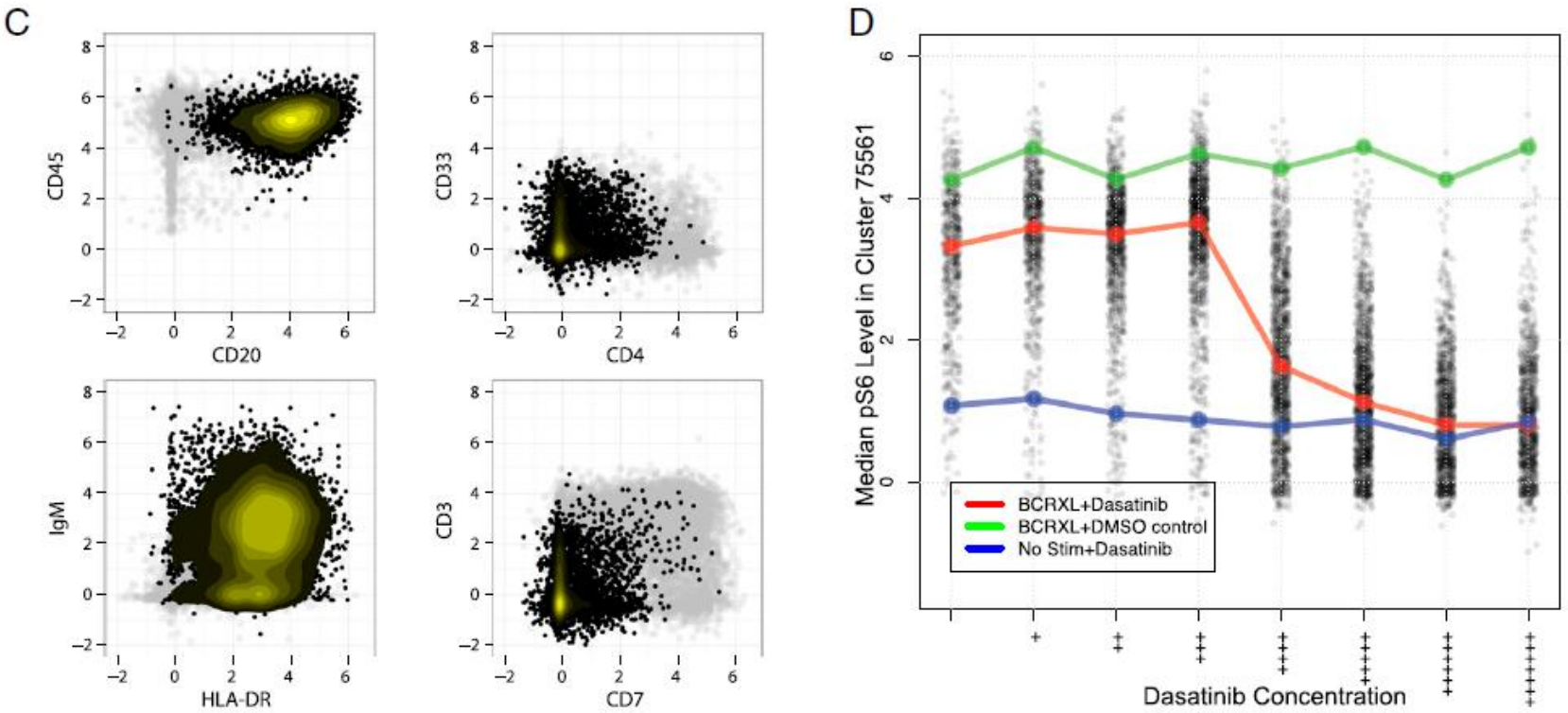
10个已知抗原和14个细胞功能标记



(A) Estimated model accuracy and feature FDRs as a function of model regularization threshold FDR <1%.

(B) The first 4 of 117 identified stratifying features between the unstimulated and stimulated samples

Identification of stratifying cell subsets

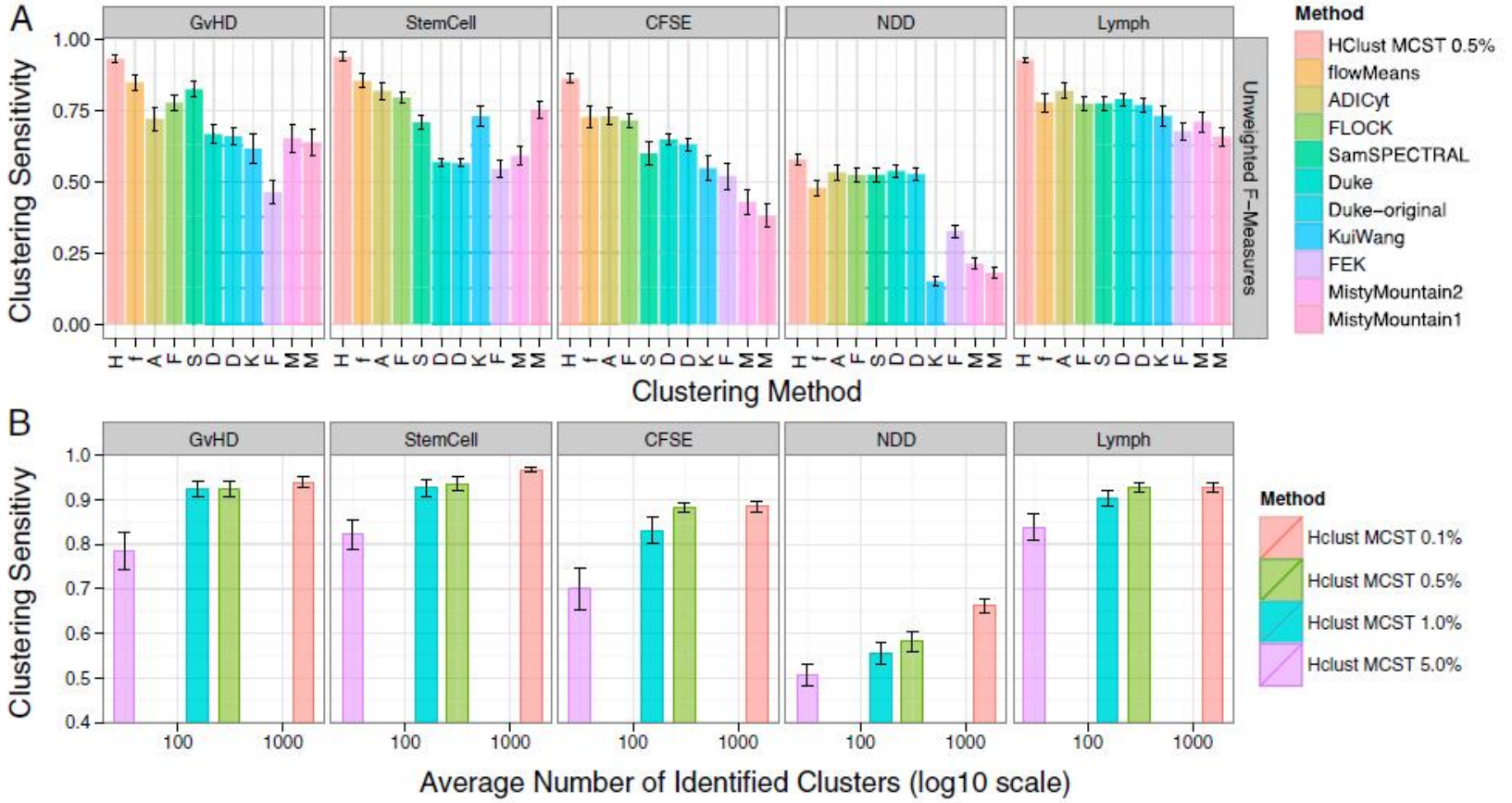


(C) Scatter plots showing lineage marker values from cells in cluster 75561.

(D) S6 phosphorylation levels as a function of dasatinib concentration in cluster 75561

在两种条件下鉴定了117个不同的细胞亚群，对这些细胞分层，确定了磷酸化水平（ phosphorylated S6 ），在75561个集群中markers CD20, CD45, and HLA-DR, but not CD3, CD7, or CD4是富集的，是B细胞的亚群。B细胞相关的通道蛋白被激活， Citrus可以识别成百上千个与响应相关的信号路径。

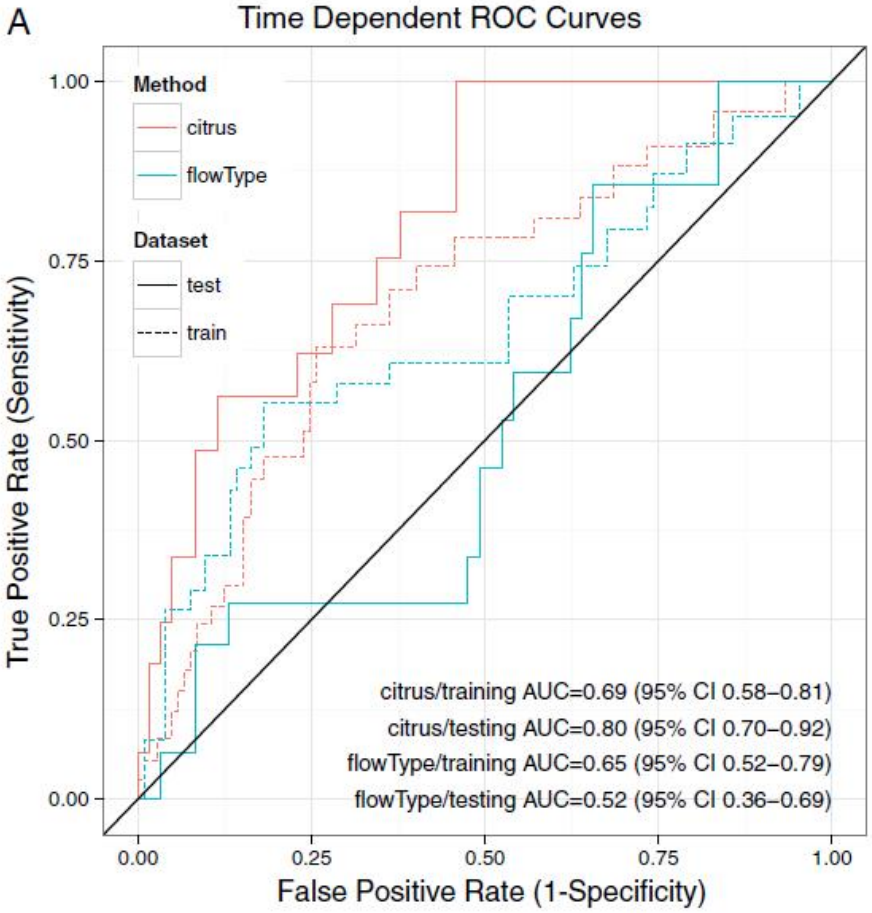
Evaluation of Citrus vs. Existing Methods

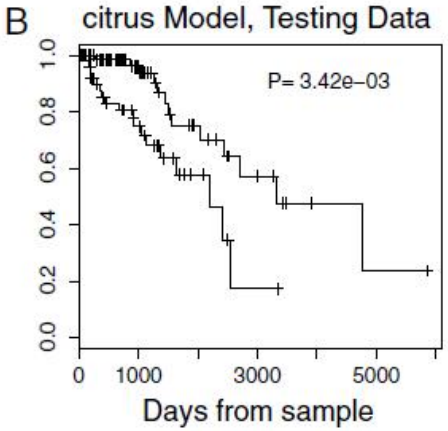


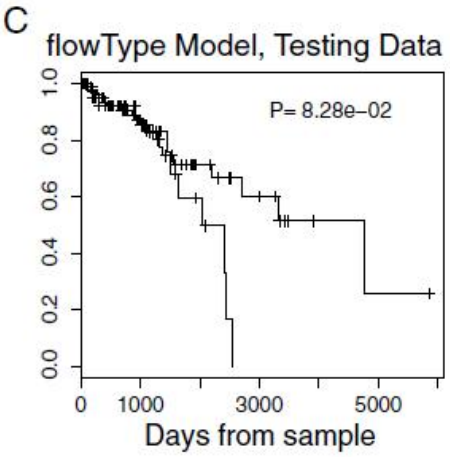
Clustering sensitivity of hierarchical clustering in FlowCAP-I datasets

(A) Clustering sensitivity measures from hierarchical clustering and other FlowCAP-I methods in FlowCAP-I datasets. (B) Clustering sensitivity as a function of number of identified clusters for hierarchical clustering. Minimum cluster size threshold (MCST), The smaller the threshold value, increase the cluster sensitivity .

Identification of prognostic cell subsets in HIV-infected patients







是反映敏感性和特异性连续变量的综合指标，ROC曲线下的面积值在1.0和0.5之间。在AUC>0.5的情况下，AUC越接近于1，说明诊断效果越好

两个模型Citrus和flowType models ： 训练组和测试组

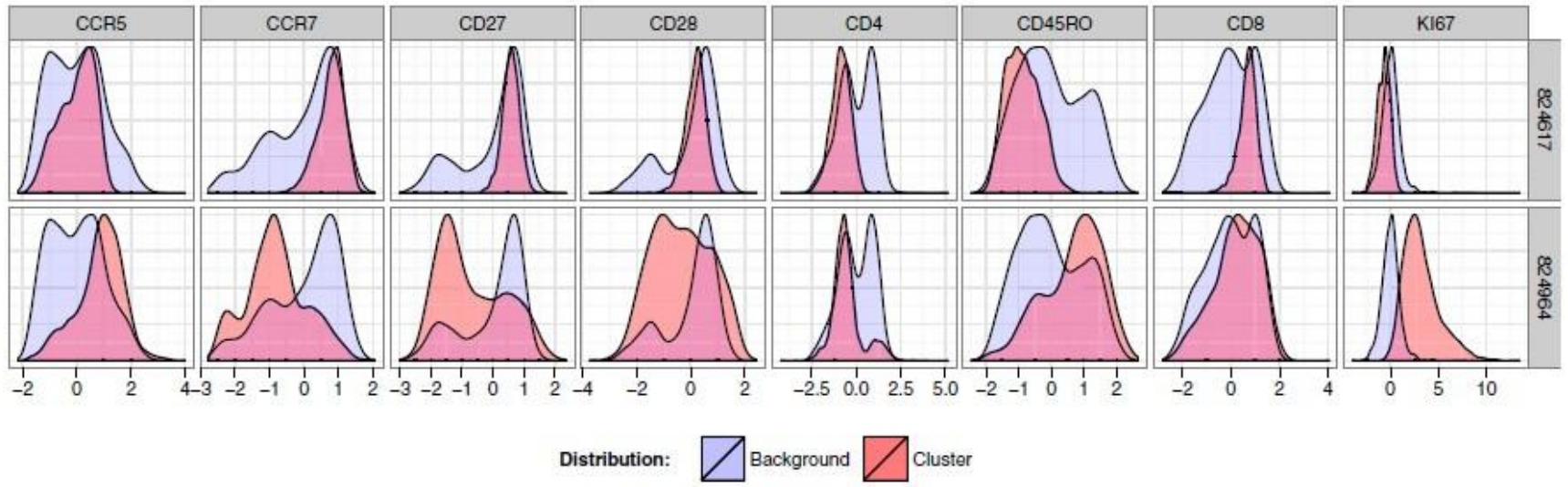
Identification of prognostic cell subsets in HIV-infected patients

Table 2. Summary of clusters frequently selected during cross-validation

Cluster ID	CV selection frequency, %	Coefficient average	Abundance average, %
824823	70	7.24	0.85
824971	70	−0.79	7.64
824715	80	−9.30	0.61
824617	100	−17.36	0.61
824964	100	15.79	1.49

HIV+AIDS 通过Citrus分析，根据表型鉴定了5个亚群细胞

D



Phenotype plots of clusters that were selected in all 10 cross-validation models. Both naive CD8+ T-Cells and Ki-67+ cells were identified as having prognostic utility in previous analyses.

Citrus was used to analyze datasets from the Acute Myeloid Leukemia (AML) and HIV Vaccine Trials Network (HVTN)FlowCAP-II challenges.

Method Sensitivity in Relationship to Analysis Parameters

Citrus分析需要每个样品中细胞的数目，聚类时功能标记，MCST参数选择。结果显示，每个样本数量选择并没有对集群有很大的影响，在这些数据的敏感性。足够的细胞被聚类，确保其准确性，50个细胞/样品以上。

MCST的选择，在回归分析中，找到最小的细胞亚群，阈值越小时，会增加聚类的敏感性，PBMC analysis采用了不同的MCST参数，MCST很小时，会识别稀有的细胞亚群中的信号通路，但是不能解释小亚群与整个数据集的关系。

文章亮点

Citrus可以处理多维数数据集，可以是多种实验条件

Citrus样品集团分析，识别罕见的细胞亚群

Citrus缩小实验偏差

Citrus采用了数据驱动的方法，解决了拓展性，主观性，敏感性以及细胞亚群间的相关性

主要结论

- Citrus自动识别细胞亚群分层，数据集由流式细胞技术产生的数据，利用层次聚类的方法，计算每个亚群的特征，正则化的回归模型，识别细胞亚群表型特征，实现细胞亚群分层。

- 随着高通量细胞计数的发展，数据集会越来越多，Citrus需要更好的维护。

Thank you for your attention