

RNA 组学: 后基因组时代的科学前沿

历史表明, 每当 DNA 研究取得重大突破后, 都会出现一个 RNA 研究的高潮. 1953 年 DNA 双螺旋结构的解析掀起了在 RNA 转录和翻译水平解读遗传信息的高潮, 导致 mRNA, tRNA 和 rRNA 的发现以及遗传密码和“中心法则”的建立. 1977 年分裂基因(split gene)的发现极大地促进了在 RNA 转录后加工水平解读遗传信息表达的过程及机制, 深入解析了 mRNA, snRNA, snoRNA 和 gRNA 的结构与功能, 提出了 RNA 剪接、编辑以及核酶(ribozyme)和“RNA 世界”等新的概念. 2001 年人类基因组计划完成, 宣告后基因组时代的开始, 同时也预示新一轮的 RNA 研究从非编码 RNA 基因(non-protein-coding RNA genes, ncRNA genes)的角度解读遗传信息的组成及其表达调控的研究高潮的到来.

人类基因组中编码蛋白质的基因数目不足 3 万个, 约占整个基因组序列的 2%, 远远低于人类基因组计划完成之前所预期的 10 万个基因, 而 98% 的基因组序列没有得到注释. 显然, 仅仅按照生物体编码蛋白质的基因数目的多少无法解释高等哺乳动物的复杂度. 科学家们认为高等生物拓宽其分子水平差异并不完全取决于基因的数目, 而主要依靠对基因表达的调控. 在高等哺乳动物中, 基因表达调控有两个主要策略: (1) 通过对基因组转录的 mRNA 前体中内含子和外显子的可变剪接加工等, 从一个基因可以产生大量的蛋白同源异形体, 有效地增加了蛋白组的复杂度. mRNA 可变剪接的调控有大量的 RNA 顺式元件和反式因子参加. (2) 占基因组序列 98% 的非蛋白质编码区实际上编码了大量的非编码 RNA. 除 rRNA, tRNA, snRNA, snoRNA 等非编码 RNA 外, 近年来发现的 miRNA, endo-siRNA 和 piRNA 等调控型的小分子非编码 RNA 在基因的转录和翻译、细胞分化和个体发育、遗传和表现遗传等生命活动中发挥重要组织和调控作用, 形成了细胞中高度复杂的 RNA 网络.

由此可见, 生命起源于原始的“RNA 世界”以后, 在现代生物细胞中仍存在大量的非编码 RNA. 各种生物非编码 RNA 的比例与生物的复杂度成正比, 表明在高等生物中, 有一个巨大且尚未被完全发现的现代“RNA 世界”. “RNA 组学”就是从基因组水平研究细胞中非编码 RNA 结构与功能的一门新的科学. “RNA 组学”是后基因组时代一个重要的科学前沿, 因为它有可能揭示一个全新的由 RNA 介导的遗传信息表达调控网络, 从而以不同于蛋白质编码基因的角度来注释和阐明人类基因组的结构与功能. 同时, 基于非编码 RNA 研究所获得的新发现, 将为人类疾病的研究和治疗提供新的技术和思路.

目前, 非编码 RNA 研究领域主要包括以下几个方面:

(1) 非编码 RNA 的系统识别与鉴定. 在现代隐蔽的 RNA 世界中, 已知的非编码 RNA 仅为冰山一角. 采用计算机 RNA 组学和实验 RNA 组学等方法, 系统地发现和注释各种模式生物中的非编码 RNA 基因并进一步阐明其生物学意义是 RNA 组学的首要任务.

(2) 细胞分化和发育中 miRNA 的结构与功能. miRNA 是一个巨大的小分子非编码 RNA 家族, 广泛存在于各种动植物甚至单细胞真核生物中. miRNA 在转录后水平调控蛋白质基因的表达谱来决定细胞分化、胚胎发育等一系列重要生命活动的进程及多样性. 目前在各种生物中已发现数千种 miRNA, 大部分 miRNA 的功能尚有待阐明.

(3) 表观遗传中的 RNA 调控. 许多非编码 RNA 都参与了基因组 DNA 转录水平的调控. 研究这些 RNA, 特别是内源 siRNA 参与的真核细胞核内异染色质的形成和基因组 DNA 修饰或加工过程, 有可能揭示表观遗传控制发生的原因及调控机制.

(4) RNA 与疾病发生. 许多新发现的小分子非编码 RNA 不仅是细胞增殖、分化和凋亡的重要调控因子, 而且与细胞的异常表型和人类重要疾病密切相关. 在许多肿瘤中可检测到特有 miRNA 基因的异常表达或 mRNA 异常可变剪接体. 一些动物病毒也编码用于逃逸宿主细胞免疫攻击的 miRNA. 比较分析正常生理和疾病发生过程中的非编码 RNA 的表达及其作用, 将从 RNA 调控的角度揭示疾病发生机制, 并为疾病诊断和治疗提供新的基因靶点和分子标记.

(5) mRNA 可变剪接的调控. 大量的剪接调控蛋白与特异的 RNA 顺式作用元件和 RNA 反式因子在 mRNA 可变剪接调控中起着重要作用. 发现和鉴定这些调控因子及其作用方式将阐明 RNA 调控蛋白质组复杂度的机制及其与细胞分化和个体发育的关系.

(6) 非编码 RNA 基因资源与 RNA 技术及其应用. 非编码 RNA 基因是新发现的遗传资源和新的生物技术制高点. 如 miRNA 和 RNA 干涉技术在干细胞维持、动植物品种选育及病害控制等方面有重要应用前景. miRNA 治疗干预作为人类重大疾病新的治疗技术, 正得到迅速的发展, miRNA 药物及靶点的研究前景十分广阔.

面对非编码 RNA 的挑战, 金由辛研究员、朱圣庚教授和我, 在 1998 年 12 月共同主持召开了“面向 21 世纪的 RNA 研究”第 109 次香山会议. 会议回顾了一个世纪以来的 RNA 研究成果及发展趋势, 指出在人类及其他物种基因组研究取得重大突破后, 21 世纪的 RNA 科学将有飞速的发展, 因此提议在我国尽快开展以 RNA 为终产物的基因组计划(即“RNA 组学”研究). 10 年来, 我国的 RNA 研究有了长足的进展, 在非编码 RNA 的各个研究领域涌现出一批优秀的学术人才. 为了介绍当前国内外非编码 RNA 研究成果及进展, 进一步促进“RNA 组学”在我国的发展, 《中国科学 C 辑: 生命科学》编辑部组织了这期“RNA 研究进展专辑”. 专辑包括 16 篇综述, 内容涉及 miRNA, siRNA, snoRNA, tRNA 等动植物小分子非编码 RNA、RNA 与细胞分化发育、RNA 与肿瘤发生及诊断治疗的靶点、核酶的结构与功能、遗传印记起源、RNA 基因的进化等多方面研究成果及发展趋势, 展示了一个宽广的 RNA 研究新领域. 撰文者有长期从事 RNA 研究的学者, 也有来自发育生物学、医学以及生物进化等方面的专家. 这也表明由于 RNA 调控在生命现象(细胞、发育、疾病)中的重要性, RNA 研究者也面临学科交叉和综合的挑战. 值得指出的是, 蛋白质编码基因和非蛋白质编码 RNA 基因是遗传信息整体中两个方面, 对这两类基因及其调控网络的交叉和相互作用的研究, 将进一步加深人们对生命本质的认识. 希望通过本专辑的出版, 不仅对非编码 RNA 研究有促进作用, 同时对生命科学的其他领域也有借鉴价值.

屈良鹤

中山大学生命科学学院, 广州 510275