

面向国家重大需求 打造国际一流研究所 ——庆祝中国科学院北京基因组研究所成立 15 周年

薛勇彪^{1,2}

1. 中国科学院北京基因组研究所, 北京 100101

2. 中国科学院遗传与发育生物学研究所, 北京 100101

DNA 是细胞核中携带生物生长指令的遗传物质。1953 年, 沃森和克里克提出 DNA 双螺旋结构模型, 推动了遗传密码的复制、转录、转译等一系列生命功能的研究, 也由此揭开了分子生物学的序幕。DNA 上的核苷酸序列最终决定了生命的遗传信息, 而解析 DNA 序列则是研究生物生长发育、环境适应和疾病发生机理的核心基础。

自 20 世纪 80 年代第一台 DNA 测序仪诞生以来, 基因组科学迎来了发展的新时期。伴随着基因组学技术的快速发展和生物信息数据的爆炸性增长, 基因组科学沿着基因组、表观组和 RNA 组等研究轨迹迅速发展起来, 一系列里程碑式的研究成果引领了生命科学领域的发展。中国是加入人类基因组计划唯一的发展中国家, 并展开了相当规模的基因组研究, 在各方的大力支持下取得了前沿地位。中国科学院北京基因组研究所(以下简称北京基因组所)的成立是人类基因组早期研究在中国的缩影。

北京基因组所成立于 2003 年 11 月, 前身为“中国科学院遗传研究所人类基因组研究中心”。成立至今, 伴随着基因组科学的整体发展, 北京基因组所通过参与国际基因组计划, 建立了具有竞争力的基因组科学、技术和数据挖掘等人才队伍与平台, 在一些关键领域取得了重要突破, 做出了具有影响力的研究工作, 积累了一定数量的数据资源和大数据管理经验, 走过了 15 年不平凡的历程。

以基因组科学核心发展轨迹和未来发展方向为

基础, 北京基因组所明确研究定位、凝练科学目标, 形成新时期发展定位: “面向我国人口健康和社会可持续发展的重大战略需求与生命科学前沿, 重点研究基因组结构、变异、功能及其演化规律, 加强基因组学与其他学科的交叉融合, 发展基因组学的新理论、新方法和新技术, 成为基因组学原始创新研究基地、创新人才培养基地和卓越科学中心, 为保障人类社会的健康发展做出重大贡献”。根据学科发展前沿和国家重大需求, 北京基因组所组建了“基因组科学与信息”和“精准基因组医学”两个院重点实验室及生命与健康大数据中心, 并形成了精准医学、大数据中心和公共安全 3 大战略布局。

在基因组结构解析方面, 北京基因组所围绕重要资源生物, 对大型、复杂基因组进行结构解析, 实现了相对完整准确的基因组注释, 参与“人类基因组计划”和“国际人类单体型图计划”, 为中国基因组学研究的发展奠定了基础。完成超级杂交水稻基因组计划, 为水稻生物学研究与分子育种实践开辟了广阔的发展空间。与沙特合作开展了一系列重要经济作物的复杂基因组解析研究, 成功绘制出高质量椰枣基因组图谱, 与荷兰等国家合作发现了一系列人类外貌表型基因, 提高了人类外貌分子表型刻画水平, 推进了该技术在法医学领域的应用发展。本期专刊中, 于军研究员撰写的《实现“终极版”核苷酸测序仪的技术要素》一文对核苷酸测序技术的发展进行了综述, 阐述了生命科学从宏观到微观

收稿日期: 2018-11-06

作者简介: 薛勇彪, 博士, 研究员, 中国科学院北京基因组研究所所长。E-mail: ybxue@big.ac.cn

DOI: 10.16288/j.ycz.18-305

网络出版时间: 2018/11/6 17:30:36

URI: <http://kns.cnki.net/kcms/detail/11.1913.R.20181106.1730.008.html>

需要进行的努力和目标,探讨了实验技术平台的发展和需求,系统介绍了第四代测序仪的可能技术要素,以及应用需求和新研究范式^[1]。胡松年团队撰写的《复杂基因组测序技术研究进展》一文从基因组测序实验方法、测序技术平台、组装算法与策略 3 个方面,介绍了复杂基因组测序组装相关的技术与方法,并结合实例介绍了复杂基因组测序的技术解决途径和发展历程,为制订合适的复杂基因组测序策略提供了参考^[2]。刘凡团队撰写的《EDARV370A 对新疆维吾尔族人群面部及耳朵形态的效应》一文对 715 例亚欧混合人群(维吾尔族)进行了外异蛋白受体基因关键错位突变 EDARV370A 与一系列面部及耳朵形态表型的关联分析,发现 EDARV370A 与 21 个面部形态的定量表型、下巴类型以及 4 个耳朵形态分类表型显著相关^[3]。

在基因组演化与变异方面,北京基因组所牵头国家自然科学基金委“微进化过程中的多基因作用机制”重大研究计划,将进化生物学和群体遗传学理论运用到肿瘤异质性形成与迁移的研究中,发现肿瘤细胞中遗传多样性水平远远大于预期,在基因组水平发现海量的遗传变异“淹没”了肿瘤治疗之效,揭示原位癌获得快速生长能力之前,肿瘤细胞就发生了转移。牵头中国科学院“中国人群精准医学研究计划”,建立了千人规模的健康人群队列,构建了中国人参比基因组及中国人群的高精度遗传变异图谱。牵头国家重点研发计划精准医学重点专项项目“精准医学大数据处理和利用的标准化技术体系建设”,参与编写发布《遗传变异分类标准与指南》,建立了微生物组信息平台,完成了医疗数据的安全分类分级,初步建成了临床数据集成、质控、存储和访问的标准化技术体系。开发了大样本全基因组数据的群体遗传学分析新方法,可用于检测群体的增长模式,有效推断一万年以内的群体大小变化的相关参数。本期专刊中,肖景发团队撰写的《中国人群参考基因组及遗传变异图谱资源库》一文介绍了基于已有的中国人群基因组数据建立的中国人群动态基因组数据库和中国人群全基因组序列的基因组变异数据库特点和未来发展与应用前景^[4]。陈华团队撰写的《BIG-Annotator: 基因组测序数据高效功能注释及其在遗传诊断中的应用》一文介绍

了研究组新开发的对全基因组单核苷酸多态位点进行功能注释的软件——BIG-Annotator,该软件提供多线程运行模式,运行更为高效,或比现有的同类软件流程提速 10 多倍,适用于人群队列研究、大样本全基因组关联分析等数据量庞大、时效性要求高的分析需求^[5]。

在基因组学功能研究方面,北京基因组所利用遗传学和基因组学分析方法,证明表观遗传信息能够完整的遗传到后代中,系列研究引领了国际表观基因组跨代遗传和重编程的理论研究。在白血病发生机制、基因组特征解析及新治疗策略研究上取得系列进展。建立了 RNA 甲基化技术,拓展了 RNA 甲基化表观转录组学研究新领域。对前列腺癌病人组织样本的 DNA 甲基化特征进行研究,筛选出特异性和灵敏性都很高的甲基化位点并且可以在尿液中进行无创检测,研究成果实现转移转化。本期专刊中,刘江团队撰写的《动物早期胚胎发育中染色质结构的继承和重编程》一文介绍了近年来研究染色质开放性和染色质三维高级结构的实验分析技术手段,染色质结构在动物早期胚胎发育过程中的变化规律及其在早期胚胎发育中的作用,染色质结构与其他表观遗传信息关系方面的重要研究进展和存在的科学问题,为表观遗传调控早期胚胎发育研究提供了参考^[6]。杨运桂团队撰写的《RNA 甲基化修饰调控和规律》一文总结了近年来作者所在实验室与合作团队及国内外同行在 RNA 甲基化表观转录组学研究中取得的主要前沿进展,这些研究成果证明 RNA 甲基化修饰类似于 DNA 甲基化,具有可逆性^[7]。王前飞团队撰写的《白血病的精准基因组医学研究与转化应用》一文介绍了北京基因组所近年来在白血病研究中取得的一系列重要进展与转化应用,对白血病的发病机制、新抑癌基因的挖掘、白血病克隆演化以及低剂量化疗方案临床疗效等方面的研究进行了综述,为白血病诊断及治疗方法的改进提供了新契机^[8]。方向东团队撰写的《KLF1 和 KLF9 对 K562 细胞红系分化的协同调控作用》一文介绍了 KLF 家族成员 KLF1 和 KLF9 之间存在相互作用,协同促进 K562 细胞红系分化和 β -珠蛋白的表达,并探讨了二者对红系分化协同调控的可能机制^[9]。郭彩霞团队撰写的《RAD51 调控 REV1 参与 DNA

双链断裂修复》一文探讨了跨损伤 DNA 聚合酶 REV1 互作蛋白 RAD51 和 RAD51C 在其参与的同源重组修复通路中的调控作用,发现 REV1 和 RAD51 在 HR 通路中存在彼此相互调控的关系^[10]。慈维敏团队撰写的《肿瘤突变特征与病理分型的关联研究》一文通过整合癌症基因组数据库中肾癌、肺癌和食管癌的外显子测序数据,分别对 3 种肿瘤通过肿瘤基因突变特征进行肿瘤病理分型聚类 and 预测,揭示突变特征作为新型分子标记物,对肿瘤的病理分型、诊断,尤其是早诊具有参考意义^[11]。

生命与健康大数据是现代生命科学产生新知识、提出新假说、开发新应用的引擎,是我国建设世界科技强国的战略资源。长期以来,我国生物大数据流失严重,数据安全问题日益严峻;缺乏数据管理共享机制,严重制约我国生物大数据的转化应用。北京基因组所针对我国当前和未来生命与健康大数据发展与应用的瓶颈问题,面向国家“大数据”战略和生物信息科学前沿,布局生命与健康大数据中心,以存好、管好、用好我国生命与健康大数据,建立国家生物信息中心为目标,已开发数十个专业型数据库,建成中国首个获国际期刊认可的组学原始数据归档系统 GSA,实现我国生命组学大数据资源的开放共享、自主管理与高效利用,得到国内外广泛认可。本期专刊中,鲍一明团队撰写的《国家级生物大数据中心展望》一文以美国 NCBI 为例介绍了国际生物大数据中心的发展历程及现状,阐明了我国建立国家级生物大数据中心的重要性、迫切性、当前历史机遇和发展前景^[12]。章张团队撰写的《生命与健康大数据中心资源》一文介绍了北京基因组所生命与健康大数据管理平台和多组学数据资源体系,该平台体系可向全球提供大数据汇交、整合与共享服务^[13]。赵文明团队撰写的《GSA: 组学原始数据归档库》一文介绍了北京基因组所生命与健康大数据中心研发的组学原始数据归档库 GSA,该数据库上线两年来已获得 30 多个国内外期刊认可,缓解了当前中国生命组学数据汇交、存储与共享难的问题,为我国国家生物信息中心建设奠定了基础^[14]。

过去 15 年,北京基因组所砥砺奋进,以国立科研机构的责任和担当为使命,面向国家重大战略需求,提出科研命题、解决科研问题、产出科研成果,

打造核心竞争力。下一个发展阶段,北京基因组所将不忘初心,继续前行,聚力聚焦国家战略,坚持面向未来,高起点、高标准,建设国际领先的创新研究团队,集学科、人才、项目、平台建设于一体,组织开展多学科协同创新,在创新道路上革故鼎新、勇立潮头,实现北京基因组所发展的新跨越。

致谢

特别感谢北京基因组研究所综合办公室周梦菱对此文的贡献。

参考文献(References):

- [1] Yu J. Possible technical elements for the ultimate nucleotide sequencers. *Hereditas (Beijing)*, 2018, 40(11): 929–937.
于军. 实现“终极版”核苷酸测序仪的技术要素. 遗传, 2018, 40(11): 929–937. [DOI]
- [2] Gao SH, Yu Hy, Wu Sy, Wang S, Geng JN, Luo YF, Hu SN. Advances of sequencing and assembling technologies for complex genomes. *Hereditas (Beijing)*, 2018, 40(11): 944–963.
高胜寒, 禹海英, 吴双阳, 王森, 耿佳宁, 骆迎峰, 胡松年. 复杂基因组测序技术进展. 遗传, 2018, 40(11): 944–963. [DOI]
- [3] Li Y, Zhao WY, Li D, Tao XM, Xiong ZY, Liu J, Zhang W, Liu HB, Ji AQ, Tang, Liu F, Li CX. The effect of *EDARV370A* on facial and ear morphologies in uyghur population. *Hereditas (Beijing)*, 2018, 40(11): 1024–1032.
李伟, 赵雯婷, 李丹, 陶现明, 熊子义, 刘京, 张微, 刘海渤, 季安全, 唐鲲, 刘凡, 李彩霞. *EDARV370A* 对新疆维吾尔族人群面部及耳朵形态的效应. 遗传, 2018, 40(11): 1024–1032. [DOI]
- [4] Song SH, Teng XF, Xiao JF. Database resources of the reference genome and genetic variation maps for the Chinese population. *Hereditas (Beijing)*, 2018, 40(11): 1048–1054.
宋述慧, 滕徐菲, 肖景发. 中国人群参考基因组及遗传变异图谱资源库. 遗传, 2018, 40(11): 1048–1054. [DOI]
- [5] HuangY, Liu Q, Chi LJ, Shi CHM, Wu Z, Hu M, Shi H, Chen H. Application of BIG-Annotator in the genome sequencing data functional annotation and genetic diagnosis. *Hereditas (Beijing)*, 2018, 40(11): 1015–1023.
黄莹, 刘琪, 池连江, 石承民, 吴祯, 胡敏, 石宏, 陈华. BIG-Annotator: 基因组测序数据高效功能注释及其在

- 遗传诊断中的应用. 遗传, 2018, 40(11): 1015–1023. [DOI]
- [6] Ke YW, Liu J. The inheritance and reprogramming of chromatin structure in early animal embryos. *Hereditas (Beijing)*, 2018, 40(11): 977–987.
- 柯玉文, 刘江. 动物早期胚胎发育中染色质结构的继承和重编程. 遗传, 2018, 40(11): 977–987. [DOI]
- [7] Yang Y, Chen YS, Sun BF, Yang YG. RNA methylation: regulations and mechanisms. *Hereditas (Beijing)*, 2018, 40(11): 964–976.
- 杨莹, 陈宇晟, 孙宝发, 杨运桂. RNA 甲基化修饰调控和规律. 遗传, 2018, 40(11): 964–976. [DOI]
- [8] Yu XX, Chen AL, Li YY, Liu D, Wang QF. Precision genomic and translational medicine for acute myeloid leukemia. *Hereditas (Beijing)*, 2018, 40(11): 988–997.
- 于雪新, 陈艾莉, 李玥莹, 刘丹, 王前飞. 白血病的精准基因组医学研究与转化应用. 遗传, 2018, 40(11): 988–997. [DOI]
- [9] Ren L, Xiao RD, Zhang Q, Lou XM, Zhang ZJ, Fang XD. Synergistic regulation of the erythroid differentiation of K562 cells by *KLF1* and *KLF9*. *Hereditas (Beijing)*, 2018, 40(11): 998–1006.
- 任岚, 肖茹丹, 张倩, 郑佳文, 张昭军, 方向东. *KLF1* 和 *KLF9* 对 K562 细胞红系分化的协同调控作用. 遗传, 2018, 40(11): 998–1006. [DOI]
- [10] Huang M, Yang YR, Sun XY, Zhang T, Guo CX. RAD51 regulates REV1 recruitment to DNA double-strand break sites. *Hereditas (Beijing)*, 2018, 40(11): 1007–1014.
- 黄敏, 杨业然, 孙晓艳, 张婷, 郭彩霞. RAD51 调控 REV1 参与 DNA 双链断裂修复. 遗传, 2018, 40(11): 1007–1014. [DOI]
- [11] Shi Y, Xu ZZ, Lu H, Ci WM. Correlation studies of distinct mutational signatures with common cancer pathological subtyping. *Hereditas (Beijing)*, 2018, 40(11): 1033–1038.
- 史悦, 许争争, 鲁欢, 慈维敏. 肿瘤突变特征与病理分型的关联研究. 遗传, 2018, 40(11): 1033–1038. [DOI]
- [12] Ma Yk, Bao YM. Prospects for national biological big data centers. *Hereditas (Beijing)*, 2018, 40(11): 938–943.
- 马英克, 鲍一明. 国家级生物大数据中心展望. 遗传, 2018, 40(11): 938–943. [DOI]
- [13] Zhang YS, Xia L, Sang J, Li M, Liu L, Li MW, Niu GY, Cao JB, Teng XF, Zhou Q, Zhang Z. The BIG Data Center's database resources. *Hereditas (Beijing)*, 2018, 40(11): 1039–1043.
- 张源笙, 夏琳, 桑健, 李漫, 刘琳, 李萌伟, 牛广艺, 曹佳宝, 滕徐菲, 周晴, 章张. 生命与健康大数据中心资源. 遗传, 2018, 40(11): 1039–1043. [DOI]
- [14] Zhang SS, Chen TT, Zhu JW, Zhou Q, Chen X, Wang YQ, Zhao WM. GSA: Genome Sequence Archive. *Hereditas (Beijing)*, 2018, 40(11): 1044–1047.
- 张思思, 陈婷婷, 朱军伟, 周晴, 陈旭, 王彦青, 赵文明. GSA: 组学原始数据归档库. 遗传, 2018, 40(11): 1044–1047. [DOI]