

微型述评

非编码的 mRNA

罗文琴 袁建刚 强伯勤

(中国医学科学院基础医学研究所 医学分子生物学国家重点实验室, 北京 100005)
(中国协和医科大学基础医学院,

摘要 非编码的 mRNA 是近年发现的一类不含典型 ORF 的 mRNA. 目前已发现或克隆的这类基因主要有: *H 19* 基因, *XIST* 基因, *XLSIRT* 基因, *His-1* 基因, *bic* 基因, *rox 1* 和 *rox 2* 基因等. 它们与胚胎发育, 肿瘤发生及 X 染色体失活密切相关.

关键词 非编码的 mRNA, *H 19* 基因, *XIST* 基因, *His-1* 基因, *bic* 基因, *rox 1* 和 *rox 2* 基因

学科分类号 Q522

传统的 mRNA 作为细胞核与细胞质之间的信使, 其主要功能是指导蛋白质合成; 最近几年, 人们发现了一类非编码的 mRNA, 它们跟传统的 mRNA 一样进行剪接和 3' 端加尾, 但并不具有典型的开放性阅读框架 (ORF). 非编码 mRNA 的基因与假基因不同, 它们两端没有反向重复序列, 转录本具有较明确的功能. 目前发现的非编码 mRNA 分布广泛, 从果蝇到人都可见. 这些非编码 mRNA 的基因一般都位于染色体的重要功能位点, 与染色体印迹现象紧密偶联, 参与细胞的生长和分化、胚胎的发育、肿瘤的形成和抑制. 非编码 mRNA 的功能及其机制与近年发现的 3' UTR 的反式 (trans-acting) 作用相似, 因此有人提出“调控 RNA (riboregulator)”的概念. 以下就此类一些主要基因作简单介绍:

1 *H 19* 基因

H 19 基因位于 11p15, 父源印迹, 其相邻下游的 *IGF2* 基因为母源印迹, 两者的表达互相影响. *H 19* 基因对 *IGF2* 基因的表达有顺式抑制作用, 而在某些情况下 *IGF2* 基因的印迹活化也会导致 *H 19* 基因的失活. 但在一些肿瘤发生早期 *H 19* 基因印迹的活化并不依赖于

IGF2 基因的灭活. *H 19* 基因的印迹现象与其 5' 的甲基化及组蛋白的乙酰化有关. *H 19* 基因为一胚胎期特异基因, 其表达与 AFP 平行, 受 Raf 及 Rif 这两个上游因子的调控. 成熟的 *H 19* mRNA 长约 2.5 kb, 经过剪接及多聚腺苷酸加尾, 但没有典型的 ORF. *H 19* mRNA 主要与胞质中某些非核糖体蛋白质形成 RNP 复合物.

H 19 基因与胚胎发育和肿瘤发生密切相关^[1]. *H 19* mRNA 是胚胎期特征的分子事件. *H 19* 基因在肾母细胞瘤和其他一些胚胎性肿瘤中早期失活, 行为类似肿瘤抑制基因. 将 *H 19* 基因导入某些肿瘤细胞系会引起细胞形态逆转、非停泊生长减慢、对裸鼠致瘤性降低; 但在另外一些情况, 肿瘤细胞中也可以检出高水平的 *H 19* mRNA, 此外在膀胱癌和乳腺癌, *H 19* 基因的再表达则是肿瘤发生的分子特征, 因此人们又认为它是一种癌胚基因. 在腺癌中 *H 19* 基因在不同细胞的表达及其表达量与肿瘤的分期和预后密切相关.

2 *XIST* 基因和 *XLSIRT* 基因

XIST 基因定位于人和鼠的 Xq13, 在 X

染色体失活中心 (XIC) 内, 仅在雌性细胞失活的 X 染色体转录, 与 X 染色体失活有关. 成熟的 *XIST* 基因产物长 15 kb, 由六个外显子组成. *XIST* mRNA 不含有长的 ORF, 编码蛋白的可能性极小. 在 *XIST* mRNA 序列中含有至少 5 种串联重复序列 (可能的 DNA 或 RNA 结合蛋白结合区), 这些串联重复序列具体的重复次数在人和鼠并不一致, 但这可能不影响其高级结构和功能上的保守性. *XIST* mRNA 全部位于细胞核中, 覆盖于失活的 X 染色体 (Barr 小体) 上.

XIST mRNA 在 X 染色体上的积累是染色体失活的必要步骤^[2]. *XIST* 基因在雌性胚胎发育过程中有父源印迹现象. Xp (paternal X) 在精子形成时被印迹, 胚胎发育早期 (2.5 dpc) 在滋养层细胞内非选择性失活, 但它在胚胎细胞中去印迹, 胚胎细胞于 5.5~8.5 dpc 建立随机 X 染色体失活. 失活 X 染色体转录的 *XIST* mRNA 数量的累积主要是因为其稳定性较活性 X 染色体转录本升高. *XIST* 基因的转录活性主要与其 5' CpG 岛的甲基化程度有关, 低甲基化会使 *XIST* mRNA 过高表达, 导致 X 染色体全部失活, 引起细胞凋亡. *XIST* mRNA 的作用可能是参与 X 染色体骨架的形成, 从而调节 X 连锁基因的表达. 在果蝇中剂量补偿效应的 X 染色体也被一非编码 mRNA (*rox* 详见后) 覆盖, 这提示它们可能具有共同的作用机制.

Xlsirt RNAs 是一群散布于爪蟾卵母细胞皮质由重复序列组成的与人和鼠的 *XIST* 基因同源的 RNAs. *Xlsirt* 基因一般由 3~13 串联重复单位 (79~81 nt) 组成, 这些串联重复单位与 mRNA 3' UTR 顺式作用相似, 介导 *Xlsirt* RNAs 在卵母细胞发育中定位到植物极的过程. 破坏 *Xlsirt* RNAs 定位将导致后进入植物极的 mRNA *Vg1* 扩散出来.

3 *His-1* 基因和 *bic* 基因

His-1 和 *bic* 基因是从逆转录病毒转化的白血病细胞克隆的两种非编码基因, 它们正常

只在局部组织低水平转录, 但当逆转录病毒在其上游位点插入提供启动子时 (promoter insertion) 基因转录被活化.

His-1 位点定位于人和鼠的 2 号染色体, 是鼠髓样白血病常见的插入位点. *His-1* 基因是在鼠的 *His-1* 位点克隆的一个与细胞表型转化相关的基因^[3], 长约 6 kb, 由 3 个外显子组成, 外显子 3 具有 8 拷贝的 ATTTA motif, 该 motif 可通过跟特异的 RNA 结合因子作用参与癌基因的去稳定和细胞因子信号传递. *His-1* mRNA 长 3 kb, 序列含有大量终止子. *His-1* 基因正常仅在局部组织少量转录, 但在鼠淋巴瘤及脉络膜肉瘤中可检出较大量的转录. 通常在时空上局限转录基因的过分表达则会引起肿瘤, *His-1* mRNA 的高表达应与髓样白血病的发生相关.

bic 位点是禽白血病病毒 (ALV) 诱导的 B 细胞淋巴瘤中常见的逆转录病毒插入位点. 由于 *bic* 位点与 *c-myc* 基因的活化紧密关联并在转移瘤中优先活化, 人们认为在 *bic* 位点中含有一基因, 它与 *c-myc* 基因协同参与淋巴瘤发生, 并对晚期肿瘤发展起作用. *bic* 基因即是从 *bic* 位点克隆的一个与该位点功能相关的新基因^[4]. *bic* 基因的表达量随不同胚胎发育阶段变化, 成体中正常情况下仅在淋巴器官中表达, 但在含有 *bic* 位点插入的肿瘤细胞中可检出高水平的含 *bic* 外显子 2 的嵌合 RNA. *bic* 基因有两个外显子, 外显子 2 有 a/b 两个选择性剪切供体, *bic* cDNA 序列没有典型的阅读框架, 空间结构分析发现, 外显子 2a 能形成广泛的二级结构, 尤其是 316~461 nt 可形成稳定的茎环结构. 有实验证明双链 RNA 可与 PKR 结合并调节其功能, 而后者广泛参与细胞的生长和分化、调节基因转录和/或抑制翻译. *bic* mRNA 是否通过这一机制作用尚需进一步证实.

4 *ROX1* 和 *ROX2* 基因

rox 1 和 *rox 2* 是从成年雄性果蝇的神经元克隆的两个 X 连锁、雄性特异、非编码的新基因^[5], 它们与 X 染色体上卵巢表达的雌性

特异基因连锁, 受剂量补偿基因 *sxl*、*mos-1*、*mos-2*、*mos-3* 等调控. *rox 1*、*rox 2* 分别位于 X 染色体的 3F 和 10B 区, 转录产物覆盖在失活的 X 染色体上. *rox 1*、*rox 2* 与 *Xist* 基因相似, 抑制与其连锁基因的表达. 雄性果蝇唾液腺核仁中与单个 X 染色体作用的剂量补偿复合体对 RNase 高度敏感, *rox 1*、*rox 2* 可能是剂量补偿复合体的组分, 通过调节 X 染色体的结构发挥其功能.

5 *NTT*, *IPW* 和 *UHG* 基因

NTT, *IPW* 和 *UHG* 基因是 1997 年新近分离的三个非编码基因, 因其发现较晚, 研究工作比较少, 故在此只作简单介绍. *NTT* 仅在某些活化的 CD⁴+ T 细胞克隆中转录, 与 INF- γ 受体基因相邻, 两者表达互相影响. *IPW* 基因 (imprinted gene in the Prader-willi syndrome) 位于 15q11~q13, 母源印迹, 在 PWS 患者中不转录, 可能与发病有关, 鼠中的同源物 *Iwp* 主要在脑中表达. *UHG* (U22 host gene) 基因位于 11q13, 成熟的 mRNA 其实是由其内含子转录而成的. *UHG* mRNA 分布在整個核仁, 形成核糖核蛋白体, 参与 rRNA 的加工和修饰.

非编码的 mRNA 分子目前发现的虽然只有几个, 但已有的实验结果提示: 它们参与许多基本的生命过程, 诸如胚胎的发育、染色体的失活、细胞的生长和分化、肿瘤的形成和抑制等, 具有重要的功能. 非编码的 mRNA 分子虽然是 mRNA 的一种, 但它们的作用机制似乎与其他 RNA 分子更接近, 这可能是因为它们在广义上都应属于非编码的 RNA 分子. 非编码的 RNA 分子可以自身直接作用, 例如 ribozyme 和反义 RNA; 它们也可以通过与蛋白质结合形成复合体作用, 例如 7SRNA 和 snRNA. 目前关于 RNA 和蛋白质的结合有一些新发现: RNA 可以帮助原本松散易降解的蛋白质形成高级结构, 使之变得稳定, 从而提高其细胞内浓度, 引起相应的生理反应. 这一发现对于传统观念是个补充: 在 RNA 和蛋白

质的相互作用中, RNA 并不是只处在被动受调控的地位, 它们可以互相调节对方的结构、定位、半衰期和功能. 从进化上看, 人们现在认为在蛋白质出现之前有一个 RNA 世界, RNA 在那个时期作为信息分子、结构分子和酶发挥着全能的作用. 非编码的 RNA 的功能提醒我们: 位于人基因组中 90% 的非编码序列作为亿万年生物进化的产物, 它们决不会也不可能无用的累赘, 它们的重要性将逐渐被了解.

参 考 文 献

- 1 Looijenga L H, Verkerk A J, de Groot N *et al.* H19 in normal development and neoplasia. *Mol Reprod Dev*, 1997, **46** (3): 419~439
- 2 Kuroda M, Meller V H. Transient Xist-ence. *Cell*, 1997, **91** (1): 9~11
- 3 Askew D S, Li J, James N. Ihle Retroviral insertion in the murine *His-1* locus activate the expression of a novel RNA that lacks an extensive open reading frame. *Mol Cell Bio*, 1994, **14** (3): 1743~1751
- 4 Tam W, Ben-Yehuda D, Hayward W S. *bic*, a novel gene activated by proviral insertion in Avian Leukosis Virus induced lymphomas, is likely to function through its noncoding RNA. *Mol Cell Bio*, 1997, **17** (3): 1490~1502
- 5 Amrein H, Axel R. Genes expressed in neurons of adult male drosophila. *Cell*, 1997, **88** (4): 459~469

Non-coding mRNA. LUO Wen-qin, YUAN Jian-gang, QIANG Bo-qin (*Institute of Basic Medical Science, Chinese Academy of Medical Sciences & Peking Union Medical College, National Laboratory of Medical Molecular Biology, Beijing 100005, China*).

Abstract Non-coding mRNA is a kind of newly discovered mRNA. Though spliced and tailed, it does not have any typical ORF. Up to date, several such genes have been cloned: *H 19* gene, *XIST* gene, *XLSIRT* gene, *His-1* gene, *bic* gene, *rox 1* and *rox 2* genes. Non-coding mRNA play important roles in the development of embryos, in the inactivation of X chromosome and in the tumorigenesis.

Key words non-coding mRNA, *H 19* gene, *XIST* gene, *His-1* gene, *bic* gene, *rox 1* gene and *rox 2* gene