北京基因组所关于原始数据管理体系（GSA Family）研发取得进展

作者：writer 来源：中国科学院

本文原地址：https://www.iikx.com/news/progress/15244.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！

8月14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（CNCB-NGDC）在*Genomics, Proteomics Bioinformatics*上，在线发表研究论文*The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types*。GSA数据库体系接受全世界科研工作者的数据提交，汇交和管理各类型的数据，并对所有公开可用数据提供免费开放访问，支撑生命科学研究。

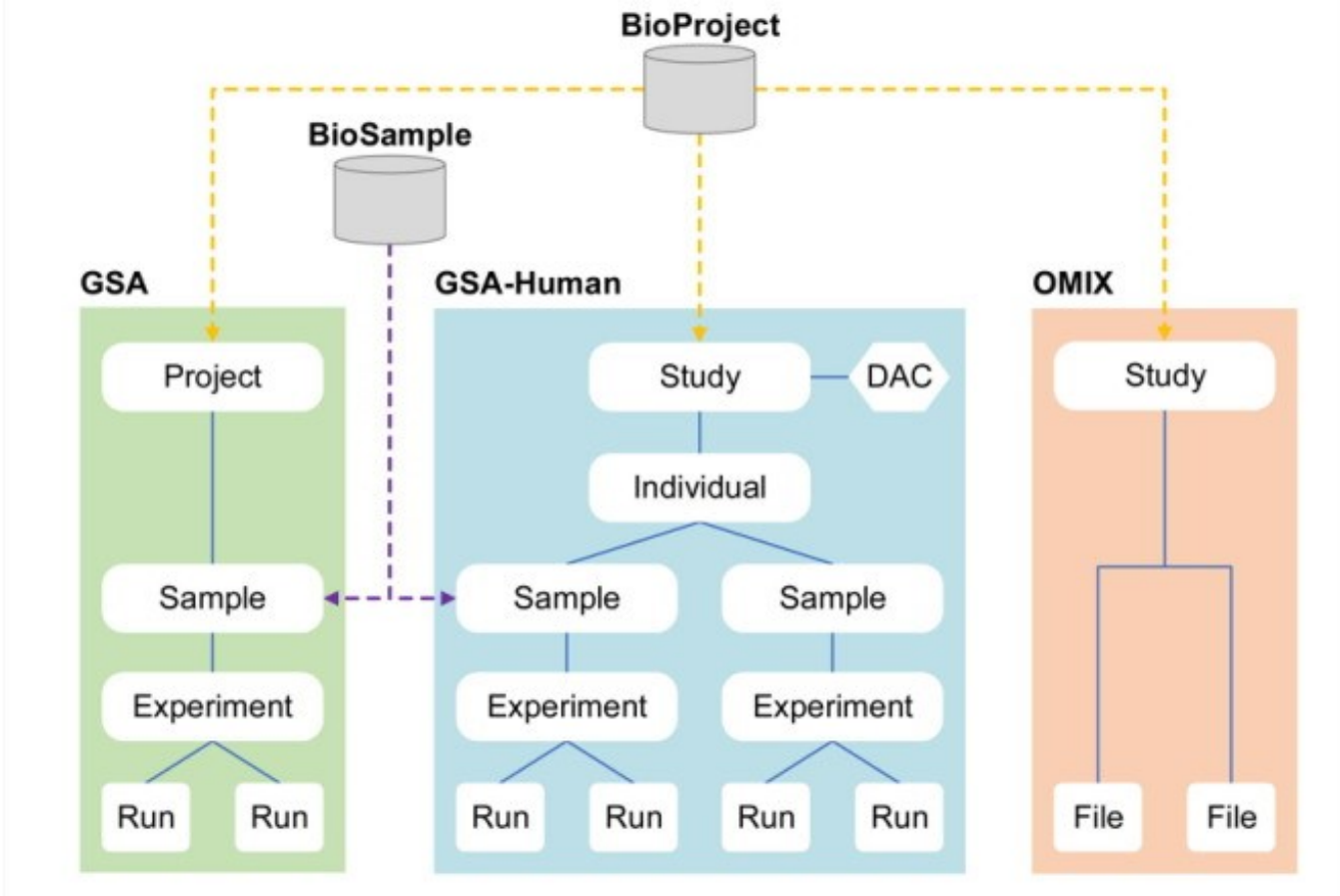
组学原始数据归档库（GSA）是生命组学原始测序数据管理的公益性数据库，旨在推动全球生命组学数据的共享与应用。近年来，随着组学数据的爆炸性增长和数据类型的多样化，以及人类遗传资源数据管理的特殊需求，CNCB-NGDC对GSA数据库进行了更新和扩展，形成了GSA数据库体系，包括GSA、GSA-Human和OMIX。

GSA数据库与2017发布的版本相比，在数据模型、系统功能和数据提交方式等方面进行了更新和功能提升；GSA-Human是存储人类遗传资源数据的数据库，可实现人类遗传资源数据的受控访问，保障人类遗传资源数据的安全性；OMIX数据库存储非原始测序数据，如环境组、表型组、代谢组等，作为上述两种数据资源库的重要补充，有效地解决了用户提交除原始测序数据外的其他类型数据的需求。

截至2021年8月14日，GSA和GSA-Human已收集的数据量达9.5 PB，OMIX上线不久数据量已达1.6 TB。GSA数据库体系已为全球111个国家/地区的用户提供数据服务，平均每天的数据下载量达4 TB，已成为Elsevier、Wiley、Taylor Francis、Cell及Springer Nature出版集团指定的核酸数据归档库，并获得领域内国内外主流期刊的认可。

研究工作得到国家重点研发计划、中科院战略性先导科技专项、中科院信息化专项等的支持，GSA归档数据使用的计算机硬件设施得到国家财政部修缮购置专项的支持。

论文链接



GSA Family数据模型

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转发