
研究开发单细胞测序分析TEs表达的工具包

作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/progress/12956.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

研究开发单细胞测序分析TEs表达的工具包。中国科学院广州生物医药与健康研究院（以下简称广州健康院）陈捷凯课题组和南方科技大学Andrew Hutchins课题组合作开发出单细胞测序分析转座元件表达的工具包scTE。相关研究3月5日发表于《自然—通讯》。据悉，生物岛实验室副研究员何江平为该论文第一作者，陈捷凯、Andrew Hutchins为论文共同通讯作者。

转座元件（TEs）是人类基因组中含量最多的遗传信息，是指一类能够在基因组内移动的DNA序列。TEs的插入和跳跃可以改变基因组遗传信息，是物种进化的重要原动力。尽管大部分TEs已经丧失了跳跃的能力，成为了基因组中的分子化石。但近年来研究发现，TEs可以通过影响染色质表观遗传修饰、转录因子结合、RNA编辑以及染色质构象等，对基因表达起着非常重要的调控作用。

据介绍，单细胞转录组测序（scRNA-seq）是研究细胞命运状态的绝佳技术。近年来，更多的单细胞测序技术以及相应的生物信息学分析方法也在不断改进，总体的目标是扩展能捕获的信息及挖掘能代表生物学功能的潜在维度。然而，目前从scRNA-seq数据分析的工具包都只定量基因的表达，分析TEs的表达还比较困难，缺乏相关研究需要的生物信息学分析的工具包，因而忽略了基因组中含量最多的TEs来源的遗传信息。

为填补这一研究的空白，研究人员开发了能够从scRNA-seq数据中同时定量基因和TEs的表达的生物信息学工具包——scTE。由于TEs为多拷贝重复序列，与常规基因不同，同一类TEs通常有成千上万个不同的拷贝遍布在全基因组，并且每个拷贝间的序列高度相似，因此，针对TEs的分析通常难以做到单位点准确定量。

为解决这一问题，同时由于常规scRNA-seq只有较短的测序读长的因素，scTE采用了针对TEs家族层面的定量策略，这一策略忽略了基因组位置信息，提高了TEs的定量准确性。研究人员利用scTE，通过分析小鼠胚胎发育和人类疾病scRNA-seq数据，发现了一系列细胞命运、疾病状态特异性高表达的TEs，提示这些TEs的表达可能与胚胎发育或疾病的发生发展有关，也证明了通过scRNA-seq分析TEs的表达很有必要。

相对scRNA-seq研究转录组而言，单细胞ATAC-seq（scATAC-seq）等单细胞基因组技术研究的对象是染色质，染色质开放性与表观遗传修饰情况在很大程度上决定了基因表达丰度，以scATAC-seq为代表的单细胞基因组学技术的开发，让获得高分辨率的单细胞精度的染色质开放/修饰图谱变为可能，有利于构建从DNA到RNA再到表型的调控网络，寻找与表型强相关的核心调控因子。

然而，以scATAC-seq为代表的单细胞基因组测序数据存在几个特点：一、高维度。每个细胞中可以检测到可能开放的区域可以高达几十万个；二、数据的稀疏性。由于技术原因，导致大量开放的区域没有检测到信号。由于以上原因，目前对scATAC-seq数据的生物信息学还存在比较大的挑战。

研究人员提出由于TEs多拷贝的特点，通过累积TEs信号可以去除数据的稀疏性，降低数据的维度，或许能够有效解决scATAC-seq数据以上难题。他们利用scTE，发现仅利用TEs信号，能够将scATAC-seq数据中主要细胞类型区分开来，提示TEs信息对scATAC-seq等单细胞基因组数据的分析是很好的补充。（来源：中国科学报朱汉斌 黄博纯）

相关论文信息：<https://doi.org/10.1038/s41467-021-21808-x>

版权声明：凡本网注明来源：中国科学报、科学网、科学新闻杂志的所有作品，网站转载，请在正文上方注明来源和作者，且不得对内容作实质性改动；微信公众号、头条号等新媒体平台，转载请联系授权。邮箱：shouquan@stimes.cn。

作者：陈捷凯等 来源：《自然—通讯》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发