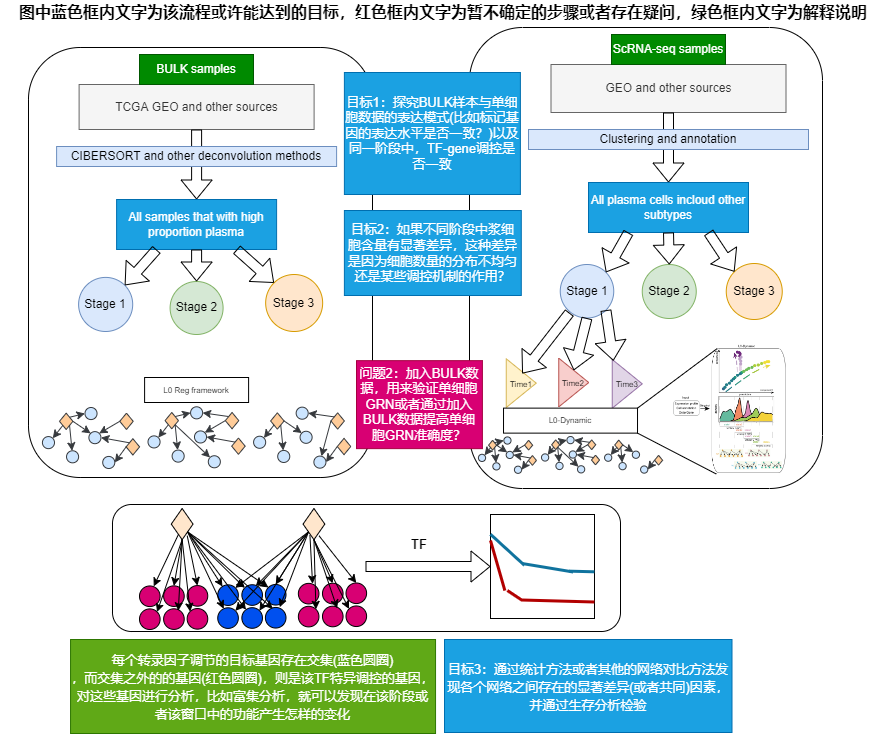
***Methods***



(图左上部分)BULK数据中虽然无法直接得到各种免疫细胞，但是可以通过CIBERSORT等去卷积方法，计算BULK样本中的免疫浸润程度，以此估计各种免疫细胞的含量，并且通过与单细胞的表达谱进行参考后，具有高度一致性。那么如何更合理的改进/调整CIBERSORT算法本身或参考表达谱便是从BULK数据中推断出B/Plasma细胞含量的关键，并且得到具有较高含量的B/Plasma细胞的样本。

(图右上部分) B细胞的不同亚型(包括Plasma)在非小细胞肺癌的不同阶段中如何分化以及发展？

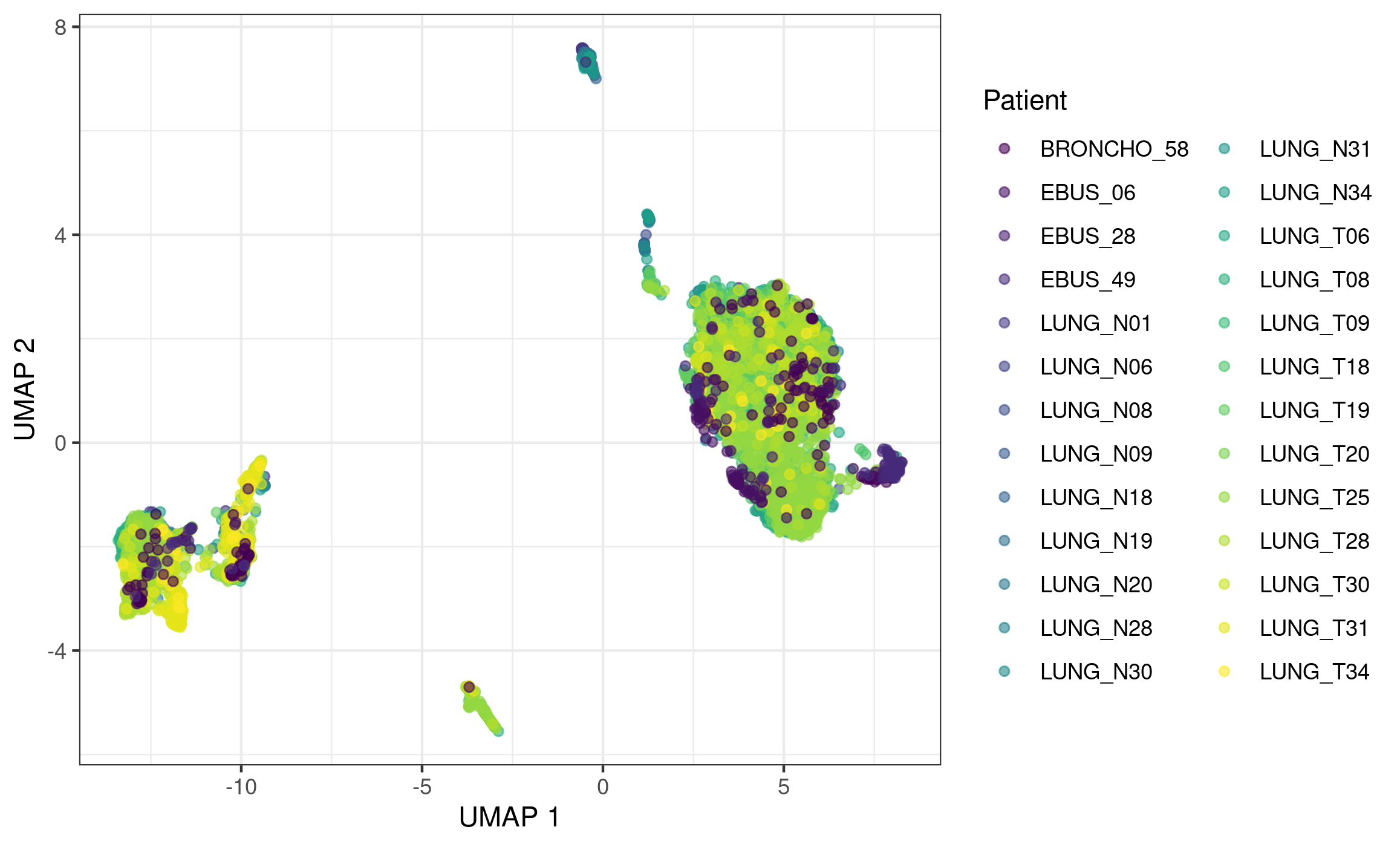
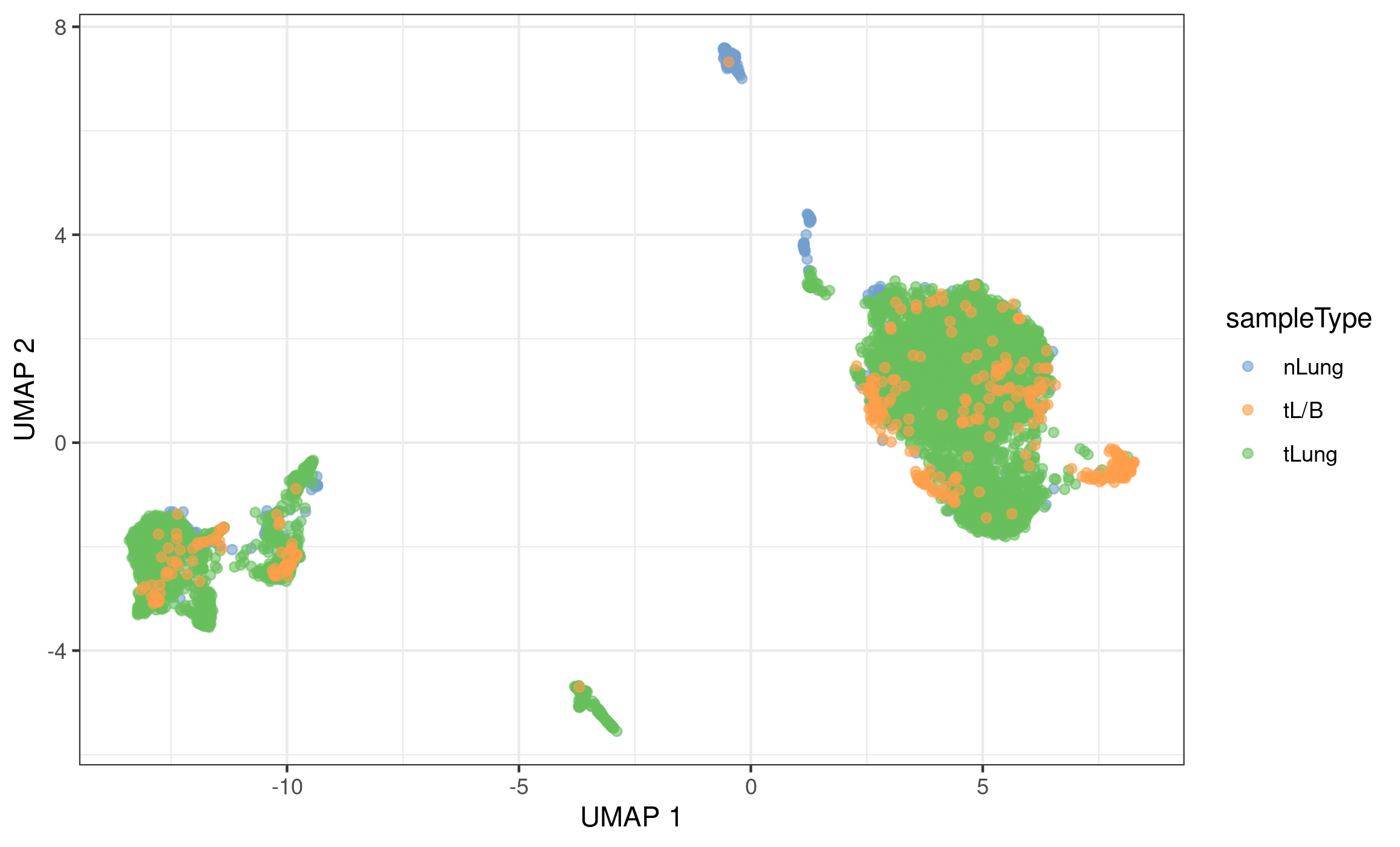
不同阶段的肺腺癌BULK及单细胞数据中各种细胞亚型的含量是否存在显著差异呢？

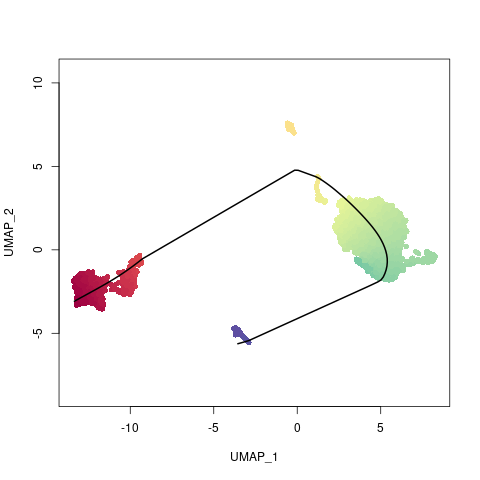
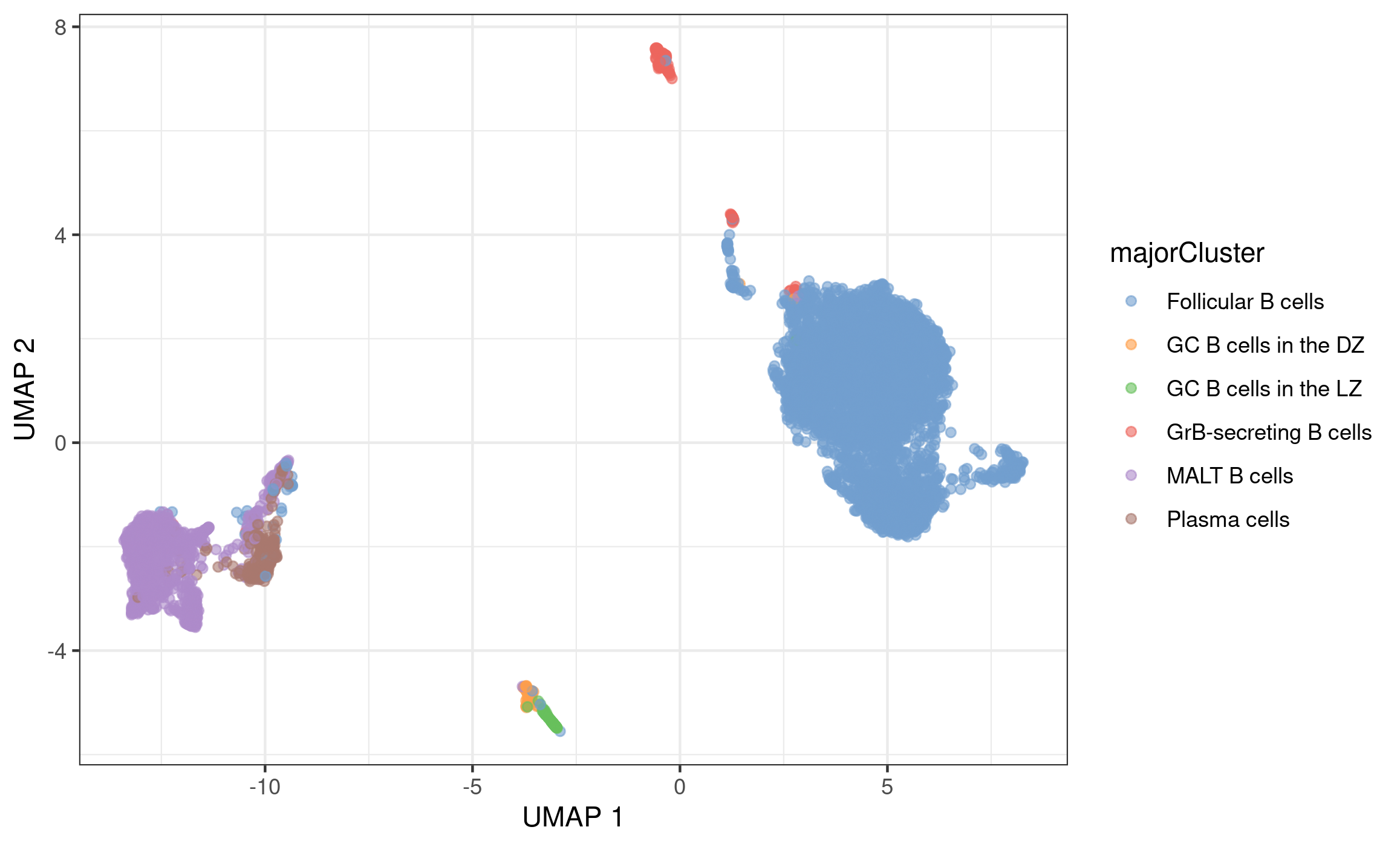
这种差异是因为细胞数量(单细胞)还是转录调控机制的作用所导致呢？

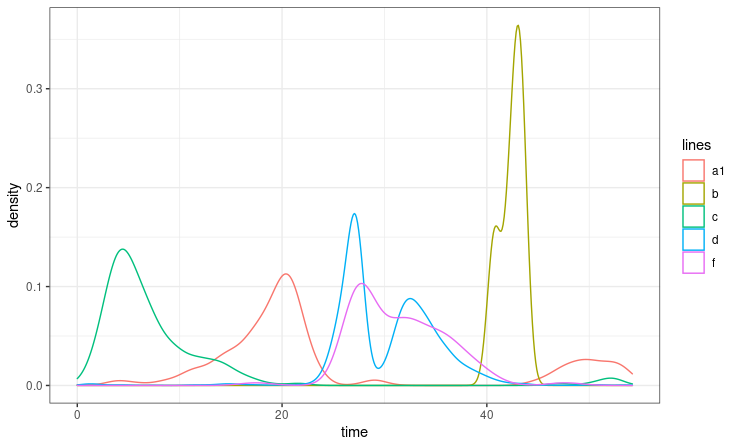
(图下部分)如果是因为转录调控机制的影响，那关键的转录因子是什么呢？

不同阶段的某种细胞亚型(比如Plasma)的伪时间是否与阶段的发展一致呢？

细胞丰度计算

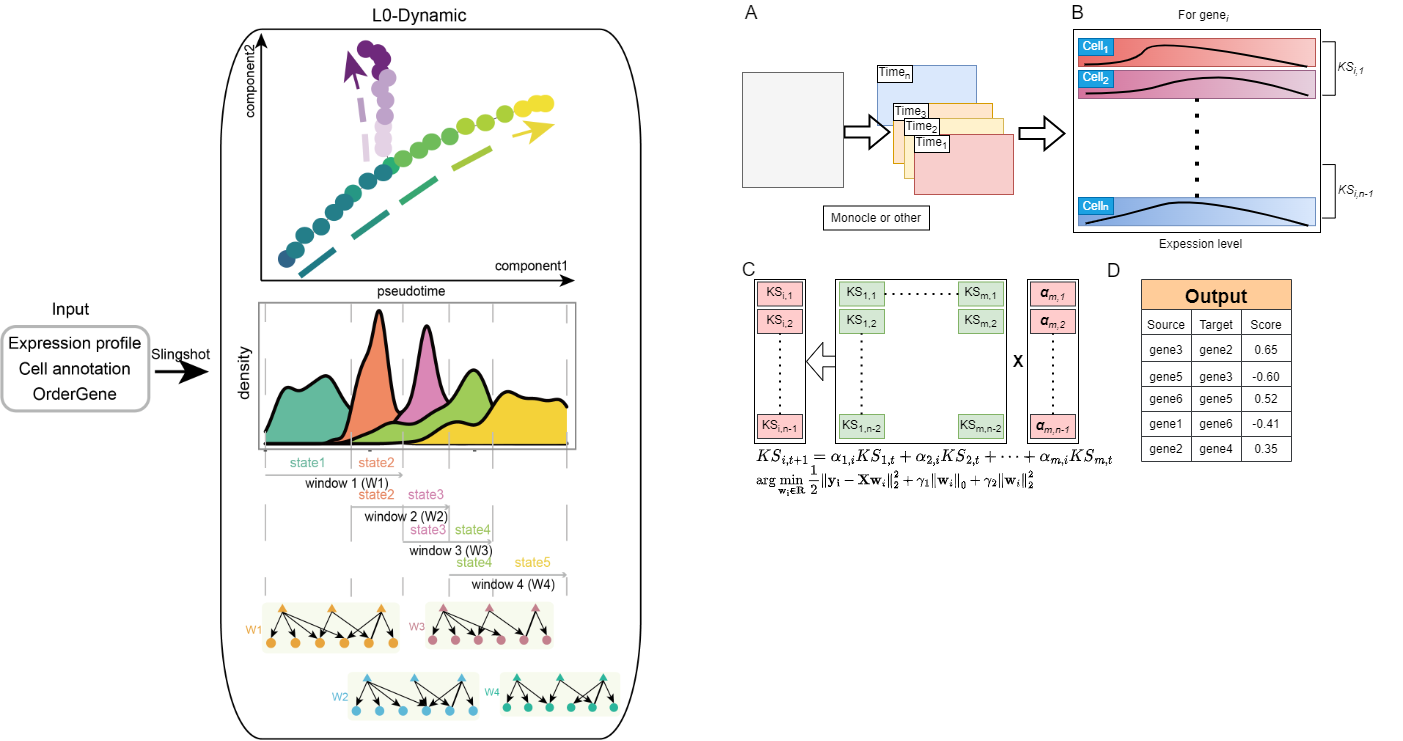






在这里使用了3种不同的阶段，nlung是正常细胞，tlung是早期非小细胞肺癌细胞，tL/B是晚期非小细胞肺癌细胞，应该是前面数据处理存在问题，存在一定的批次效应(图1，蓝色的正常细胞)，但是根据后续结果可以看出，最开始提出的思路是可以成立的，即通过动态窗口(图5)来对相邻阶段的细胞构建调控网络(下图左)。

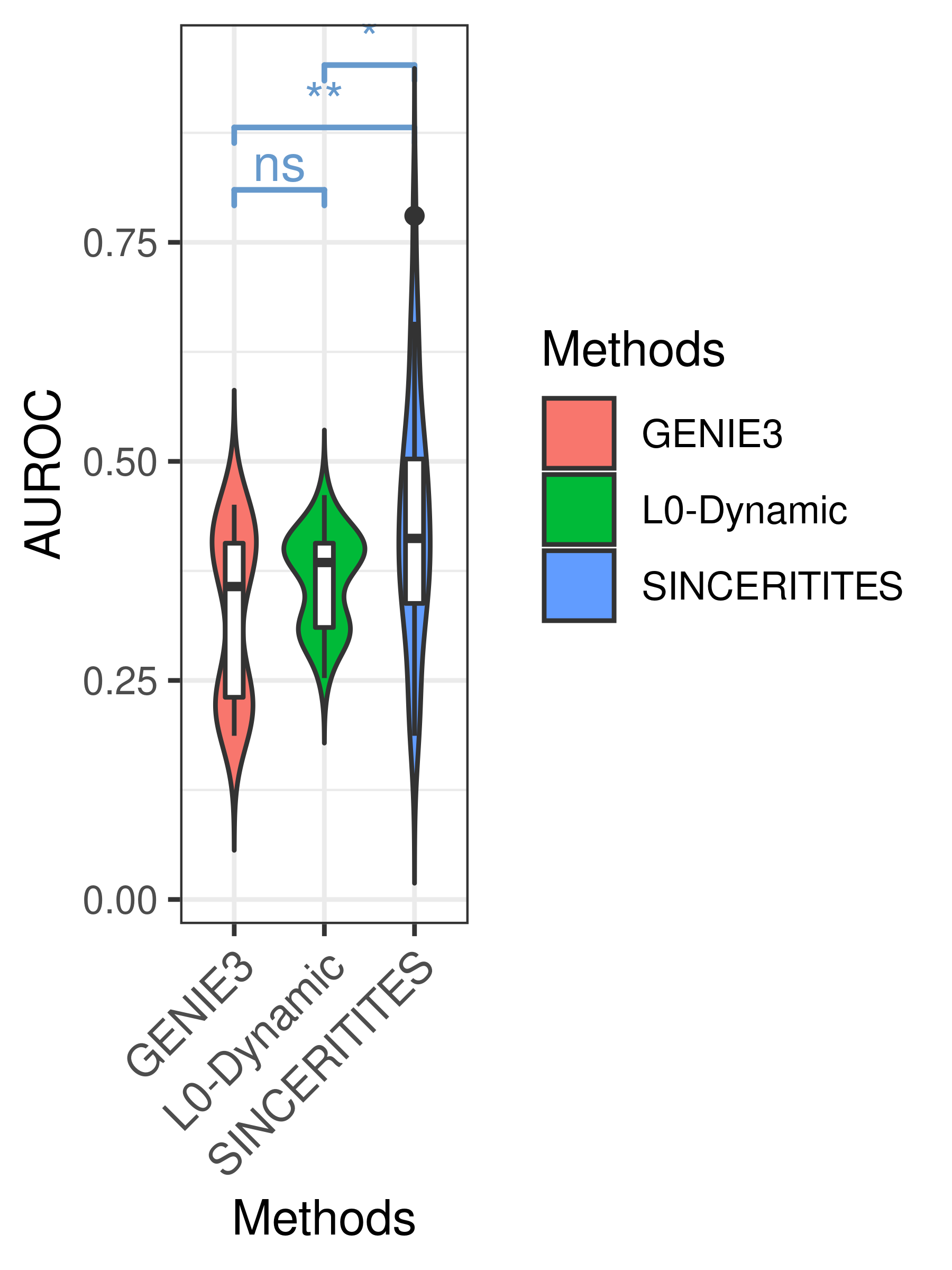
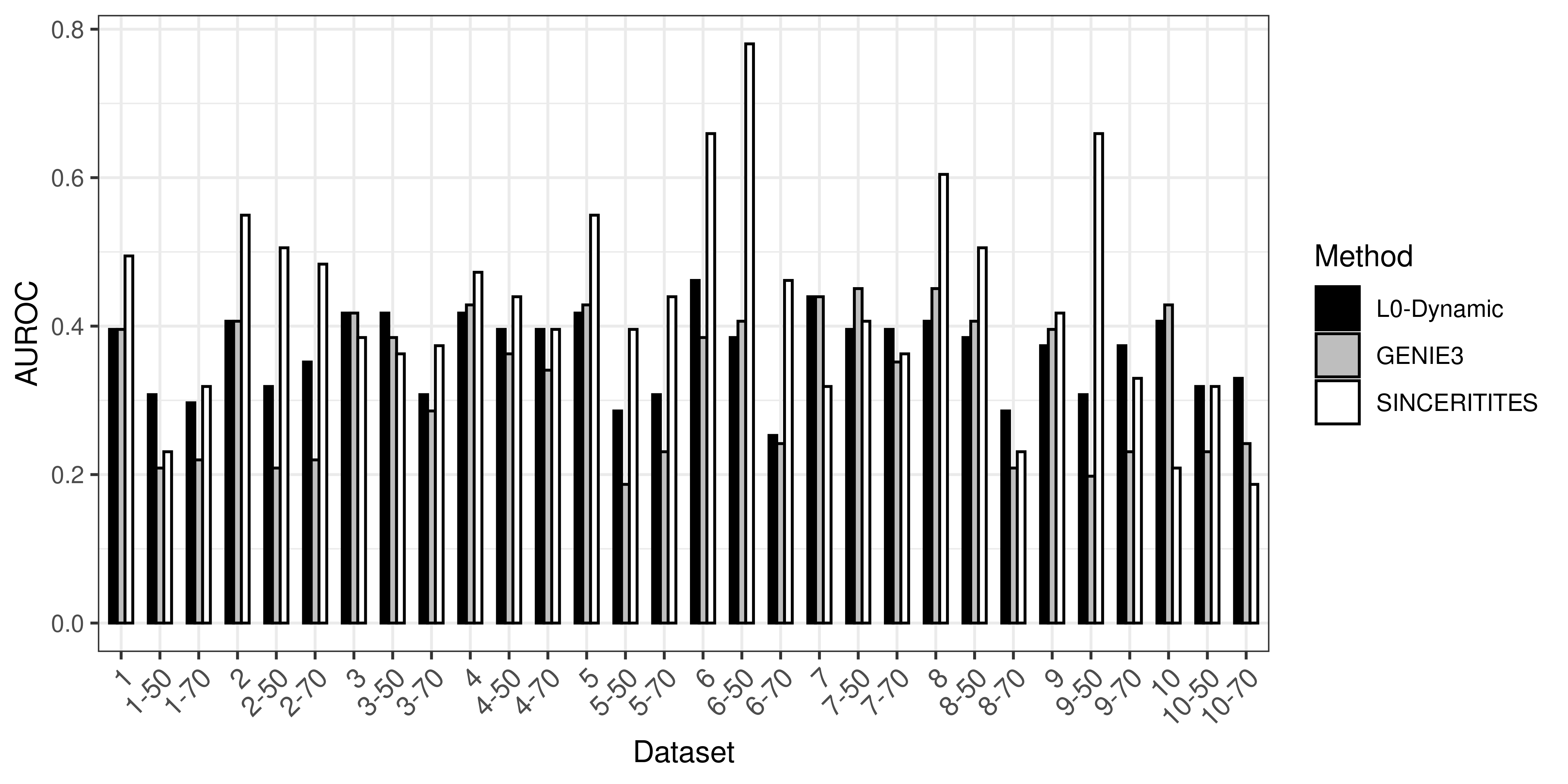
在进行拟时间分析时，需要筛选基因(一般是高可变基因或差异基因，tradeSeq algorithm)，是否可以考虑使用先前提出的结合熵和L0的特征选择方法(之前是选择可能影响各阶段发展的基因，在这里可以用在两个状态state1及state2所形成的window1中。这个思路还没有验证，但是有类似研究，通过改进特征选择方法以此来提高拟时间分析的准确度)。



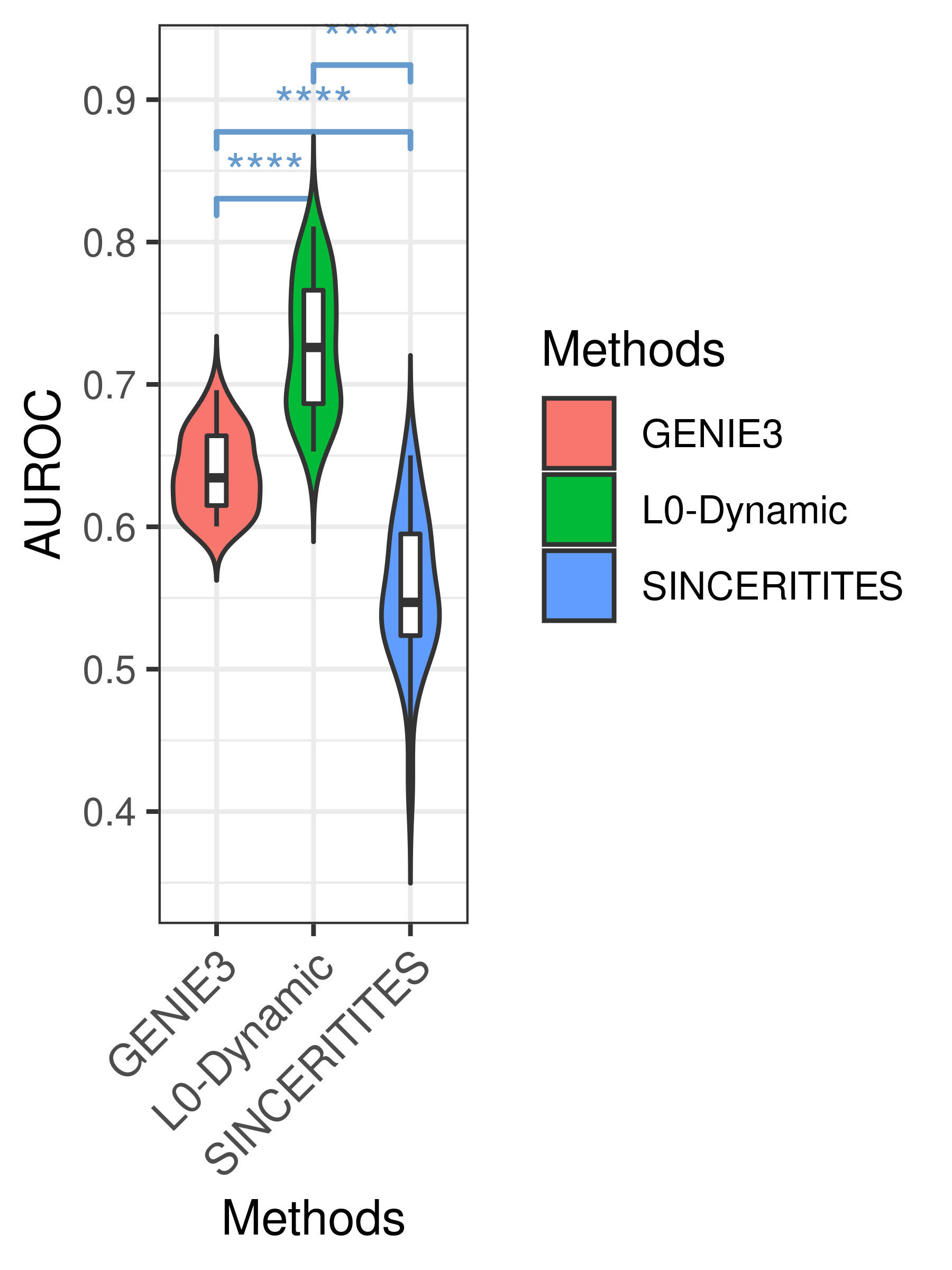
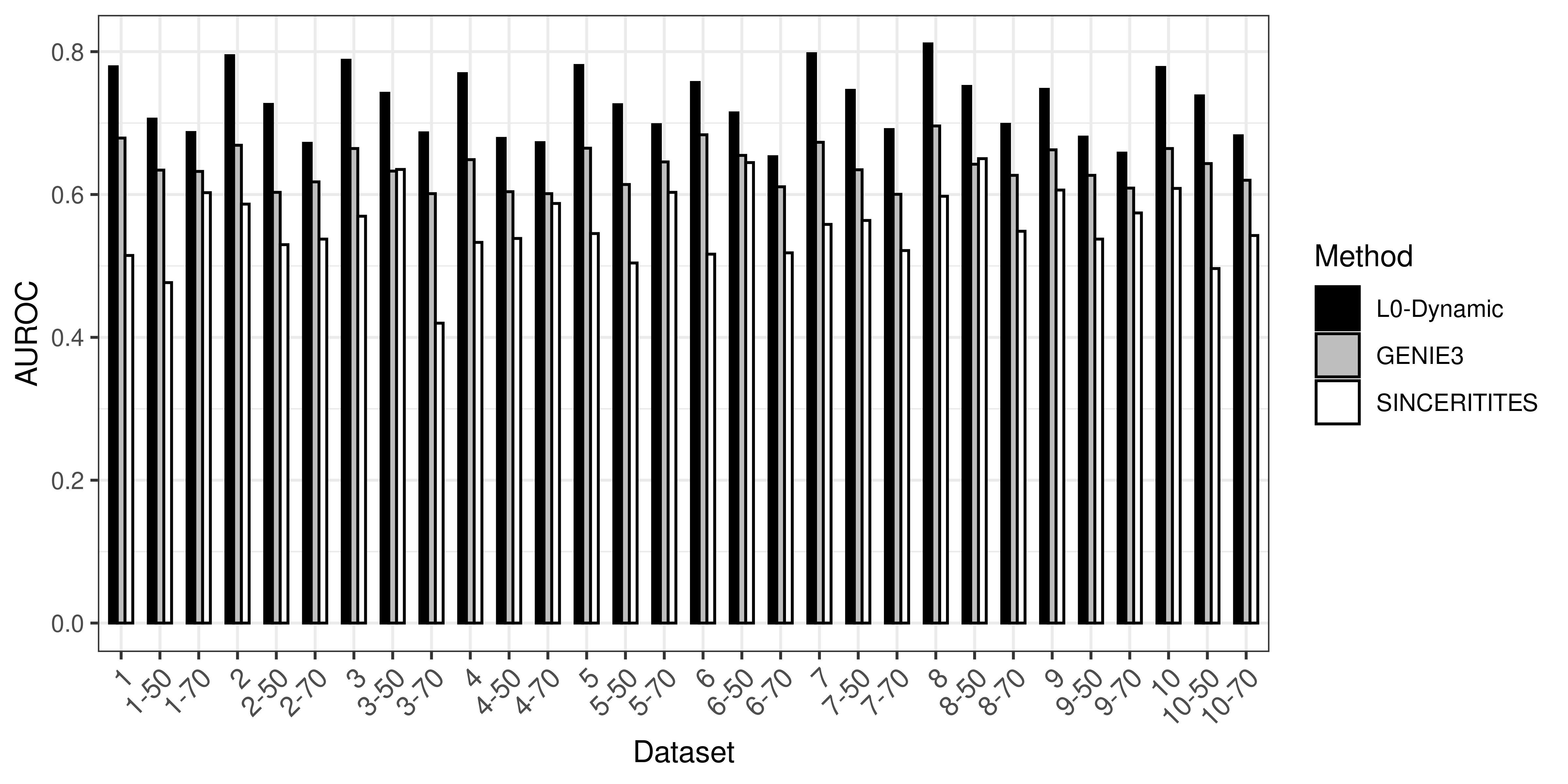
上左图是L0-Dynamic，即结合Slingshot或其他拟时间分析方法，对某一种细胞或者某种细胞及其亚型(主要作用)，将该细胞群划分为多个states，对于邻近的两个states可以组成一个window(类似于滑动窗口，不过不需要手动设定细胞数等参数)，通过L0Reg framework对该window构建调控网络。后续有与其他方法的详细的实验对比。

上右图是基于目标基因求邻居细胞之间KS值的方法，根据实验结果，该方法性能低于L0-Dynamic，并且计算非常耗时。

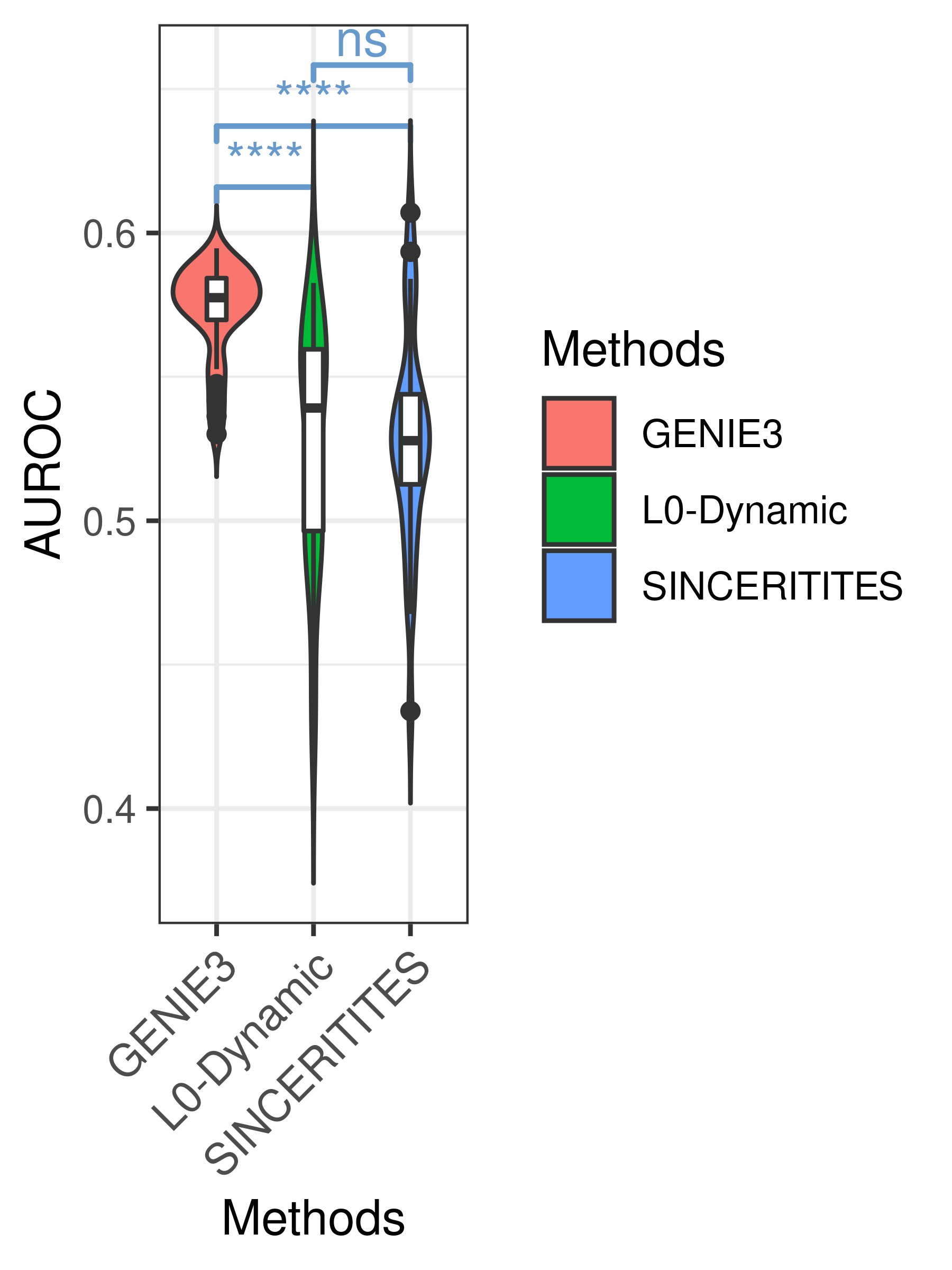
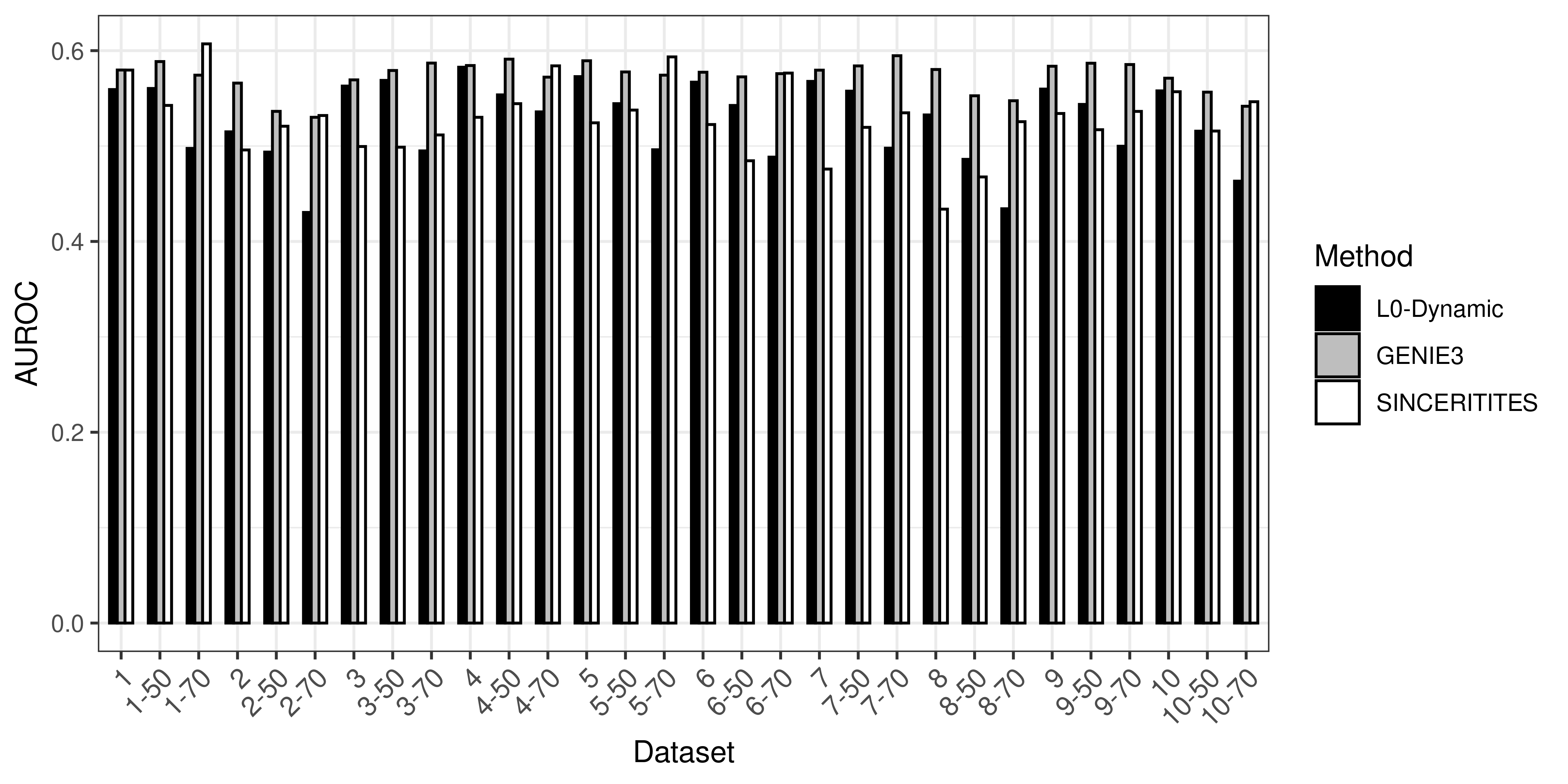
***TEST***



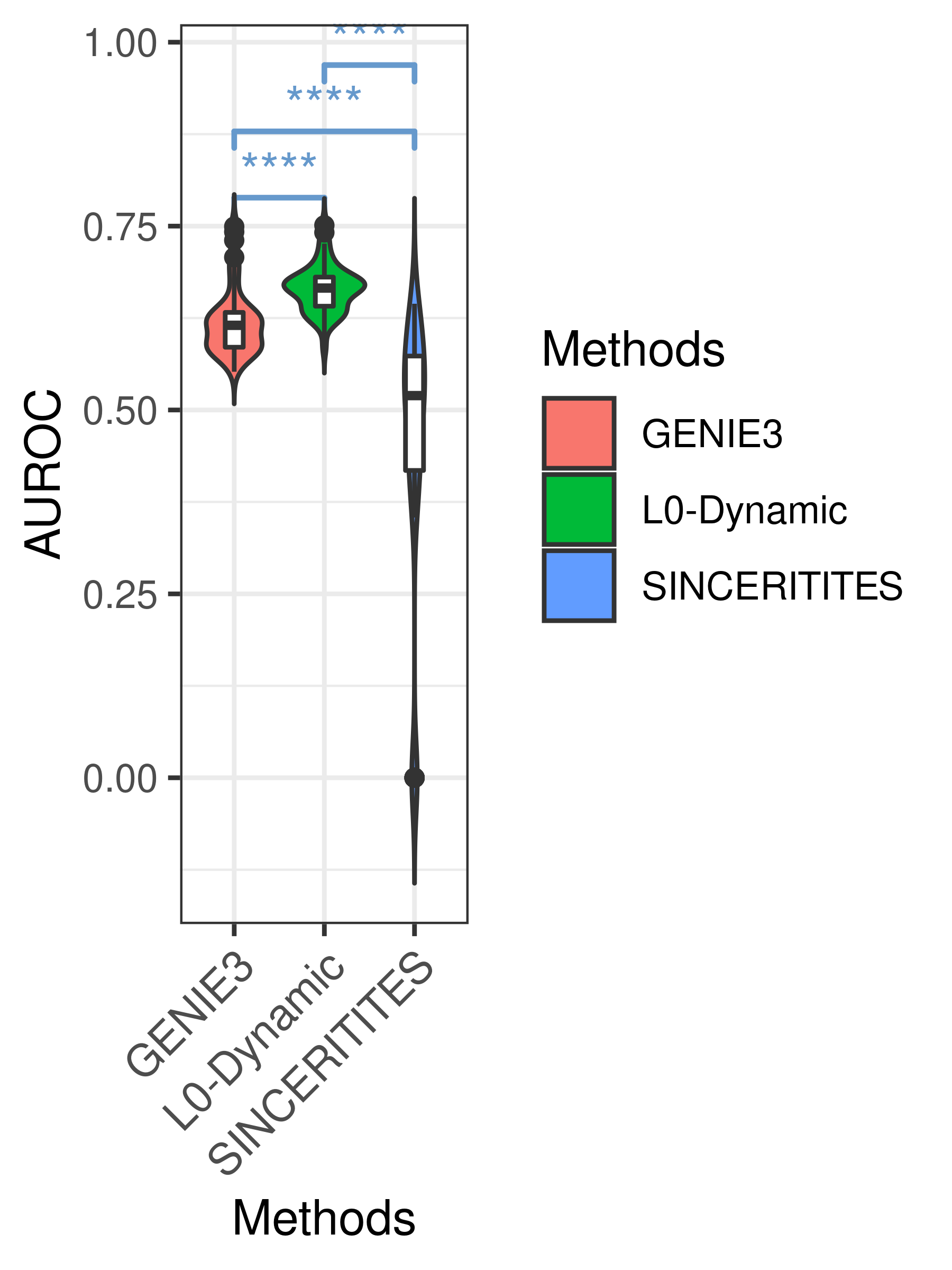
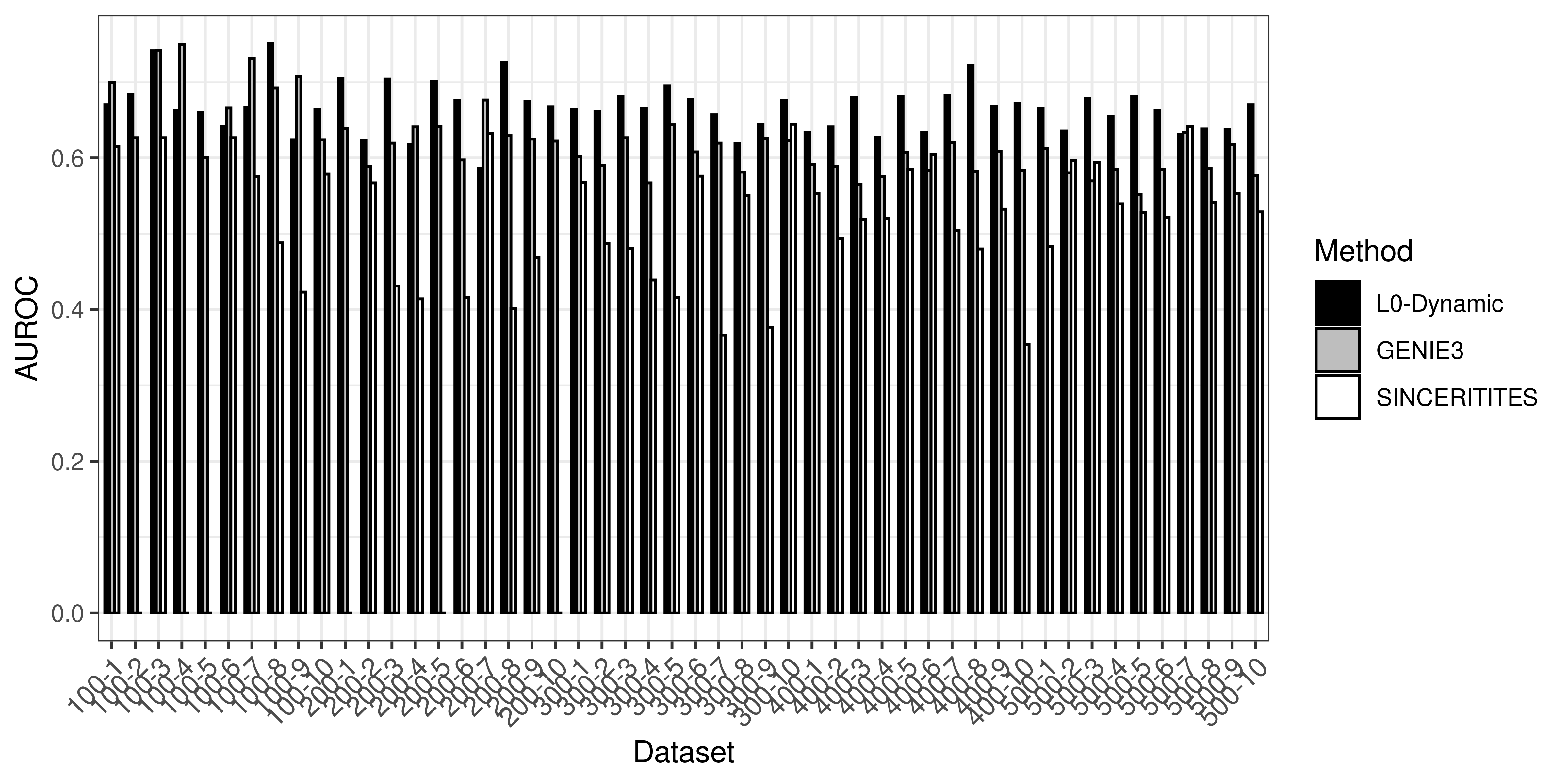
模拟数据(在该数据集上，三种方法表现均一般，SINCERITITES表现相对更好，该模拟数据集由微分方程(ODE)模拟得到，对于这三种方法，数据分布可能并不合适)

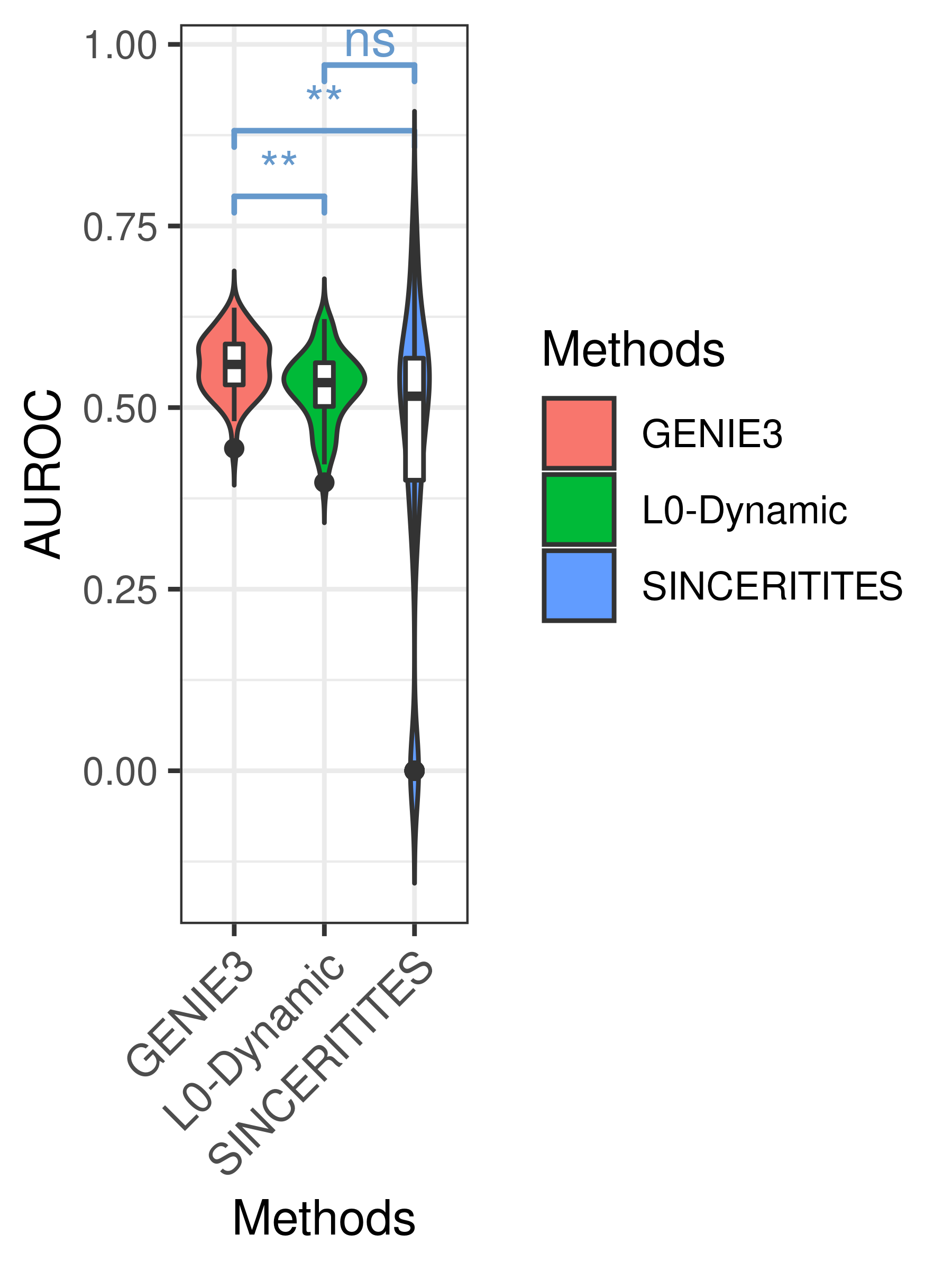
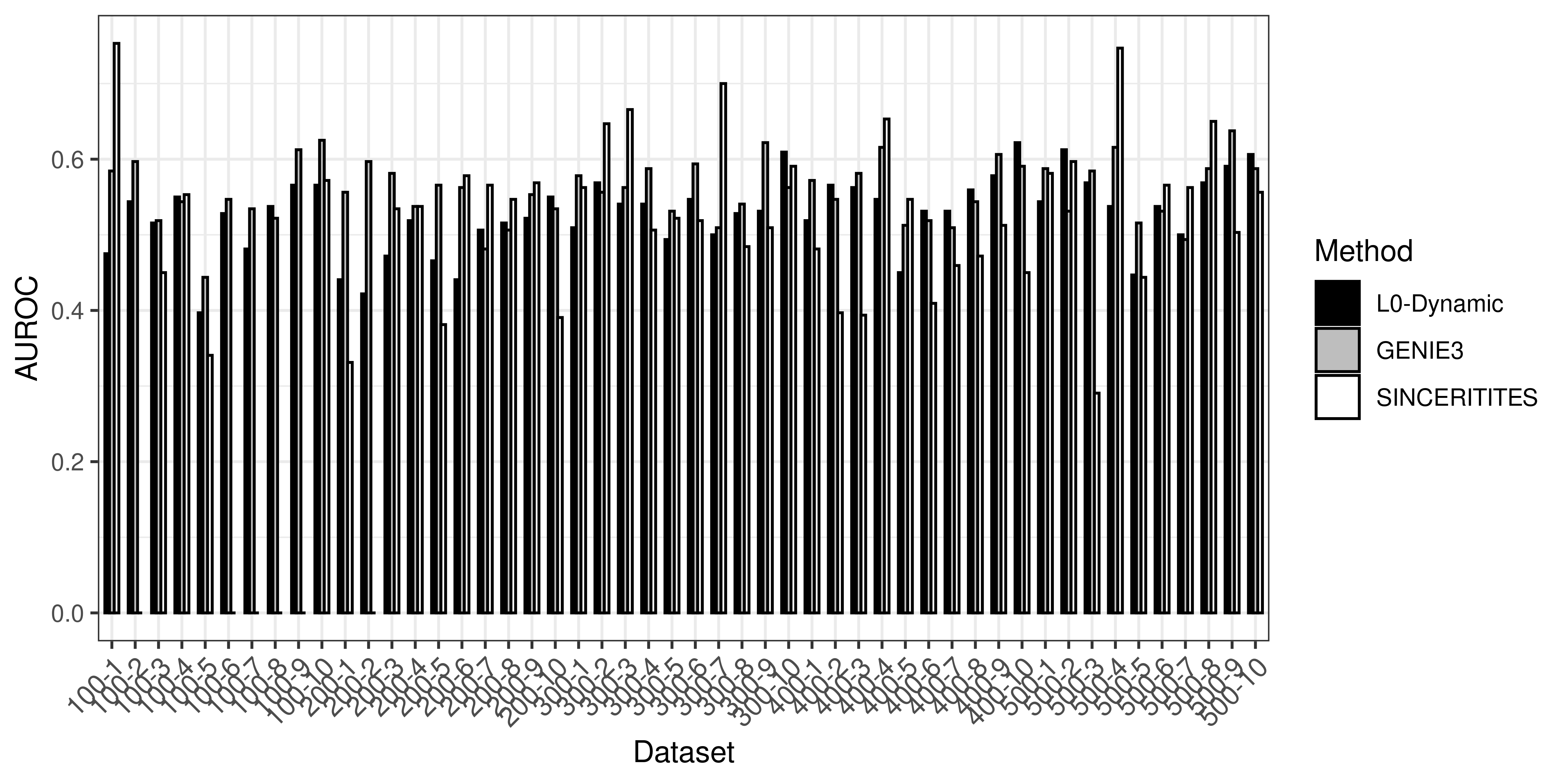


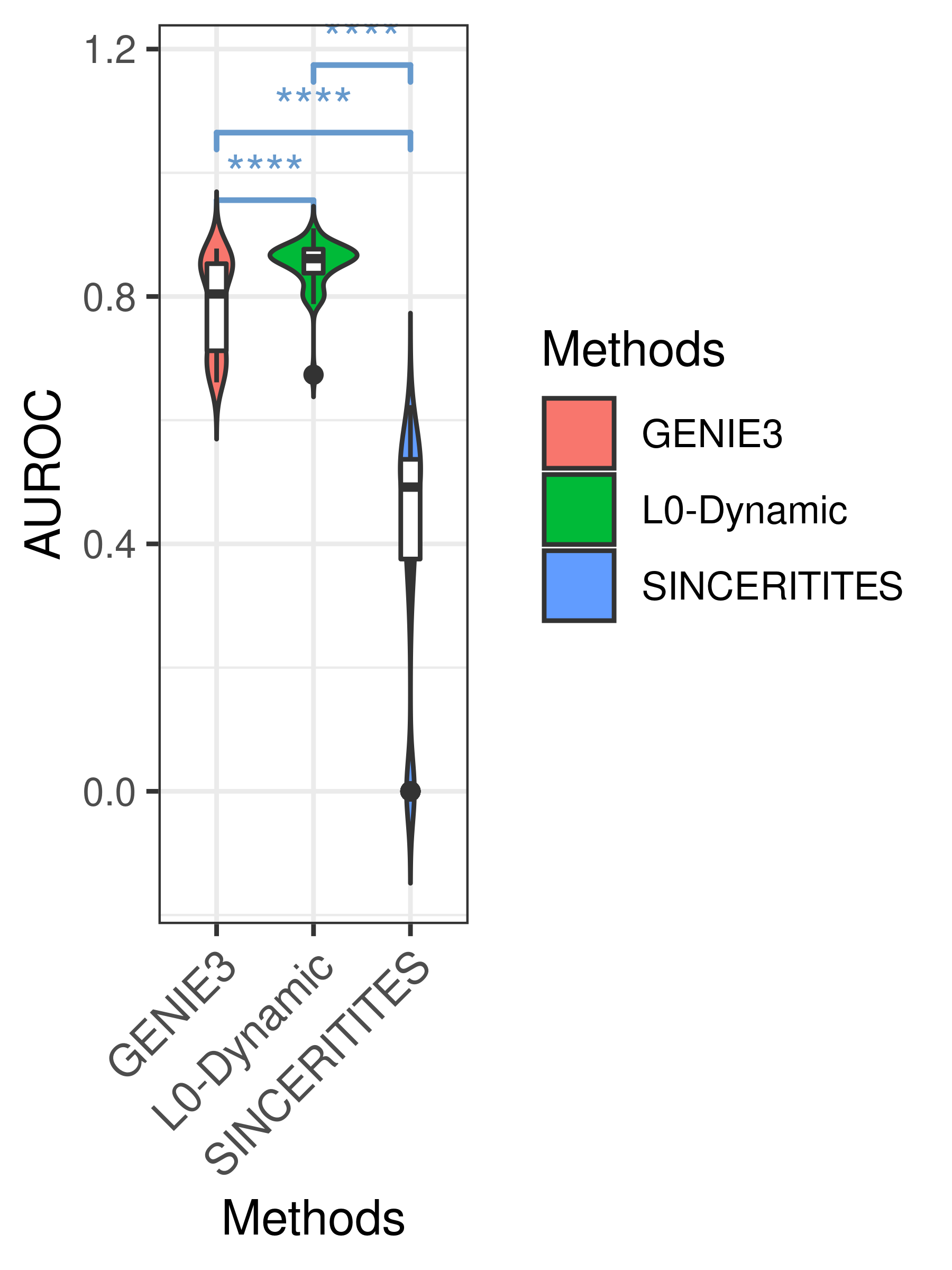
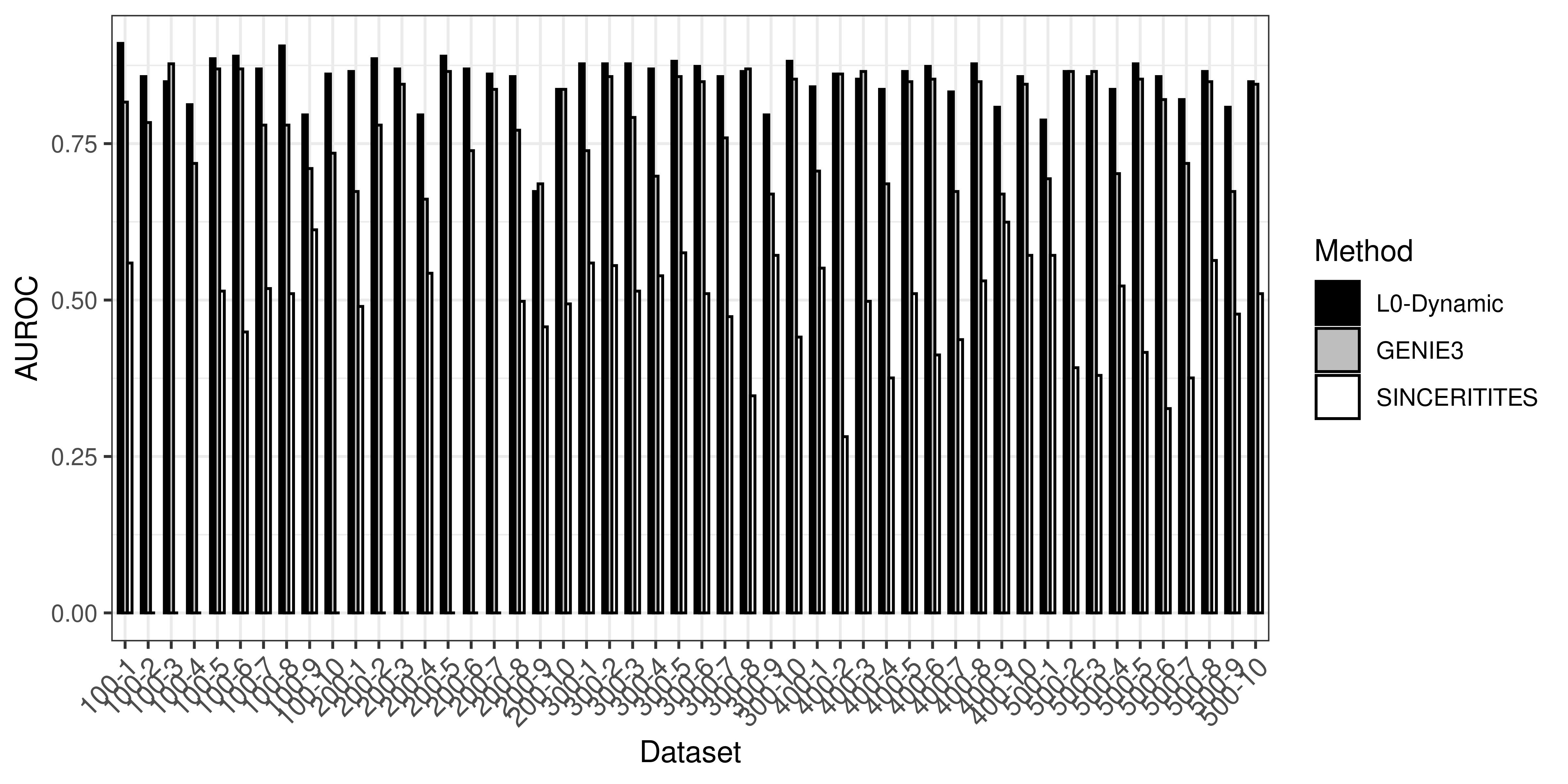
人类相关数据(在该数据集上L0-Dynamic表现最好，且三种方法随着dropout率的增加，性能降低)

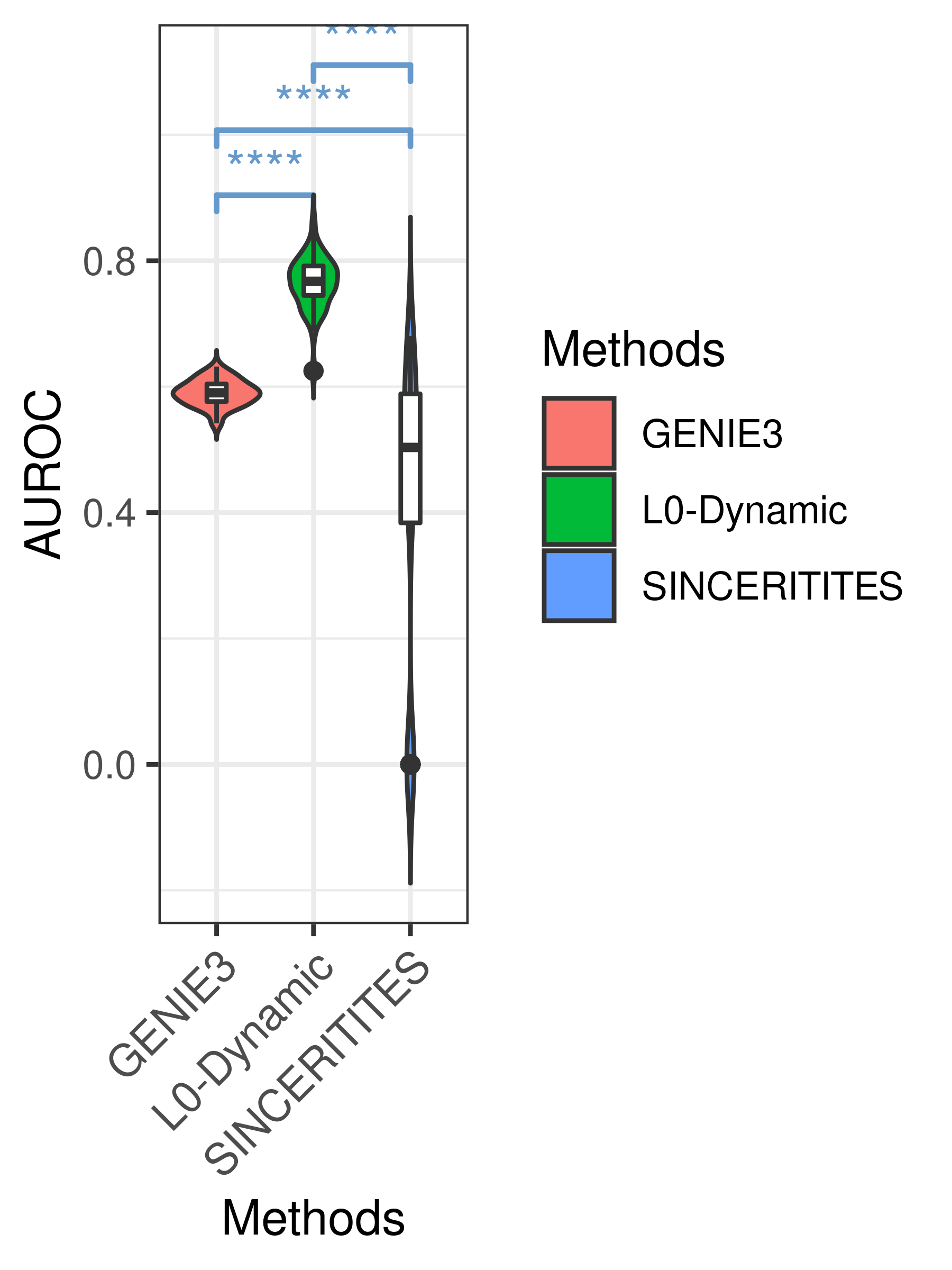
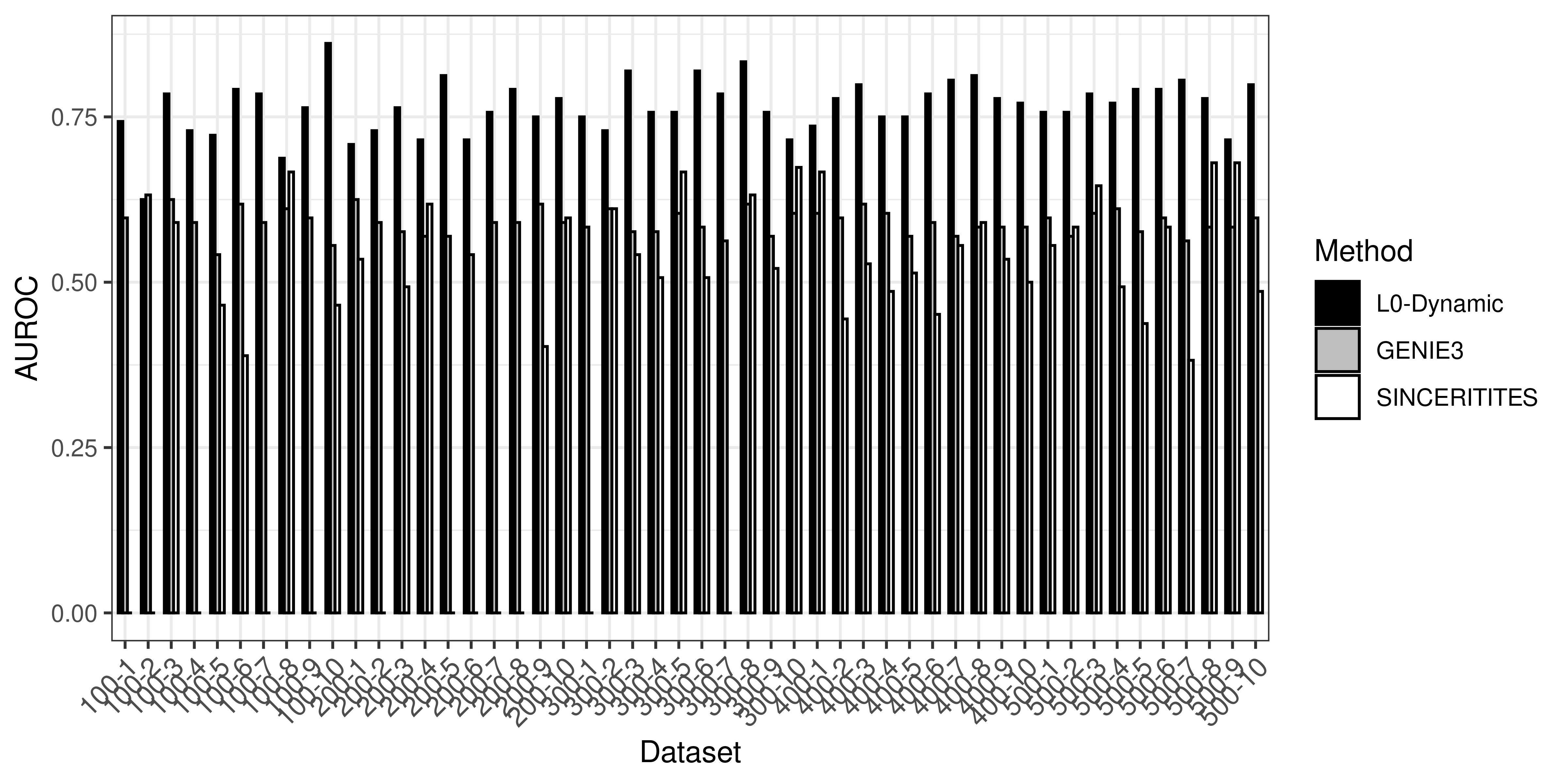


小鼠相关数据(在该数据集上L0-Dynamic与GENIE3优于SINCERITITES，并且GENIE3整体优于L0-Dynamic)









根据以上3组实验，可以看出L0-Dynamic和GENIE3即使不依赖于时间信息，也可以推断出更好的网络，并且综合来看L0-Dynamic性能在多数数据集上优于GENIE3，并且具有更稳定的性能。