北京生命科学研究院成功构建环形RNA大数据整合挖掘和分析平台

作者：writer 来源：中国科学院

本文原地址：https://www.iikx.com/news/progress/9385.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！

4月28日，中国科学院北京生命科学研究院赵方庆团队在国际期刊*Genome Biology*发表题为*CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes*的研究论文。该研究基于现有的海量转录组数据，采用多维数据智能整合分析手段，成功解析了跨物种、多组织、大样本的环形转录本表达特征和进化规律，为探索真核生物复杂多变的环形RNA全貌和产生机制提供了强有力的数据支持。

近年来，环形RNA作为一类新型的内源性非编码RNA在生物系统调控和疾病发展过程中的作用不断被发现和扩展。高通量测序技术的快速发展和广泛应用，更是将环形RNA研究带入了大数据时代，使之迅速成为RNA研究领域的热点之一。环形RNA转录组数据的大量积累，给研究人员带来了新的机遇和挑战：如何从转录组数据海洋中高效筛选和获取具有重要生物学功能的环形RNA分子。物种信息的日益丰富为全面解读复杂的环形RNA转录调控过程打开新的突破口：基于多物种的进化保守性分析将有助于筛选出具有潜在功能的环形转录本，而多组学数据的整合分析则可以从不同层次解析环形RNA的表达调控过程。

研究人员通过整合自有及公共转录组数据，获得覆盖6个物种（人、猴、小鼠、大鼠、猪和鸡）的19个组织类型，共计1070个转录组数据集，构建了目前覆盖物种最广、数据最齐全的环形RNA整合数据资源平台circAltas (http://circatlas.biols.ac.cn)。该平台收录超过100万个高质量的环形RNA分子，其中>80%具有全长转录本序列。此外，通过整合功能组学数据和注释信息，为环形RNA数据挖掘和功能研究提供了重要的数据资源和技术保障。

在上述数据的基础上，他们进一步提出了新的保守环形RNA识别方法和保守性多层次评估机制。通过结合全局比对和反向剪接位点的局部比对特征，筛选出超过12万保守的环形RNA，并进一步结合物种间、组织间和个体间的表达一致性对其保守性进行打分（Multiple Conservation Score），直观反映出环形转录本在不同层次的保守性和进化规律，对功能环形RNA分子的筛选具有重要意义。此外，研究人员还结合保守性和表达量信息，对收录的环形RNA进行重新命名，并提供多个环形RNA数据库间的名称查询和转换功能，厘清了环形RNA领域存在的命名混乱问题。此外，利用重建的环形RNA全长序列，该团队首次通过大规模分析其可能的ORF和IRES序列，去预测其翻译成蛋白质的潜力。进一步结合CLIP等多组学数据，构建环形RNA和mRNA、miRN

A及RBP的表达调控网络，并结合网络中众多调控元件的注释信息对环形RNA的功能进行预测。该研究为环形RNA的功能挖掘和注释提供了重要的分析工具。

该工作由赵方庆课题组的博士研究生吴婉莹和助理研究员冀培丰完成，并获得了国家自然科学基金委、科技部重点研发计划及中科院的经费支持。赵方庆团队在前期的工作中建立了环形RNA识别和质控、转录本组装、可变剪接识别及定量等一系列方法和工具，相关研究发表在*Genome Biology* (2015, 2020)、*Nature Communications* (2016，2020)、*Briefings in Bioinformatics* (2017)、*Trends in Genetics* (2018)、*Genome Medicine* (2019)、*Cell Reports* (2019)和*Bioinformatics* (2020)。这些研究丰富了人们对环形RNA的表达和功能的认识，为深入了解这一崭新类型的非编码RNA分子奠定了方法学基础。

论文链接

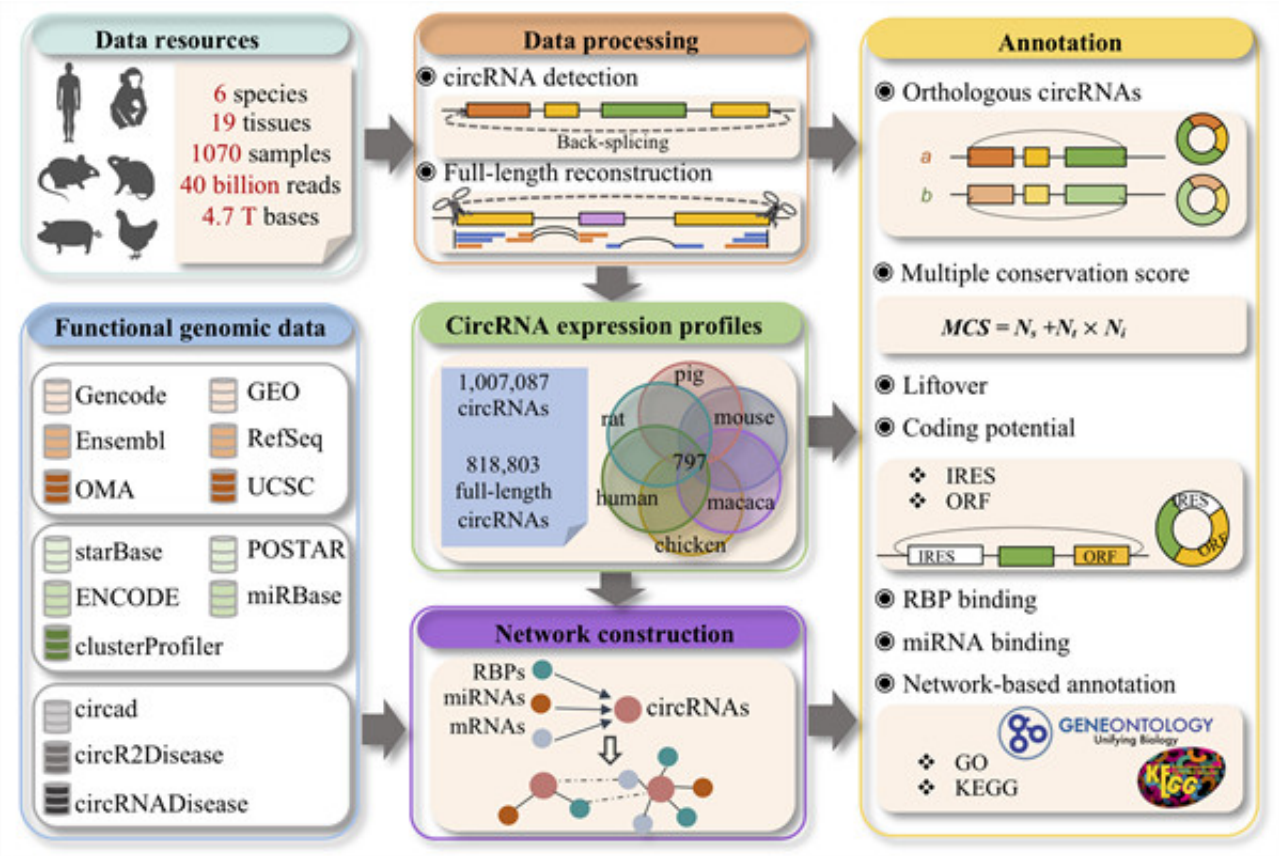


图1.环形RNA整合数据资源和挖掘平台——circAtlas

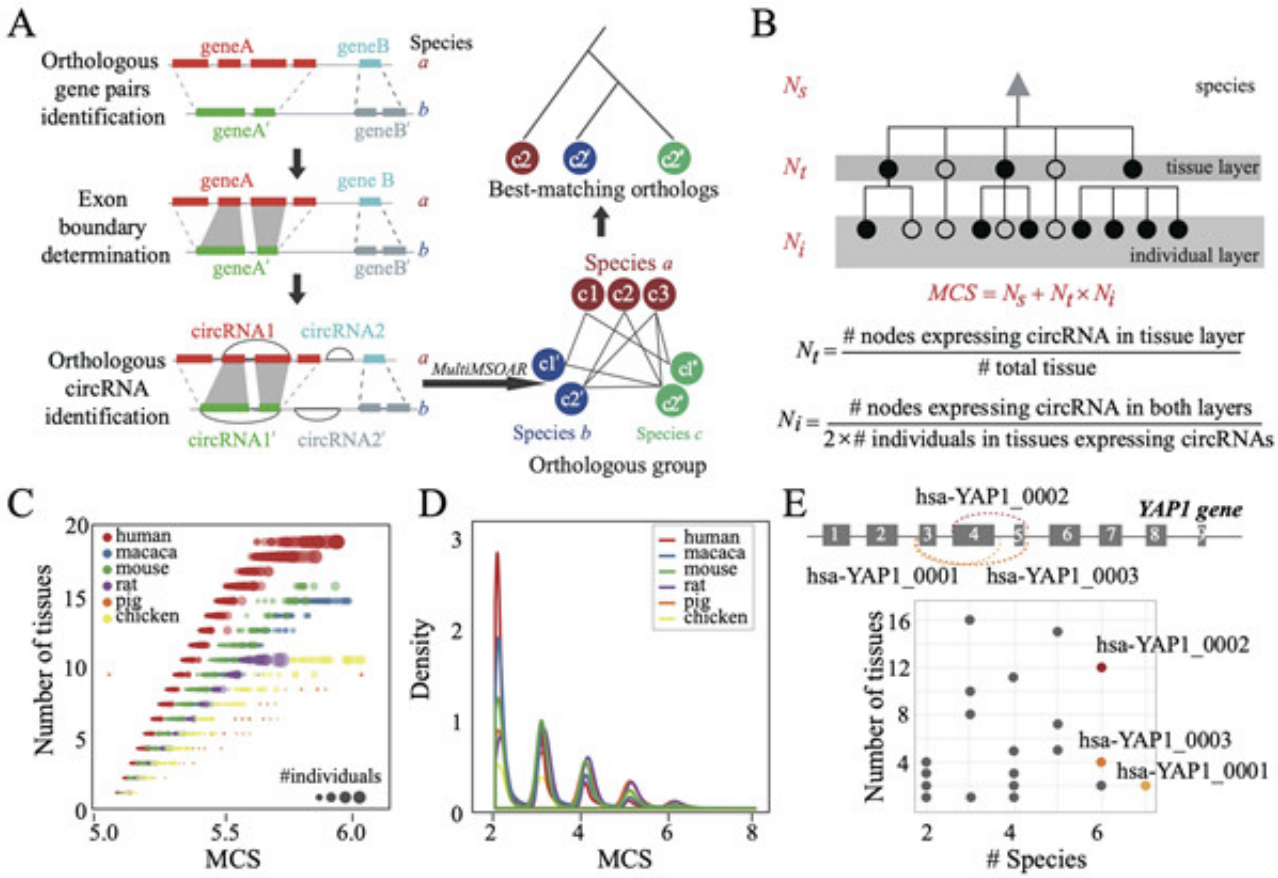


图2. 保守性环形RNA的识别和打分策略

研究团队单位：北京生命科学研究院

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