

文章编号: 1004-0374(2010)07-0641-08

长非编码 RNA 的作用机制及在肿瘤发生发展中的意义

柏庆然, 宋旭*

(四川大学生命科学学院功能基因组与生物信息学研究中心, 成都 610064)

摘要: 功能基因组学的飞速发展将越来越多的目光引向了对非编码转录产物功能的研究。在人的转录组中, 存在着一类长度大于 200 nt, 但并不具备编码蛋白质功能的基因转录产物, 即长非编码 RNA (long noncoding RNA, lncRNA)。相比于小分子 RNA, 它们仍是目前基因组转录产物中较为陌生的部分。在整个基因组转录产物中, lncRNA 所占的比例远远超过编码 RNA 所占的比例。不同于编码 RNA, lncRNA 的保守性要差得多, 然而在其分子内部, 却含有较为保守的局部区段, 且其表达具有时空特异性, 这些现象都提示了 lncRNA 具有重要的生理生化功能。越来越多的研究表明, lncRNA 在基因表达调控方面发挥着十分重要的作用, 与物种进化、胚胎发育、物质代谢以及肿瘤发生等都有着紧密的联系, 其功能的深入研究将使目前对细胞的结构网络和调控网络的认识带来革命性的变化, 具有不可估量的科学和临床价值。该文将着重讨论 lncRNA 在不同层面对基因表达的调控机制以及在肿瘤发生发展中的意义。

关键词: lncRNA; 基因表达调控; 肿瘤发生

中图分类号: Q522; R730.231 **文献标识码:** A

The roles of long noncoding RNAs and its implications in tumorigenesis

BAI Qing-ran, SONG Xu*

(Center for Functional Genomics and Bioinformatics,
College of Life Science, Sichuan University, Chengdu 610064, China)

Abstract: One of the most important discoveries from studies at the genomic scale is that there are much more abundant non-coding transcripts in different species than previously imagined. Long noncoding RNAs (lncRNAs) are defined as transcripts longer than 200 nt without coding capacities. Compared to short ncRNAs, the functions of lncRNAs are less understood. Although the overall similarity are relatively low, strongly conserved elements are found in homologous lncRNAs among different species. The spatio-temporal expression profiles of some lncRNAs indicate they may have magnificent physiological and biochemical functions. Recent studies show that lncRNA plays a key role in the regulation of gene expression, development, oncogenesis and evolution. In this review, we mainly discuss the roles of long noncoding RNAs at the level of transcription, and its implications in tumorigenesis.

Key words: lncRNA; regulation of gene expression; tumorigenesis

1 LncRNA 简介

在某些情况下或者是至少在一种特定的细胞类型中, 人类的常染色体中几乎每一对核苷酸对都会发生转录现象^[1]。能够稳定存在的转录产物中信使 RNA (messenger RNA, mRNA) 不超过 2%, 其余绝大部分为非编码 RNA (noncoding RNA, ncRNA)^[2]。

目前关于 ncRNA 的分类, 有两种较常用的方法, 一种是依据 ncRNA 表达特点及功能, 将其划分为组成型 ncRNA (constitutive ncRNA) 和调节型 ncRNA (regulatory ncRNA); 另一种是依据 ncRNA 分子大小

收稿日期: 2010-05-05

*通讯作者 E-mail: xusong@scu.edu.cn

也就是所含碱基数量的多少,划分为 lncRNA(碱基数>200)和小分子非编码 RNA。在人体内, lncRNA 在数量上占全部 ncRNA 转录的大部分。

lncRNA的编码基因在基因组范围内的分布非常广泛,可能位于编码 mRNA 基因的外显子或内含子中,也可能位于编码 mRNA 基因之间的序列。lncRNA 与 mRNA 有很多相似的特点和重叠的区域:因为基因组中所谓的转录热点常常既转录 mRNA,也转录 lncRNA,蛋白质编码基因两条链中的任意一条都可能转录 lncRNA^[1]。而同一个蛋白质编码基因可以转录出不同的 mRNA 或者是 lncRNA^[3]。一部分 lncRNA 和 mRNA 可由外显子通过可变剪切组合而成^[3],并且在 3' 末端添加有多聚腺苷酸^[4]。此外, lncRNA 的亚细胞定位也比较复杂,可位于细胞核内,也可位于细胞质中,加之 lncRNA 分子构成序列上的复杂性,大大增加了对其进行研究的难度。

lncRNA 与 mRNA 的相似性揭示了其可能的起源,但目前关于 lncRNA 的起源的确切过程尚不清楚,有不同的模型对其进行阐释:(1) lncRNA 是由蛋白质编码基因演变而来的。有可能在早期进化的过程中,原本编码蛋白质的基因,由于开放阅读框(open reading frame, ORF)遭到突变而失去原有功能,使得残留的片段变成了具有一定调控功能的 lncRNA。支持这一假说最有力的实验证据来源于在细胞发育过程中调控 X 染色体失活的 Xist^[5]。(2) lncRNA 有可能来自于染色体的重排,使得原先在染色体上相距较远的非转录片段彼此靠近,进而产生出一个新的含有多个外显子的 lncRNA。(3) 由逆转录转座作用产生的非编码基因的重叠有可能会产生功能性的 lncRNA^[6]。在 lncRNA 内,一些相距较近连续重复序列可能来源于随机的复制事件,但也有可能是来自于转录元件的插入,这些插入则有可能赋予 lncRNA 以功能:因为转座元件常常会与转录起始位点发生重合,故也可能对插入基因的整个转录表达产生影响^[7]。

不同于 mRNA 和小分子 RNA, lncRNA 在不同物种之间的保守性较差。这可能对进化快速适应的结果^[8],因为 lncRNA 可能在进化压力下比蛋白质编码基因更容易改变,蛋白质编码基因只要发生非同义突变,就会导致整个 ORF 的改变,而 lncRNA 则对于突变具有更大的包容性,这将有助于形成物种间或物种内的差异表型^[9]。虽然在全序列的保守性较差,然而 lncRNA 在某些短区段内的保守性非常高,相比于其他区段,它们的碱基置换率、外来

片段插入率以及自身片段丢失的概率都很低,核酸剪切位点相当稳定,在 lncRNA 的稳定转录和剪切层面上体现了其功能序列的完整性。这些高保守元件的特定核苷酸序列或空间结构往往对 lncRNA 功能的发挥至关重要,其他区段则可能维持其二级结构的稳定或对 lncRNA 功能的发挥起辅助作用。

2 lncRNA 是重要的基因转录表达调控元件

lncRNA 具有局部高度保守的序列元件,特定的空间二级结构,组织或细胞特异性表达以及明确的亚细胞定位,提示了这些在细胞总转录物中占极大比例,但不具蛋白质编码能力的转录产物可能具有重要的生物学功能。以前认为蛋白质是生物调控的执行元件,现在发现 lncRNA 通过与蛋白质相互作用,参与到生物调控网络中。在真核生物中,基因的转录在时空上具有很强的特异性,且处于严密而精确的调控之下。随着对 ncRNA 以及基因表达调控网络研究的深入, RNA 通过与蛋白质或 DNA 之间特异性的动态相互作用来调控基因表达的行为逐渐为人所认识^[10]。lncRNA 可以作用于转录激活因子或转录抑制因子,作用于 RNA polymerase II (RNAP II) 的不同组分,也可以直接与 DNA 双螺旋相互作用^[10],从不同方面影响基因转录表达,这说明了在真核细胞中,存在着由 lncRNA 参与构成的复杂而重要的基因表达调控网络,巧妙地调节基因表达。

lncRNA 调控基因表达的方式存在着丰富的多样性,表现在 lncRNA 依赖的基因表达调控机制多种多样。在不同的机制中, lncRNA 扮演的角色可能不同,其自身既可作为主要的转录调控因子,又可作为共调控因子之一(co-regulator),与其他组分一道在基因表达过程中发挥调控作用。lncRNA 可在多个层次上对基因表达进行调控,依据目前研究所得到的信息, lncRNA 对于基因表达调控的层次大体上可以分为:(1) 表观修饰水平调控;(2) 转录水平调控;(3) 转录后水平调控。

2.1 表观修饰水平调控

lncRNA 在基因的表现修饰水平的调控上具有重要意义,通过与各种染色质修饰酶的相互作用,对染色质进行化学修饰,改变其构象,激活或抑制相关基因的表达^[11]。尤其是在胚胎发育阶段, lncRNA 参与引起可遗传的等位基因后期表达沉默、表现性状的维持,对于多细胞动物的正常发育和细

胞分化至关重要。

在人类发育过程中, 数以百计的lncRNA在人类同源异型框(homeobox, Hox)基因座的选择性表达中发挥调控作用, 它们决定着这些同源异型基因座染色质结构域中组蛋白甲基化修饰会否发生、染色质结构是否允许RNA聚合酶转录等。其中一种由HOXC基因座转录的Hox转录物反义RNA(HOTAIR), 会通过募集染色质重构蛋白复合物PRC2, 诱导HOXD基因座产生抑制性的染色体结构, 在HOXD基因座上长达40 kb的范围内抑制转录的发生^[12]。很多染色质重构蛋白间接调控基因表达的现象都可以用这种机制解释, 目前认为, 在这种机制中, 起核心作用的是lncRNA: 因为染色质重构蛋白具有的DNA定位效应很差, 对于所调控的染色质区段的特异性识别主要是由lncRNA完成的。

然而, lncRNA介导的表观修饰水平调控通路存在着许多看似矛盾的现象。比如在经典的X染色体失活(X-chromosome inactivation, XCI)机制假设中认为, 失活是由X染色体特异性失活转录物(X inactivation-specific transcript, Xist)在失活的X染色体上的拷贝的表达所触发的。将要失活的X染色体会在胚胎发育早期由其上X染色体失活中心(X inactivation center, XIC)的*Xist*基因座转录Xist, Xist随后覆盖在该X染色体有限的关键位点上, 引发一系列的染色体结构的修饰变化, 使得相关基因失活^[13]。而在一项利用雌性小鼠着床前胚胎的实验中发现, 其父源X染色体的XCI可以在没有父源*Xist*基因或是父源*Xist*基因存在缺陷的情况下启动。但XCI启动后如果仍旧没有Xist参与, X染色体最终会再活化。这说明Xist的作用或许只是稳定X染色体的失活状态^[14]。在另一项利用小鼠胚胎干细胞及小鼠胚胎成纤维细胞的研究中还发现, 在有活性的X染色体上, Xist与对XCI具有特异性抑制作用的另一种lncRNA Tsix(Xist的翻译产物)形成RNA二聚物, 随后被Dicer酶加工产生内源性的siRNA。而在失活的X染色体上, Xist的定位、积累以及对失活染色体进行染色质结构修饰的过程需要Tsix和Dicer酶的参与。早前的研究认为Tsix的功能是特异性的抑制XCI, 但在此过程中Dicer和Tsix的具体作用机制和功能尚未明了^[15]。

这说明目前关于lncRNA在表观修饰水平调控中功能的认识相当有限, 其中很多关键的问题包括lncRNA作用的时空特异性、作用对象, 自身的分

子状态、包括参与其调控的协同作用因子的种类与功能等等都需要进行更为细致的研究并进而提出更为合理的解释。

2.2 转录水平调控

当lncRNA自身作为主要的转录调控因子时, 其功能发挥主要依赖于lncRNA基因与所调节的靶基因在基因组中的相对位置或序列特征, 从而直接改变靶基因染色体结构, 或者造成多个转录因子间的作用冲突使lncRNA基因和靶基因转录这两个独立的事件互作, 调控其转录和表达^[16]: 以酵母基因*SER3*为例, 在它的上游存在lncRNA基因*SRG1*, *SRG1*的3'端序列与靶基因启动子重叠, 当*SRG1*被活跃转录并过表达时就会抑制靶基因*SER3*的转录^[17]: 因为在转录延伸过程经过启动子序列时, 转录延伸因子会由于转录lncRNA占据启动子与转录起始因子结合的空间, 造成启动子下游靶基因序列转录的抑制^[16]。转录干扰现象的存在似乎说明了转录过程本身在转录调控过程中具有极为重要的生物学意义。

上游lncRNA的转录也可能增强靶基因与RNA聚合酶的亲和力。位于人 β -球蛋白上游lncRNA的序列存在与 β -球蛋白基因的重叠, 当lncRNA转录时会改变染色体的结构, 逆转原先存在的转录抑制结构, 使转录因子可以与靶基因启动子结合, 促进靶基因的转录^[18]。

有些在靶基因上游转录的lncRNA可以直接与靶基因序列作用, 通过两种分子上的碱基互补作用, lncRNA会与靶基因启动子形成稳定的DNA-RNA三联复合物结构, 同时还会与转录因子相互作用, 使得由靶基因启动子与转录因子形成的转录起始复合物解聚, 从而抑制靶基因的转录。这种情况存在于人二氢叶酸还原酶(dihydrofolate reductase, DHFR)的表达调控中, DNA-RNA三联复合物的形成使得转录因子IIB(general transcription factor IIB, TF IIB)不能与DHFR基因启动子结合开启转录^[19]。虽然有人预测, 这种DNA-RNA三联复合物结构可能普遍出现在人类基因启动子中^[20], 但通过与启动子基因形成三联复合物而调控基因表达是否是一种普遍的lncRNA依赖的基因调控机制, 还有待于进一步的证实。

同时, lncRNA作为转录过程中的共调控因子, 与其他组分一起, 在基因转录中发挥作用。其具体作用机制包括与转录因子及其组分作用, 改变转录因子及其组分的活性、亚基聚合状态等, 诱发

转录因子的空间结构改变,使之失去与靶基因的亲和能力或者催化活性;与除转录因子外的其他反式作用因子作用,影响其与靶基因的相互作用,间接调控靶基因的转录。在转录调控过程中,直接与lncRNA相互作用的共调控因子有相当一部分是蛋白转录因子,而具有RNA结合能力的蛋白质在整个哺乳动物蛋白质中有1 000余种,它们与lncRNA的相互作用极大地丰富了转录调节的手段。

lncRNA可以将转录调控因子募集到邻近的靶基因启动子上,调控其表达。在受到金属离子射线等细胞损伤信号的刺激下,细胞应激性地启动位于细胞周期蛋白(cyclin D1, CCND1)基因启动子上游的lncRNA转录,转录产物会与一种RNA结合蛋白TLS(translocated in liposarcoma)结合,将其募集到CCND1的启动子上,并调节其与p300和CREB结合蛋白的结合,通过结合后的变构效应改变抑制的组蛋白乙酰转移酶活性,最终导致CCND1的转录表达下降^[21]。

细胞中有些RNA结合蛋白同时还具有DNA结合结构域,它们结合在靶基因的启动子上,或启动靶基因的表达,或抑制靶基因表达。与lncRNA的结合,改变了这些蛋白与靶基因的位置关系,使得靶基因表达状态改变^[22]。在人黑素瘤细胞中发现人类逆转座子基因,如LIPA16、MER11C,人类非编码基因MALAT-1及人类线粒体基因HN等,它们的非编码转录产物能够结合PSF蛋白,使后者的空间构象发生改变,将其蛋白从原癌基因GAGE6的启动子上释放下来,解除对GAGE6的抑制,使之表达。这种新近发现的调控机制在胚胎发育、细胞分化、肿瘤发生以及病毒感染过程发挥极为重要的作用^[23]。

另一些lncRNA的调控功能依赖于与蛋白质形成核糖核蛋白复合物。在神经发育中起重要作用的lncRNA Evf2即是如此。Evf2由位于所调控的靶基因Dlx5和Dlx6间的超保守序列元件转录而来,与转录因子Dlx2结合形成复合物,激活Dlx5/Dlx6的增强子,上调两者的转录。虽然激活的具体机制暂时还不清楚,不过有推测认为,激活可能是Evf2的序列直接特异性的定位到拥有同源性序列的靶基因的增强子上而完成^[24]。哺乳动物基因组中类似的超保守元件既可能被转录又具有增强子的功能,说明这可能是哺乳动物生长过程中调控关键发育基因表达的一种普遍策略^[25]。

调节转录因子亚基的聚合状态也是lncRNA调控转录的手段之一。HSR1(heat shock RNA-1)是一类与细胞热敏感性相关的lncRNA,与真核转录延伸因子1A一起作用,诱导热休克因子(heat-shock factor1, HSF1)三聚体化。以三聚体形式存在的HSF1方可结合到热休克基因的启动子上,启动其表达^[26]。

lncRNA还可以通过改变转录因子的亚细胞定位来影响靶基因转录。例如NRON(noncoding repressor of NFAT)可能通过与核转运因子(nuclear transport factor)的作用,限制转录因子NFAT(nuclear factor of activated T cells)进入核内转录相关基因^[27]。需要指出的是,通过预测,NRON上可能大量分布高保守的茎环结构,而NRON的功能及与核转运因子的相互作用或许依赖于这些茎环结构。

在lncRNA作为共调控因子参与的基因转录调控中,很多调控机制和lncRNA与RNAP II的相互作用有关。很多作用于RNAP II的lncRNA是由RNAP III转录^[28],因而在这部分lncRNA与RNAP II之间,不存在偶联的相互作用关系,也就不存在直接的反馈调节,有利于在细胞遭遇压力时及时终止蛋白质编码基因的转录。RNAP II作为真核生物中一种普遍的转录因子,对蛋白质编码基因的转录起着至关重要的作用。lncRNA可以通过很多途径调控RNAP II的活性,包括与转录起始复合物相互作用影响其对启动子的选择,进而选择性地激活或抑制基因表达。不仅如此,lncRNA还能通过与构成RNAP II依赖的转录复合物的基本组分相互作用,影响总体转录。例如,在细胞受到压力的情况下,7SK RNA可以与延伸因子P-TEFb(positive transcription elongation factor b)在另一类RNA结合蛋白HEXIM1(hexamethylene bisacetamide-induced protein-1)的参与下形成复合物,降低P-TEFb的活性,抑制其磷酸化RNAP II与转录活性相关的磷酸化位点,从而抑制转录^[29]。类似的情况还存在于lncRNA Alu调控RNAP II的机制中,不同在于Alu直接与RNAP II相结合抑制其转录活性。Alu中独立介导与聚合酶结合的区段在整个哺乳动物基因组中都有大量而广泛的分布,在很多已知的lncRNA中都发现了这些区段的存在,表明它们可能在进化过程中被整合入lncRNA^[30],也提示了大量的包括活跃转座元件的lncRNA可能通过靶定RNAP II来调节基因组中特定基因的转录。

2.3 转录后水平调控

类似于小分子RNA的调控机制, lncRNA介导的转录后调控主要以自身序列与靶mRNA序列互补配对为基础。lncRNA与mRNA经过序列配对形成的RNA二聚物可以掩蔽mRNA分子上与转录加工过程相关因子的作用位点, 从而达到调控mRNA转录后加工的目的。受其调控的转录后加工过程包括了剪切、转运、翻译和降解等。

最早发现的受lncRNA直接调控可变剪切的靶mRNA是Zeb2。5'端非编码区中内含子剪切会受到其异位表达的反义lncRNA的调控, 后者会与5'端非编码区中内含子发生互补, 使得含有翻译起始位点的内含子片段能够逃过剪切而被保留, 确保Zeb2基因的表达^[31]。与之类似, 甲状腺激素受体ErbA α 2 mRNA的可变剪切也与反义lncRNA有关, 虽然lncRNA仅部分序列与之互补, 但仍能调控剪切过程, 剪切产物随后产生相互拮抗的两种蛋白质。

3 lncRNA调控肿瘤发生

随着对lncRNA在细胞生物学中的功能不断的深入研究, lncRNA与疾病的关系得到越来越多的关注。很多研究已经表明, lncRNA与很多疾病的发生有关, 如lncRNA所调控的基因的异常表达会产生严重的病理变化。包括肿瘤在内的很多疾病都存在lncRNA的异常表达, 尽管目前对lncRNA的致病机理还知之甚少, 但大量由临床观察和实验得来的证据显示, lncRNA的异常表达是疾病发生的重要原因之一。通过全基因组序列分析发现, lncRNA中特定的超保守元件在人类肿瘤细胞中存在着广泛表达。后续研究则发现这类超保守元件与正常细胞的原癌基因有关, 起着抑制细胞凋亡的作用。它们在正常细胞中有很多拷贝, 分布在染色体的脆性位点和肿瘤相关区域。这说明这类超保守元件本在正常个体发育中发挥重要作用, 而其异常表达则会导致细胞发生恶性转化^[32]。OCC-1(结肠癌过表达基因-1)RNA在结肠癌细胞中过量表达^[33]。而在前列腺肿瘤中, 一种名为PCGEM1的lncRNA的过量表达, 导致了肿瘤细胞的增殖和克隆形成^[34]。MALAT-1 RNA是另一种已知的肿瘤相关lncRNA, 多篇文献报道了它的高表达与非小细胞性肺癌、乳腺癌、结肠癌、前列腺癌、肝癌等肿瘤的恶性化程度密切相关。而MALAT-1同源基因在小鼠肝癌细胞中也是高表达的^[35]。

虽然lncRNA的异常表达在各类肿瘤中相继被

发现, 但是它们在肿瘤发生、发展中所扮演的角色和所发挥的功能尚未被深入了解。宋旭实验室致力于研究lncRNA在肿瘤发生中的作用, 于2002年发现小鼠lncRNA VL30可以与PSF(PTB-associated splicing factor)蛋白结合诱发肿瘤的新途径。研究揭示了在哺乳动物中高度保守的PSF蛋白的新身份——一种极为重要的肿瘤抑制蛋白。PSF蛋白有一个DNA结合结构域(DNA binding domain, DBD)和一个RNA结合结构域(RNA binding domain, RBD), PSF蛋白的DBD通过与原癌基因启动子结合抑制其转录来抑制肿瘤发生, 而VL30等lncRNA与PSF蛋白的RBD结合形成lncRNA/PSF蛋白复合物, 使PSF蛋白从原癌基因启动子解离下来, 解除PSF蛋白对原癌基因表达的抑制^[23]。随后, 发现除了小鼠VL30 RNA, 还有5种来源于人的lncRNA可与PSF蛋白结合调控其功能, 分别为LIPA16、MER11C逆转座子RNA、MALAT-1 RNA、来源于线粒体的HV RNA, 以及一未知基因转录产物。它们可以与PSF蛋白结合, 将其从原癌基因GAGE6的启动子上释放, 激活GAGE6转录。每一种lncRNA都能够单独诱导细胞恶性转化, 而且其表达量与肿瘤恶性转化程度密切相关。多种肿瘤相关蛋白都同时具有DBD和RBD, 这些蛋白质可能也能被lncRNA所调控, lncRNA与蛋白相互作用调控肿瘤发生发展过程可能是一个普遍机制。

目前关于lncRNA与包括肿瘤在内的疾病相关联的证据大多来自于lncRNA表达水平上的差异, 此可为疾病诊断和治疗提供依据和靶点。而由ncRNA序列突变造成功能紊乱还有待进一步的研究。至于lncRNA的致病机制, 除上述模型外, 还存在别的类型, 如在很多肿瘤中发现lncRNA作为基因沉默元件与抑癌基因转录产物结合, 造成表达沉默, 诱发肿瘤^[36]; 或lncRNA形成特殊的发卡结构, 干扰靶基因mRNA的正常剪切, 造成病变等。

4 lncRNA研究展望

lncRNA的研究已然成为现代分子生物学领域研究正在兴起的重要热点, 但其发展的历史却相对较短。在相当长的一段时间内, 由于研究方法和技术的局限, 导致对RNA的认识不足, 随着高通量测序技术的成熟和应用以及生物信息学的蓬勃发展, 包括lncRNA在内的ncRNA具有基因表达调控功能相继被发现, 这既有些出人意料, 但却又在情理之中。RNA数量庞大, 调控的方式多种多样, 已知

的调控模型就包括与靶 mRNA 序列互补配对、改变靶基因空间结构、作用于调控靶基因的转录因子等形式。丰富的调控手段可能构成对靶基因调控的多条通路,对细胞生理稳定性的维持有重要意义。为什么进化的过程保留了以 RNA 为基础的调控系统?这个问题的答案可能是多方面的,我们推测可能有一些原因:一些生理过程的生化特性决定了必须要有 RNA 的参与,例如端粒的复制;与蛋白因子的相比, RNA 具有许多相似的生化性质,例如复杂的高级结构,这也就决定了 RNA 可以行使和蛋白因子相似的生物功能。随着生命科学的发展,更多的 RNA 参与的调控过程将会被发现。RNA 除了可与蛋白因子直接作用调控其功能,还可以通过碱基配对的形式识别其他核酸分子,这一生化特性决定了其在某些调控方面的特定优势:(1)高等生物在对病毒应急反应时,产生具有生理活性的 RNA 识别病毒核酸要比产生具有生理活性的蛋白过程更快, RNA 调控能短时间内发挥生理效应无疑增强了细胞的抗逆性和适应性;(2)蛋白转录因子通常识别 8~12 个碱基序列,但它们在基因组上的定位远远低于随机出现的几率。这些因子的定位很可能是由 RNA 通过与相应 DNA 的配对引导的。在酵母和植物中,这一转录调控机制普遍存在;在哺乳动物中该机制是否同样普遍存在还有待研究。

近年来,内源性小分子 RNA 的研究已取得了令人瞩目的成果,包括 RNAi 在内的以小分子 RNA 的基因表达调控功能为基础的工程学手段问世,大大推进了生命科学和医学的发展。相比于其他 RNA, lncRNA 在表达调控中的特色主要与自身分子长度有关:第一, lncRNA 既可以通过核酸序列的一级结构与其他 DNA、RNA 相互作用,又可以通过较长的核苷酸链形成复杂而多样的分子内高级空间结构与蛋白因子相互作用,作用方式多样,作用对象广泛;第二, lncRNA 的长链状分子能够提供较大的空间位置同时与多个分子相接触,这使得 lncRNA 可以作为功能性平台,通过自身的不同区域,可以与不同的蛋白质、DNA、RNA 相结合,将它们整合在合适的空间范围内,发挥生物学功能;第三,除了部分区段以外, lncRNA 较长的分子序列的整体保守型较差,对变异的包容性较大,有利于维持 lncRNA 的功能长期存在,在物种对进化上的适应上具有重大的意义。

除了基因调控功能外, lncRNA 的功能还涵盖

了染色体高度有序的动态结构变化、端粒生物学及亚细胞结构定位等领域^[37]。对 lncRNA 的鉴定使我们对整个功能性基因调控的基础进行重新解读。转录组研究的持续进展揭示了 lncRNA 在基因组表达上的重要功能。lncRNA 功能的注释对于现在很多的研究具有现实意义。例如端粒对于染色体稳定有着重要的作用,与衰老和肿瘤密切相关^[38]。目前发现由端粒基因座转录的 lncRNA,其长短不一,含有重复程度不同的端粒重复序列^[39],可能与端粒染色质修饰有关。体外检测发现这些 lncRNA 能够抑制端粒酶的活性^[40],表明 lncRNA 或许在端粒生物学研究中发挥重要的作用。又如在 HeLa 细胞中,丁型肝炎病毒的基因组 RNA 可以特异地与 PSF 结合,推测某些致癌病毒的基因组 RNA 可能通过与内源性的 PSF 结合,解除对原癌基因的抑制诱发肿瘤^[41],对病毒诱发癌症机理的研究具有相当的提示意义。

lncRNA 在漫长的进化中被赋予了复杂的结构、信息等方面的功能,通过与 DNA、RNA、蛋白质的相互作用,在生命活动调控网络中扮演着十分重要的角色。lncRNA 在生物中的丰度随着组织复杂性的增加而增加,证明了 RNA 的进化与调控功能的扩展是真核生物基因控制复杂性的基础,但目前所鉴定的功能性 lncRNA 与依据生物信息学手段所推测的数目相比仅仅是冰山一角,我们对 lncRNA 的了解还存在很多空白,比如功能性 lncRNA 在所有 lncRNA 中所占的比例;它们的分布范围及作用机制;它们的进化机制、进化速率,还有如何在进化中获得或失去功能等等。此外,很多编码蛋白质的 RNA 除了编码功能,还发现具有其他的功能,即具有作为 lncRNA 发挥功能的能力,不仅如此,很多迹象表明,任何一种 RNA,无论其是否具有蛋白质编码能力,都存在着自身固有的功能性信息。这提示了基因组可能存在一套以 RNA 为基础的信息系统,而这套系统要远比想象的复杂。如果包括 lncRNA 在内的大多数 ncRNA 具有功能的话,那么它们无疑会对关于生物基因组复杂性的理解产生意义重大而深远的影响,那些关于 lncRNA(包括发挥除编码蛋白质以外功能的 mRNA)功能的研究会对生物学长久以来关注的进化、发育和疾病问题带来新的解答。

[参 考 文 献]

- [1] Birney E, Stamatoyannopoulos JA, Dutta A, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 2007, 447

- (7146): 799-16
- [2] Bertone P, Stolc V, Royce TE, et al. Global identification of human transcribed sequences with genome tiling arrays. *Science*, 2004, 306(5705): 2242-6
- [3] Kampa D, Cheng J, Kapranov P, et al. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res*, 2004, 14(3): 331-42
- [4] Tian B, Hu J, Zhang H, et al. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res*, 2005, 33(1): 201-12
- [5] Duret L, Chureau C, Samain S, et al. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science*, 2006, 312(5780): 1653-5
- [6] Hutchinson JN, Ensminger AW, Clemson CM, et al. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics*, 2007, 8: 39
- [7] Conley AB, Miller WJ, Jordan IK. Human cis natural antisense transcripts initiated by transposable elements. *Trends Genet*, 2008, 24(2): 53-6
- [8] Pang KC, Frith MC, Mattick JS. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet*, 2006, 22(1): 1-5
- [9] Pollard KS, Salama SR, King B, et al. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet*, 2006, 2(10): e168
- [10] Goodrich JA, Kugel JF. Non-coding-RNA regulators of RNA polymerase II transcription. *Nat Rev Mol Cell Biol*, 2006, 7(8): 612-6
- [11] Sanchez-Elsner T, Gou D, Kremmer E, et al. Noncoding RNAs of trithorax response elements recruit *Drosophila* Ash1 to Ultrabithorax. *Science*, 2006, 311(5764): 1118-23
- [12] Rinn JL, Kertesz M, Wang JK, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, 2007, 129(7): 1311-23
- [13] Boumil RM, Lee JT. Forty years of decoding the silence in X-chromosome inactivation. *Hum Mol Genet*, 2001, 10(20): 2225-32
- [14] Kalantry S, Purushothaman S, Bowen RB, et al. Evidence of Xist RNA-independent initiation of mouse imprinted X-chromosome inactivation. *Nature*, 2009, 460(7255): 647-51
- [15] Ogawa Y, Sun BK, Lee JT. Intersection of the RNA interference and X-inactivation pathways. *Science*, 2008, 320(5881): 1336-41
- [16] Martens JA, Laprade L, Winston F. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* *SER3* gene. *Nature*, 2004, 429(6991): 571-4
- [17] Osato N, Suzuki Y, Ikeo K, et al. Transcriptional interferences in cis natural antisense transcripts of humans and mice. *Genetics*, 2007, 176(2): 1299-306
- [18] Gribnau J, Diderich K, Pruzina S, et al. Intergenic transcription and developmental remodeling of chromatin subdomains in the human β -globin locus. *Mol Cell*, 2000, 5(2): 377-86
- [19] Martianov I, Ramadass A, Serra Barros A, et al. Repression of the human dihydrofolate reductase gene by a noncoding interfering transcript. *Nature*, 2007, 445: 666-70
- [20] Goni JR, de la Cruz X, Orozco M. Triplex-forming oligonucleotide target sequences in the human genome. *Nucleic Acids Res*, 2004, 32(8): 354-60
- [21] Wang X, Arai S, Song X, et al. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature*, 2008, 454(7200): 126-30
- [22] Garen A, Song X. Regulatory roles of tumor-suppressor proteins and noncoding RNAs in cancer and normal cell functions. *Int J Cancer*, 2008, 122: 1687-9
- [23] Li L, Feng T, Lian Y, et al. Role of human noncoding RNAs in the control of tumorigenesis. *Proc Natl Acad Sci USA*, 2009, 106(31): 12956-61
- [24] Feng J, Bi C, Clark BS, et al. The Evt-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev*, 2006, 20(11): 1470-84
- [25] Ashe HL, Monks J, Wijgerde M, et al. Intergenic transcription and transinduction of the human β -globin locus. *Genes Dev*, 1997, 11(19): 2494-509
- [26] Shamovsky I, Ivannikov M, Kandel ES, et al. RNA-mediated response to heat shock in mammalian cells. *Nature*, 2006, 440(7083): 556-60
- [27] Willingham AT, Orth AP, Batalov S, et al. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science*, 2005, 309(5740): 1570-3
- [28] Dieci G, Fiorino G, Castelnuovo M, et al. The expanding RNA polymerase III transcriptome. *Trends Genet*, 2007, 23(12): 614-22
- [29] Yik JH, Chen R, Nishimura R, et al. Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol Cell*, 2003, 12(4): 971-82
- [30] Mattick JS. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *BioEssays*, 2003, 25(10): 930-9
- [31] Beltran M, Puig I, Peña C, et al. A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev*, 2008, 22(6): 756-69
- [32] Calin GA, Liu CG, Ferracin M, et al. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell*, 2007, 12(3): 215-29
- [33] Pibouin L, Villaudy J, Ferbus D, et al. Cloning of the mRNA colon carcinomas. *Cancer Genet Cytogenet*, 2002, 133: 55-60
- [34] Fu X, Ravindranath L, Tran N, et al. Regulation of apoptosis by a prostate-specific and prostate cancer-associated noncoding gene, PCGEM1. *DNA Cell Biol*, 2006, 25(3): 135-41
- [35] Ji P, Diederichs S, Wang W, et al. MALAT-1, a novel noncoding RNA, and thymosin β 4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*, 2003, 22(39): 8031-41
- [36] Yu W, Gius D, Onyango P, et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*, 2008, 451(7175): 202-6
- [37] Amaral PP, Mattick JS. Noncoding RNA in development.

- Mamm Genome, 2008, 19: 454-92
- [38] Blasco MA. Telomere length, stem cells and aging. *Nat Chem Biol*, 2007, 3(10): 640-9
- [39] Azzalin CM, Reichenbach P, Khoriantseva L, et al. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*, 2007, 318 (5851): 798-801
- [40] Schoeftner S, Blasco MA. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol*, 2008, 10(2): 228-36
- [41] Greco-Stewart VS, Thibault CS, Pelchat M. Binding of the polypyrimidine tract-binding protein-associated splicing factor (PSF) to the hepatitis delta virus RNA. *Virology*, 2006, 356(1-2): 35-44