科学家开发从低分辨率Hi-C数据中提取高分辨率3D基因组结构的新方法

作者：writer 来源：中国科学院

本文原地址：https://www.iikx.com/news/progress/1462.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！

高等真核生物的核内活动，从基因转录、DNA的复制到DNA损伤修复等，都与基因组的三维空间结构息息相关。因此，3D基因组研究成为近年来的前沿热点领域。Hi-C是目前最常见的全基因组检测染色质三维结构的组学技术。然而，由于Hi-C具有分辨率相对较低、测序量要求极高的特点，在类似精准医学人群队列这样的大规模项目中，开展高精度的3D基因组研究在经济上几乎无法实现。如何低成本快速地获得高分辨率的染色质空间结构，成为大规模开展3D基因组研究的一大瓶颈。

中国科学院北京基因组研究所张治华研究组与中科院软件研究所李昂生研究组合作，开发了利用低分辨率的Hi-C结合其他表观组数据，预测高分辨率的染色质结构域和染色质相互作用的新方法，使得在大样本中，快速、低成本获得染色质高精度结构成为一种可能。该研究以Decoding Topologically Associating Domains with Ultra-low resolution Hi-C Data by Graph Structural Entropy为题于8月15日在线发表于《自然-通讯》(Nature Communications)上。

该方法主要应用低分辨率的Hi-C预测高分辨率的染色质拓扑结构域(TAD)。基于李昂生团队之前发展的结构信息熵理论，该研究创造性地把Hi-C数据看作是一个相互连接的网络，开发了deDoc算法。新算法显著区别于当前的其他方法，可以直接使用原始测序数据而不需要归一化。正确的归一化方法对其他的软件来说十分关键，不恰当的归一化往往会得到糟糕甚至是错误的结果。此外，新算法对数据总量的依赖非常低。测试发现，即使只使用低至十个单细胞的Hi-C数据聚合，也可以很清晰地鉴定出类似拓扑结构域的结构。由于deDoc的上述两个重要特征，使得deDoc可以成为在大人群队列中进行高精度三维基因组研究的重要工具。

张治华团队一直致力于染色质三维结构研究，此前研究开发了精确预测染色质相互作用的算法CISD_loop，该算法利用真核生物核小体在基因组上排布的不均匀性，从核小体的排布方式上来推断高精度的染色体相互作用位点，进一步引入低分辨率的HI-C数据来预测染色质的相互作用。通过欠抽样实验，发现只要极低分辨率的Hi-C数据，CISD_loop就可以高分辨率预测染色质的相互作用。

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转发