



Junk DNA 的功能诠释

李明振 张 铭 明镇寰*

(浙江大学生命科学学院, 杭州 310012)

摘要 在庞大的基因组序列中数量占绝对优势的序列因为不编码蛋白质或 RNA 产物, 一直被人们称为 junk DNA. 事事讲究经济效率的生物在长期的进化中, 应该不会让大量无功能的“垃圾”堆积在充满活力的生命细胞中. 近年来的研究已揭示 junk DNA 具有重要的功能, 随着研究的深入, 一定会发现越来越多的 junk DNA 决非垃圾, 而是基因组的宝贵财富.

关键词 无功能 DNA, 重复序列, 保守的非基因序列

学科分类号 Q522

蛋白质在生物中发挥着各种生理功能, 所以那些在基因组中不编码蛋白质、也不编码 rRNA 或 tRNA 的 DNA 序列就被人们统称为无功能的 DNA, 即 junk DNA. 人类基因组计划酝酿的早期, 在人们对这些所谓的 junk DNA 的功能还缺乏认识的时候, 有人曾极力主张只测定那些编码蛋白质的序列, 以便更有效地获取基因组的绝大部分信息. 近年来随着研究的逐渐深入, junk DNA 的功能也日益被认识, 昔日的基因组“垃圾”已成为生物学家致力寻求的基因组的财富, 人类基因组全序列的测定也更显露出其当初决定的正确.

1 Junk DNA: 基因组内序列占绝对优势者

在人类基因组测序完成后, 人们发现, 在基因组 30 亿碱基对的序列中, 只有大约 1.5% 的序列用以编码蛋白质, 而大约占 98% 的基因组序列都属于 junk DNA 之列. 在这些序列中包括与基因表达有关的各种调控序列 (但所占比例很小)、基因的内含子、基因间的大量非编码序列. 其中后者所占比重最大, 它们都是一些重复序列, 是 junk DNA 的主要组成部分.

2 内含子: 作用奇特的酶的编码者

内含子存在于基因的初始转录物, 但在随后成熟加工中被除去, 属于 junk DNA 之列. 但在 20 世纪 80 年代中期就发现一种酵母的内含子具有编码一种奇特的、后来被称为归巢核酸内切酶的功能. 当其与另一种不具有该内含子的酵母细胞交配时,

可以在受体酵母细胞 DNA 上与供体细胞完全相同的部位精确地将该内含子插入. 至今已在酵母、藻类、病毒、高等植物的线粒体和叶绿体基因组的 I 类内含子中发现了几百种这类功能的酶^[1]. 它们识别 15 ~ 40 bp 的序列, 可以作为一种精确切割的工具在基因治疗中将治疗基因插入特定的位点. 有研究者已精确地将 2 个分别来自嗜热菌和藻类的天然归巢内切酶, 拼接成一个能识别特定的 22 碱基对序列的新酶. 运用计算机辅助蛋白质设计和突变技术已有可能产生无数这类酶的变种, 通过筛选而获得对某一特定 DNA 靶位点进行切割的酶.

类似地, II 类内含子编码的酶可以精确地识别 30 ~ 35 bp 的序列, 而且内含子转录的 RNA 在插入位点特异性上具有重要的作用, 从而可以通过对内含子 RNA 的修饰而将内含子重新定位. 目前已掌握对内含子 RNA 的序列进行修饰、将其重新靶定于一特定基因的规律, 运用这些规律编成的计算机程序可将特定的内含子在所选择的最佳位点插入感兴趣的基因.

编码蛋白内含子 (intein) 的 DNA 序列也被归入 junk DNA 之列. 至今已在 50 多种生物中发现了蛋白内含子. 蛋白内含子切除发生在蛋白质进行折叠的同时, 其机制与内含子的相似. 利用蛋白内含子编码序列的插入和剪接规律, 近年来已发展了一种特殊的实验技术, 可以使得蛋白质的纯化方法变

* 通讯联系人.

Tel: 0571-88273604, E-mail: zhming@mail.hz.zj.cn

收稿日期: 2004-03-22, 接受日期: 2004-04-30

得简便有效, 同时还可处理因具有毒性而难以用工程菌生产的一些蛋白质.

3 转座因子: 基因组进化的推动者

转座因子是具有能整合到基因组新位点能力的 DNA 序列, 它是基因组内重复序列的主要组成部分, 几乎占了哺乳动物基因组一半的序列, 而在某些植物中更可高达基因组的 90%. 转座因子可分为 DNA 转座因子、自主的逆转录转座子和非自主的逆转录转座子. 近年来的研究表明, 它们作为重组热点, 通过提供基因组洗牌机制和为新的转录调节元件、多腺苷酸化信号和蛋白质编码序列提供即用基序的来源, 与围绕它们的基因组其他序列相互作用, 极大地增加了生物的进化能力^[2].

Alu 元件属于非自主的逆转录转座子, 在人类基因组内的拷贝数可达 140 万个, 占基因组总序列的 10% 以上. 人类基因组中大约有 5% 的可以进行选择性剪接的外显子起源于 Alu 序列. 这些 Alu 起源的外显子形成遵循的机制, 总能使基因组中同时存在另一个没有 Alu 元件而其余序列都相同的编码区, 因而这一过程的发生并不扰乱基因组编码原有蛋白质的功能^[3]. 基因复制是生物体产生新基因的一种重要手段. Alu 元件作用的机制使得在一个基因复制后, 其中的一个拷贝保持其原有的功能而另一个能自由地进化, 结果增强了基因组的编码能力和调节的多面性而不损害其基因组的完整性, 从而在进化上发挥重要的作用.

重复元件可以看作基因组的备件场所, 储存着用以自然进化实验所需的各种备件.

4 端粒: 染色体稳定和复制过程完整的维护者

端粒 DNA 是真核细胞染色体末端的一个特殊结构, 由简单的 DNA 高度重复序列组成. 端粒 DNA 的主要功能是保护染色体不被核酸酶降解, 防止染色体相互融合, 解决 DNA 复制的末端隐缩以保证染色体的完全复制. 同时, 端粒又是基因调控的一个特殊位点, 可抑制位于端粒附近基因的转录活性. 正常细胞随着细胞分裂活动的进行, 端粒 DNA 逐渐缩短, 当缩短到一定程度时, 染色体结构被破坏, 细胞进入衰老期并以死亡而告终.

与端粒 DNA 类似的是中心粒 DNA, 在染色体的结构上起特定的作用. 中心粒 DNA 也是高度重复的非编码 DNA.

5 保守的非基因序列: 进化中比蛋白质编码序列更保守者

保守的非基因序列 (conserved non-genic sequences, CNGs) 在生物的进化中具有特殊的功能. 最近对人类和小鼠基因组序列的比较分析表明, 两个物种的基因组序列中有 5% 是保守的^[4]. 除了占基因组大约 1.5% 的编码蛋白质序列和少量的非编码 RNA (non-coding RNA, ncRNA) 序列外, 大多数就是这类 CNGs. 而且, 这类 CNGs 在哺乳动物中具有很大的保守性, 人 21 号染色体中 220 个 CNGs 的绝大多数, 都可在从猴子到鸭嘴兽的哺乳动物基因组中找到相应的序列^[5]. CNGs 序列的保守程度令人惊奇, 它们比典型的蛋白质编码序列和已知的 ncRNA 序列都要保守得多. 在 12 种哺乳动物中都存在的最保守的 CNGs, 在哺乳动物进化谱系中所呈现的核苷酸替代速率还不及蛋白质编码序列的一半.

估计在人类基因组中有 6 万个这样的 CNGs, 几乎为蛋白质编码基因的 2 倍. 这些 CNGs 可能的功能是什么? 可以想到的有两种: 一种是作为基因表达调控区, 因为 CNGs 具有蛋白质结合区的一些典型结构特点, 但其中很多序列看来又过于保守而不太可能成为调节蛋白质的结合位点; 另一种是作为染色体的结构成分, 但同样难于理解的是染色体的结构成分是否需要如此高程度的保守性? 所以, CNGs 的真正功能还有待于人们对包括哺乳动物以外的其他脊椎动物和无脊椎动物的进一步研究.

但有一点是毋庸置疑的, 这些 CNGs 必定不是真正的 junk DNA, 而是基因组中具有重要功能的序列, 因为序列的保守往往是与特定的功能密切相关的.

参 考 文 献

- 1 Wickelgren I. Spinning junk into gold. *Science*, 2003, **300** (5626): 1646 ~ 1649
- 2 Kazazian H H. Mobile elements: driver of genome evolution. *Science*, 2004, **303** (5664): 1626 ~ 1632
- 3 Makalowski W. Not junk after all. *Science*, 2003, **300** (5623): 1246 ~ 1247
- 4 Waterston R H, Lindblad-Toh K, Birney E, *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature*, 2002, **420** (6915): 520 ~ 562
- 5 Dermitzakis E L T, Reymond A, Scamuffa N, *et al.* Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). *Science*, 2003, **302** (5647): 1033 ~ 1035

Annotation for The Functions of Junk DNA

LI Ming-Zhen, ZHANG Ming, MING Zhen-Huan *

(College of Life Science, Zhejiang University, Hangzhou 310012, China)

Abstract The prevailing sequences in the massive genome sequences, which do not code for proteins or RNAs, have been named junk DNA. It is unreasonable for organisms, in which there is the principle of maximum economy, to accumulate nonfunctional garbage in vital living cells during their long-term evolution. Recent researches have demonstrated that so-called junk DNA possesses important functions and it is believed that more and more junk DNA will be proved to be not garbage but the treasures of the genome.

Key words nonfunctional DNA, repetitive sequences, conserved non-genic sequences

* Corresponding author. Tel: 86-571-88273604, E-mail: zhming@mail.hz.zj.cn

Received: March 22, 2004 Accepted: April 30, 2004

2004 年国际生物芯片技术论坛即将召开 2004 年 10 月 21 ~ 24 日, 北京中关村生命科学园

会议由清华大学、生物芯片北京国家工程研究中心、科技部、教育部、国家自然科学基金委员会等单位共同主办, 由生物芯片北京国家工程研究中心具体筹办。

会议议题将主要包括以下内容:

(1) DNA、蛋白质、细胞及组织微阵列芯片技术; (2) 微流体芯片及缩微芯片实验室技术; (3) 芯片药物筛选技术; (4) 生物信息学技术。

会议已邀请到三十余位国际上最具权威性的生物芯片专家来做大会特邀报告, 其中既有从事生物芯片前沿性探索研究的科研院所的著名教授, 也有从事生物芯片研发的知名商业公司的总裁或部门经理, 是国内学者和投资机构进行交流和学习的良机。

会议网站: <http://www.capitalbiochip.com/IFBT2004/>

联系人: 生物芯片北京国家工程研究中心, 任琛

地址: 北京市海淀区清华大学生物科学与技术系 301 室, 邮编: 100084

电话: 010-62772239, 13651155414 传真: 010-62773059 E-mail: ychen@capitalbiochip.com