北京基因组所开发长非编码RNA数据库LncBook

作者：writer 来源：中国科学院

本文原地址：https://www.iikx.com/news/progress/2651.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！

北京基因组所开发长非编码RNA数据库LncBook。近日，由中国科学院北京基因组研究所开发的人类长非编码RNA(long non-coding RNA, lncRNA)数据库LncBook正式上线。该项研究成果以LncBook: a curated knowledgebase of human long non-coding RNAs 为题在国际学术期刊《核酸研究》(Nucleic Acids Research)在线发表。

近年来，有关lncRNA的研究一直是国际热点，研究表明，lncRNA在多种生物过程中发挥了重要作用并与疾病的发生密切相关，但lncRNA的注释信息及质量仍远远落后于蛋白编码基因。LncBook数据库不仅提供了丰富的高质量人类lncRNA数据集，还进行了大规模的多组学数据分析，以及系统的功能与疾病注释，为功能实验研究及生物信息学分析提供了大量可用信息与数据。

基于严格的审编标准，LncBook整合已有lncRNA数据并鉴定新的lncRNA，共获得270,044个lncRNA转录本。在此基础上，LncBook在lncRNA表达、甲基化、变异、miRNA-lncRNA相互作用这些多组学层面进行大规模深层次的数据分析。在表达层面，绘制lncRNA在32或者53种正常人类组织中的表达图谱，并鉴定出49,115个高度组织特异(tissue-specific)和819个持家(housekeeping)lncRNA;在甲基化层面，构建lncRNA在9种癌症中正常与癌症状态下的promoter及body区的甲基化图谱;在变异层面，基于dbSNP数据库SNP位点注释lncRNA区域92,725,757个SNP最小等位频率(基于千人基因组数据)、ClinVar与COSMIC疾病关联信息;预测了128,392,451个lncRNA-miRNA的相互作用条目。上述结果以图或表的形式展示在LncBook数据库中，相关信息均可免费下载。基于以上数据，LncBook还预测了97,998个潜在的疾病关联lncRNA。此外，在LncRNAWiki的基础上LncBook对1,867个文献报道lncRNA进行了系统的功能与疾病信息注释。

LncBook作为重要的lncRNA资源库，提供了目前为止数据量最为丰富的人类lncRNA数据。作为LncRNAWiki数据库的补充，LncBook具备友好的查询、浏览与可视化功能。用户可通过ID/symbol、功能、疾病名称等检索lncRNA信息，浏览指定lncRNA的多组学信息，并通过ftp下载所有相关注释信息与分析结果。此外，LncBook还提供了可用于lncRNA序列比对、分类、编码能力预测等研究的工具，方便在线分析。

该研究与沙特阿卜杜拉国王科技大学(King Abdullah University of Science & Technology，KAUST)教授Vladimir Bajic合作开展。研究得到中科院战略性先导科技专项、中科院国际伙伴计划、中科院“十三五”信息化专项等的资助。

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