
北京基因组所构建出多物种转录图谱综合数据库

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/16490.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！

随着高通量测序技术的发展，转录组测序（RNA-seq）已成为系统研究基因转录及转录后水平调控状态的常规方法，并在多个物种中得到广泛应用。海量转录组数据以前所未有的速度产生，以数据驱动为导向的大规模数据整合、挖掘与解析面临挑战。为更充分展现转录组数据蕴含的丰富信息，服务生物学基础研究领域需求，构建标准化数据分析流程和结构化元信息审编模型，建立面向多物种基因转录特征的数据集成与管理资源十分必要。

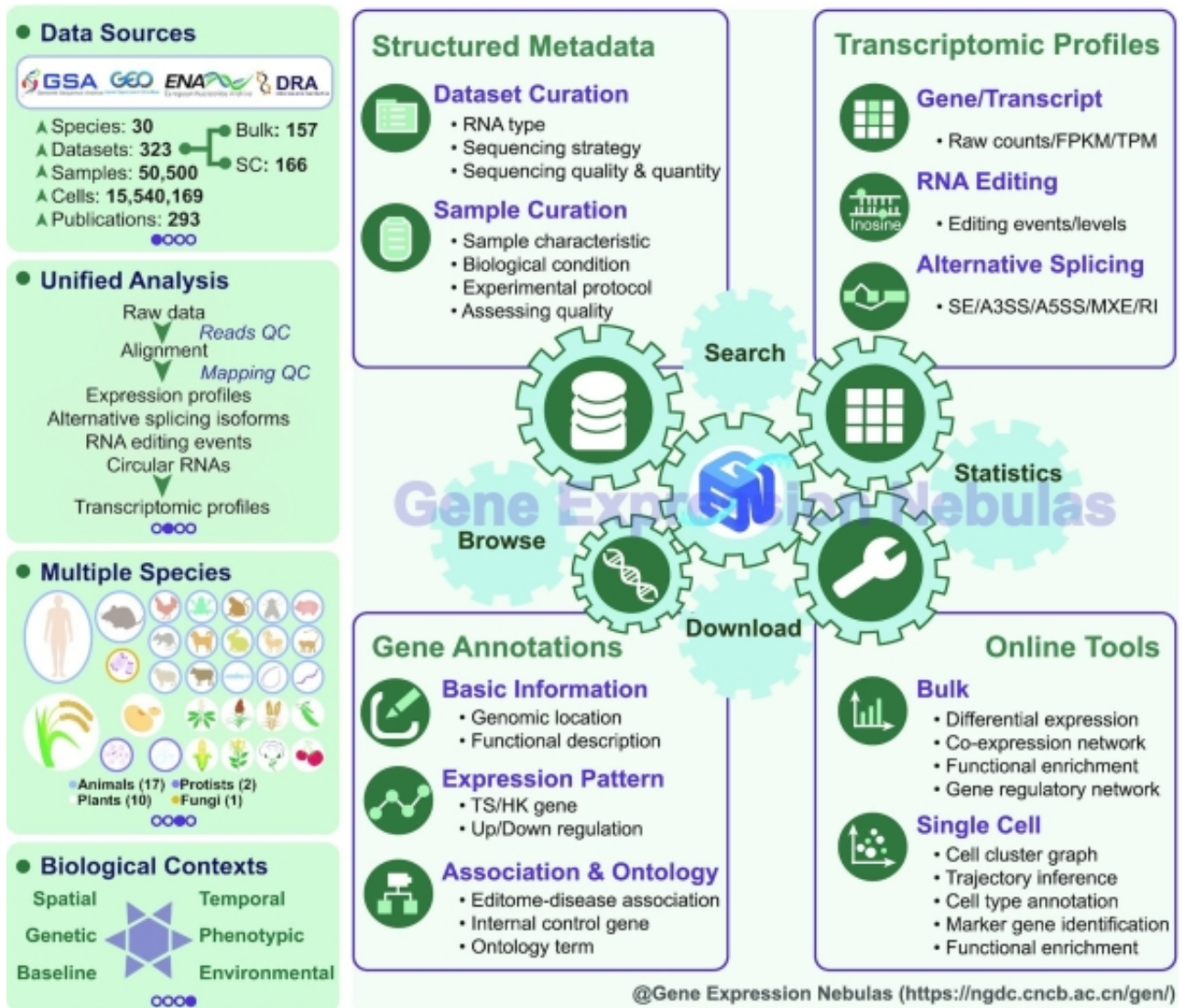
近日，中国科学院北京基因组研究所（国家生物信息中心）国家基因组科学数据中心在Nucleic Acids Research上，发表了题为Gene Expression Nebulas (GEN): a comprehensive data portal integrating transcriptomic profiles across multiple species at both bulk and single-cell levels的论文，建立了基于常规转录组测序（Bulk RNA-seq）和单细胞转录组测序（scRNA-seq）数据挖掘解析的多物种转录图谱整合型数据库[Gene Expression Nebulas \(GEN\)](#)。GEN应用结构化审编模型和标准化数据处理流程，对组织和细胞水平转录组测序数据集进行统一分析，实现了多物种多层次转录调控水平信息的系统整合。目前，GEN整合了323个高质量转录组数据集，涵盖30个物种的50,500个样本和15,540,169个细胞，提供基准参考、遗传、表型、环境、时间、空间六类生物学场景下的转录图谱，为生物学领域科研人员剖析基因遗传调控结构和功能机制提供基础资源。

基于严格的数据质控标准，GEN审编来自GSA、GEO、ENA和DRA数据库的高质量原始转录组测序数据和详细元数据信息，并利用自主搭建的标准化流程分析处理相应数据，为用户提供包括基因/转录本表达、环形RNA表达、RNA选择性剪接和RNA编辑四个层面的转录图谱。同时，GEN为30个物种的1,191,846个基因提供丰富的注释信息，包括基本注释（如基因组位置、生物类型、功能描述），以及基于基因表达数据的定量（不同实验条件下的表达水平）和定性（差异表达所处的生物学场景）的增值注释。此外，为方便下游个性化分析，GEN还为用户提供表达谱数据分析及可视化的在线及离线工具，包括基于Bulk RNA-seq表达谱的差异表达分析、加权基因共表达网络分析、功能富集分析和基因调控网络推断，以及基于scRNA-seq表达谱的质量控制、数据标准化、缩放和回归、降维、基于图的聚类、细胞簇标记基因识别、细胞标记、细胞轨迹推断和细胞类型注释等多项分析功能。

GEN对用户免费开放，具备友好的浏览、检索与可视化功能，方便用户探索多生物学场景下基因/转录本的表达及转录特征。研究团队将持续维护并定期更新GEN，以不断整合更多物种的转录图谱数据资源和集成更丰富的数据分析功能。

研究工作得到中科院战略性先导科技专项、国家重点研发计划、中科院青年创新促进会等的资助。

论文链接



GEN数据库内容和功能模块概览

研究团队单位：北京基因组研究所

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