

## 特约综述



我们实验室主要应用生物化学、分子生物学、基因组学以及细胞、干细胞生物学的多种方法,探索哺乳动物细胞中长非编码RNA的产生机制和功能,并着重研究其在细胞核亚结构功能和人源胚胎干细胞命运决定中的分子调控作用。

<http://www.sibcb.ac.cn/PI.asp?id=138>

## 探索长非编码RNA在哺乳动物细胞中的功能秘密

向剑锋 殷庆飞 陈玲玲\*

(分子生物学国家重点实验室,中国科学院上海生命科学研究院上海生物化学与细胞生物学研究所,上海 200031)

**摘要** 近年来,越来越多的长非编码RNA在不同的物种中被相继发现。该文总结了相关领域的最新研究进展,对长非编码RNA在表观遗传学调控、转录调控、microRNA网络调控、细胞核亚结构等方面的功能机制以及其与多种疾病的发生关联进行了简要的总结。

**关键词** 长非编码RNA; 基因表达调控; 细胞核亚结构; sno-lncRNAs

## Long Noncoding RNAs: Regulatory Molecules in Mammalian Cells

Xiang Jianfeng, Yin Qingfei, Chen Lingling\*

(State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China)

**Abstract** It has become apparent that a number of organisms express abundant amounts of long noncoding RNAs (lncRNAs) that lack open reading frames. Rather than accumulating silently in the cell, recent research has shed new light onto the biological significance of lncRNAs. Here, we highlight some recent advances in our understanding of their important roles in nuclear architectures and in regulation of gene expression in mammalian cells.

**Key words** long noncoding RNA; gene expression regulation; nuclear architecture; sno-lncRNAs

### 1 前言

随着基因组测序计划的完成以及新一代深度测序技术的应用,人们发现哺乳动物细胞中多于95%的转录序列为非编码RNA(noncoding RNA, ncRNA)<sup>[1]</sup>。这些非编码RNA包括人们所熟知的“管家(housekeep-

ing)”非编码RNA(如transfer RNAs、ribosomal RNAs和small nuclear RNAs等)、小非编码RNA(如microRNAs和piRNAs等),更多的是尚未深入研究的长非编码RNA(long noncoding RNA, lncRNA)。简单地讲, lncRNA是指一类含有200个核苷酸以上,但是却不具备

中国科学院干细胞先导计划(批准号: XDA01010206)和国家自然科学基金(批准号: 31271376)资助的课题

\*通讯作者。Tel: 021-54921021, E-mail: linglingchen@sibcb.ac.cn

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No.XDA01010206) and the National Natural Science Foundation of China (Grant No.31271376)

\*Corresponding author. Tel: +86-21-54921021, E-mail: linglingchen@sibcb.ac.cn

网络出版时间: 2013-03-04 16:13 URL: <http://www.cnki.net/kcms/detail/31.2035.Q.20130304.1613.010.html>

蛋白质编码能力的长链RNA分子。已知lncRNA在哺乳动物细胞基因组中广泛转录, 例如它们可以转录自基因的上游区域[如PALRs(promoter-associated long RNAs); eRNAs(enhancer RNAs)]<sup>[2-3]</sup>、基因间区域(如long intergenic ncRNAs, lincRNAs)<sup>[4-5]</sup>或与基因形成天然反式转录本(如nature antisense transcripts, NATs)<sup>[6-7]</sup>。另外, 哺乳动物中也发现了来自剪接后内含子区域并具有重要生物学功能的3'末端非polyA结尾的lncRNAs<sup>[8-11]</sup>。最新ENCODE数据分析表明, 人类基因组中含有9 000多种lncRNAs<sup>[12-13]</sup>, 然而生化鉴定和功能研究尚处于起步阶段, 目前仅有大约100种已知功能的lncRNAs。这些lncRNAs可以广泛参与细胞中多种重要的功能调控, 包括对基因表达的遗传与表观遗传调控<sup>[14-16]</sup>、基因表达的转录水平调控<sup>[17-18]</sup>、细胞核亚结构的形成<sup>[19]</sup>(图1, 详述见参考文献[20-21])以及发育的调控<sup>[22-23]</sup>、干细胞多能性<sup>[24-25]</sup>和体细胞重编程的调控<sup>[26]</sup>等。我们将结合近年来的研究成果, 在表观遗传学调控、转录调控、microRNA网络调控、细胞核亚结构以及与疾病的关系等方面, 对长非编码RNA作一简要介绍。

## 2 lncRNA在表观遗传学中的功能

目前, 已知lncRNA可以通过多种方式实现基因表达调控, 其中一种是通过改变表观遗传信息来实现的。人们根据lncRNA与其调控靶基因的相对位置关系将lncRNA调控靶基因的方式分为顺式(in cis)和反式(in trans)两种。转录产生的lncRNA调控自身基因所在染色体上基因表达的方式称为顺式调节; 而将远距离调控其它染色体上基因表达的方式则称为反式调节。目前, 已知通过顺式调控方式发挥功能的有*Xist*<sup>[16]</sup>、*Air*<sup>[14]</sup>、KCNQ1 overlapping transcript 1(*Kcnqlot1*)<sup>[20]</sup>、HOXA transcript at the distal tip(*HOTTIP*)<sup>[27]</sup>、*Mistral*<sup>[23]</sup>等, 这些lncRNAs通过招募其它与染色体修饰相关的蛋白质复合物, 对附近基因实现表观遗传调控(图1A)。通过反式发挥功能的lncRNAs相对较少, 目前已知有重要功能的包括*HOTAIR*<sup>[28]</sup>、*lincRNA-p21*<sup>[29]</sup>等。lncRNA调节基因表达的方式多种多样, 以下我们将对一些代表性的lncRNAs进行简要介绍。

### 2.1 *Xist*

雌性哺乳动物细胞中有两条X染色体, 而雄性动物细胞中只有一条X染色体。为了使X连锁基因

的表达水平与雄性哺乳动物持平, 雌性哺乳动物在胚胎发育早期随机或印记(imprinting)失活一条X染色体, 这种现象被称为“X染色体失活”。研究发现, 多个lncRNA(如*Xist*、*RepA*等)参与了X染色体失活过程<sup>[30]</sup>。X染色体失活前期, X染色体失活中心(X chromosome inactivation center, XIC)转录产生一条长约1.6 Kb的lncRNA *RepA*, 并招募PRC2(polycomb repressive complex 2)到*Xist*基因转录位点附近, 参与*Xist*基因启动子区域的组蛋白H3的27位赖氨酸三甲基化(H3K27me3), 该甲基化方式可促进*Xist*基因的大量表达, 产生的*Xist* RNA通过未知的机制覆盖在整条X染色体上, 并结合大量与基因沉默相关的蛋白质, 介导X染色体失活<sup>[16,20]</sup>。

### 2.2 *Air*和*Kcnqlot1*

lncRNA *Air*和*Kcnqlot1*都产生自印记基因区域, 它们只从父源等位基因转录<sup>[16,31]</sup>, 并对相应基因印记区域的形成至关重要。

*Kcnqlot1*转录自*Kcnql*基因的反义链, 其启动子位于*Kcnql*基因的第10个内含子上, 启动子的DNA甲基化决定了*Kcnqlot1*的转录情况<sup>[32]</sup>。研究表明, *Kcnqlot1*可以结合并招募H3K9-和H3K27-特异的组蛋白甲基转移酶G9a和PRC2, 通过参与组蛋白甲基化修饰, 改变染色质构象, 调节基因表达。*Air*也具有类似的产生机制和功能: 转录自父源*Igf2r*基因反义链的转录本*Air*全长108 Kb, 与*Igf2r*有部分重叠, 主要滞留在细胞核中<sup>[33]</sup>, 可以招募甲基转移酶G9a到靶基因的启动子区域, 介导组蛋白H3K9甲基化, 以组织特异和等位基因特异的方式介导雌鼠胎盘中父源染色体上相关印记基因的沉默<sup>[14]</sup>。

### 2.3 *HOTTIP*

最近研究表明, *HOX*基因簇不仅表达一系列重要的蛋白质, 不同*HOX*基因簇的表达调控与附近lncRNAs的表达也密切相关。例如*HOTTIP*是转录自*HOXA*基因簇的一条lncRNA, 研究发现其可以远距离调控*HOXA*基因簇的表达。*HOTTIP*可结合接头蛋白WDR5, 介导WDR5/MLL复合物定位到*HOXA*基因区域上, 使组蛋白H3K4发生三甲基化, 从而诱导*HOXA*基因簇的表达。*HOTTIP*时空特异性表达对于*HOXA*基因簇的精确调控具有十分重要的作用<sup>[27]</sup>。另外, 在小鼠胚胎干细胞分化过程中, lncRNA *Mistral*也是通过相似的途径对其附近的*HOXA6*和*HOXA7*基因区域发挥正调控作用<sup>[23]</sup>。

### 2.4 *HOTAIR*

与上文中提到的lncRNA不同,转录自*HOXC*基因簇的*HOTAIR*是通过反式调控方式改变表观遗传信息发挥功能的。*HOTAIR*转录自*HOXC*基因簇,在躯体的末端和尾端的细胞中表达。Gupta等<sup>[34]</sup>研究发现,在原发性或转移性乳腺肿瘤细胞中*HOTAIR*也高表达,并促进肿瘤细胞的浸润和转移。转录产生的*HOTAIR*可结合PRC2,通过促进组蛋白H3K27三甲基化,从而抑制基因表达<sup>[28,34]</sup>。Tsai等<sup>[35]</sup>的研究发现,除*HOTAIR* 5'端的300 nt可以与PRC2结合、抑制基因表达外,其3'端的700 nt还可以结合LSD1、CoREST、REST等蛋白质,介导组蛋白H3K4去甲基化,促进基因表达。

### 2.5 *lincRNA-p21*

*LincRNA-p21*位于*CDKN1A(p21)*基因的上游,它的转录受到p53蛋白质的调控。在细胞受到DNA损伤时,p53蛋白质直接结合在*lincRNA-p21*的启动子上,诱导*lincRNA-p21*的表达。产生的*lincRNA-p21*结合hnRNP-K,并改变其功能,参与p53依赖的凋亡应答途径,下调部分基因,发挥负调控功能<sup>[29]</sup>。有意思的是,最新研究证明在缺少RNA结合蛋白质HuR的情况下,细胞质中的*lincRNA-p21*更为稳定,并可以通过碱基互补原则,特异结合一些靶mRNA分子,使得这些mRNA从核糖体上脱落下来,从而抑制mRNA的翻译<sup>[36]</sup>。这些研究也显示,长非编码RNA在细胞内的不同定位决定了其功能的多样性。

综上,很多已知功能的lncRNAs在转录之后,通过靶向结合表观遗传修饰蛋白因子(如PRC2、MLL复合物等),从而对特定基因区域的表观遗传信息进行修饰,使得该染色体区域被抑制或激活,从而影响该区域的基因表达。这些lncRNAs在表观遗传调控中发挥了重要作用,如上文提到的X染色体失活、基因印迹、发育调控以及DNA损伤修复等。

## 3 lncRNA在转录调控中的作用

除在表观遗传水平调控基因表达外,lncRNA还可以在转录水平上直接促进或抑制靶基因的转录。已有研究表明<sup>[21]</sup>,lncRNA不仅可以招募转录抑制复合物沉默基因,还可以作为“配体”招募转录因子,参与基因表达调控(图1B)。

目前,研究得比较清楚的是,lncRNA参与周期蛋白D1(cyclin D1, CCND1)和二氢叶酸还原酶(DHFR)的调控。Wang等<sup>[18]</sup>研究发现,DNA损伤可诱导*CCND1*基因5'端调控区域产生一系列长度不

一的非编码RNAs,这些lncRNAs可结合RNA结合蛋白脂肪肉瘤转运蛋白质(translocated in liposarcoma, TLS),使得TLS的蛋白质结构发生改变,调节TLS的活性,受调节的TLS进而抑制CREB结合蛋白质CBP和P300组蛋白乙酰转移酶的活性,从而抑制CCND1的转录并抑制细胞进入细胞周期。同样是顺式抑制基因转录,转录自*DHFR*基因启动子的lncRNA则以不同的机制参与了该基因的转录抑制。*DHFR*基因上游含有两个启动子,较远的次要启动子(minor promoter)起始转录产生的lncRNA可与较近的主要启动子(major promoter)形成稳定的RNA-DNA复合物,并与转录因子TFIIB直接相互作用,进而解离主要启动子的转录前起始复合物,从而抑制主要启动子起始的转录,抑制*DHFR*基因的表达<sup>[17]</sup>。

lncRNA除可以抑制基因转录外,还可以激活基因转录,如*Evf-2*。*Evf-2*是一条长约3.8 Kb的长非编码RNA,转录自同源结构域蛋白(homeodomain proteins)*Dlx5/6*基因的一个超保守区域,它的表达和*Dlx5/6*基因的表达呈正相关。Feng等<sup>[22]</sup>研究发现,*Evf-2*可以与同源结构域蛋白*Dlx2*形成稳定复合体,共同增加*Dlx5/6*增强子的转录活性,促进*Dlx5/6*基因表达。事实上,脊椎动物*Dlx*基因家族在神经发育、迁移以及面部和四肢发育过程中都起到关键作用。*Evf-2*在*Dlx5/6*基因表达中的功能再一次表明长非编码RNA在哺乳动物发育过程中起到重要的调节作用。

## 4 lncRNA参与microRNA网络调控

由上可知,很多lncRNAs在染色质重塑和基因转录中是通过与蛋白质相互作用发挥功能的。但是,也有一些lncRNAs发挥功能并不依赖这种调节方式,而是通过调控microRNA实现功能。已有研究表明,lncRNA不仅可以作为小RNA的前体<sup>[37]</sup>,还可通过结合microRNA参与细胞中microRNA调控网络,从而调节更多相关功能基因的表达,人们在此基础上提出了ceRNA(competitive endogenous RNA)假说。

CeRNA假说是建立在经典的microRNA调控机制基础之上的。MicroRNA在细胞内主要通过降低mRNA稳定性和抑制蛋白质翻译来抑制靶基因的蛋白质水平。细胞中一个基因往往受多个microRNA调节,一个microRNA也可调控多个基因的表达。CeRNA假说认为,细胞内存在的长链RNA上含有种类和数量不等的microRNA结合位点,含有相同

microRNA结合位点的RNA被称为ceRNA。CeRNA可以相互竞争结合同种microRNA, 其中一些RNA竞争性结合microRNA, 可降低自由分布的microRNA浓度, 从而在一定程度上降低被microRNA调控的基因的抑制程度<sup>[38]</sup>。实验证明, ceRNA可以是假基因(pseudogene)<sup>[39]</sup>, 也可以是lncRNA<sup>[40]</sup>。

研究发现, 在肌肉细胞分化过程中lncRNA *MD-1*受到严格的调控。在未分化的肌细胞中, miR-133和miR-135可分别结合*MAML-1*和*MEF2C*基因的mRNA, 并抑制其表达。*MD-1*含有这两个microRNA的结合位点, 分化过程中高表达的*MD-1*竞争性结合细胞中的miR-133和miR-135, 降低了这两种microRNA的游离浓度, 从而解除了对肌肉分化相关蛋白质因子的抑制作用, 促进细胞分化<sup>[40]</sup>。

细胞中的每条ceRNA都可能含有大量且不同的microRNA结合位点, 可以竞争结合多种microRNA, 同时由于每种microRNA可以调控多种不同的靶基因, 因此每条ceRNA可以通过microRNA的调控网络实现与多种基因调控的“交叉对话(crosstalk)”(图1C)。在这一过程中, lncRNA可以像“海绵”一样, 吸附不同microRNA分子(molecular sponge), 参与microRNA网络调控。同样的, 如果lncRNA序列中含有多个蛋白质结合位点, 该lncRNA还可以吸附相应的蛋白质, 从而调控蛋白质的定位和功能(见本文第6节)。

## 5 lncRNA与细胞核亚结构功能

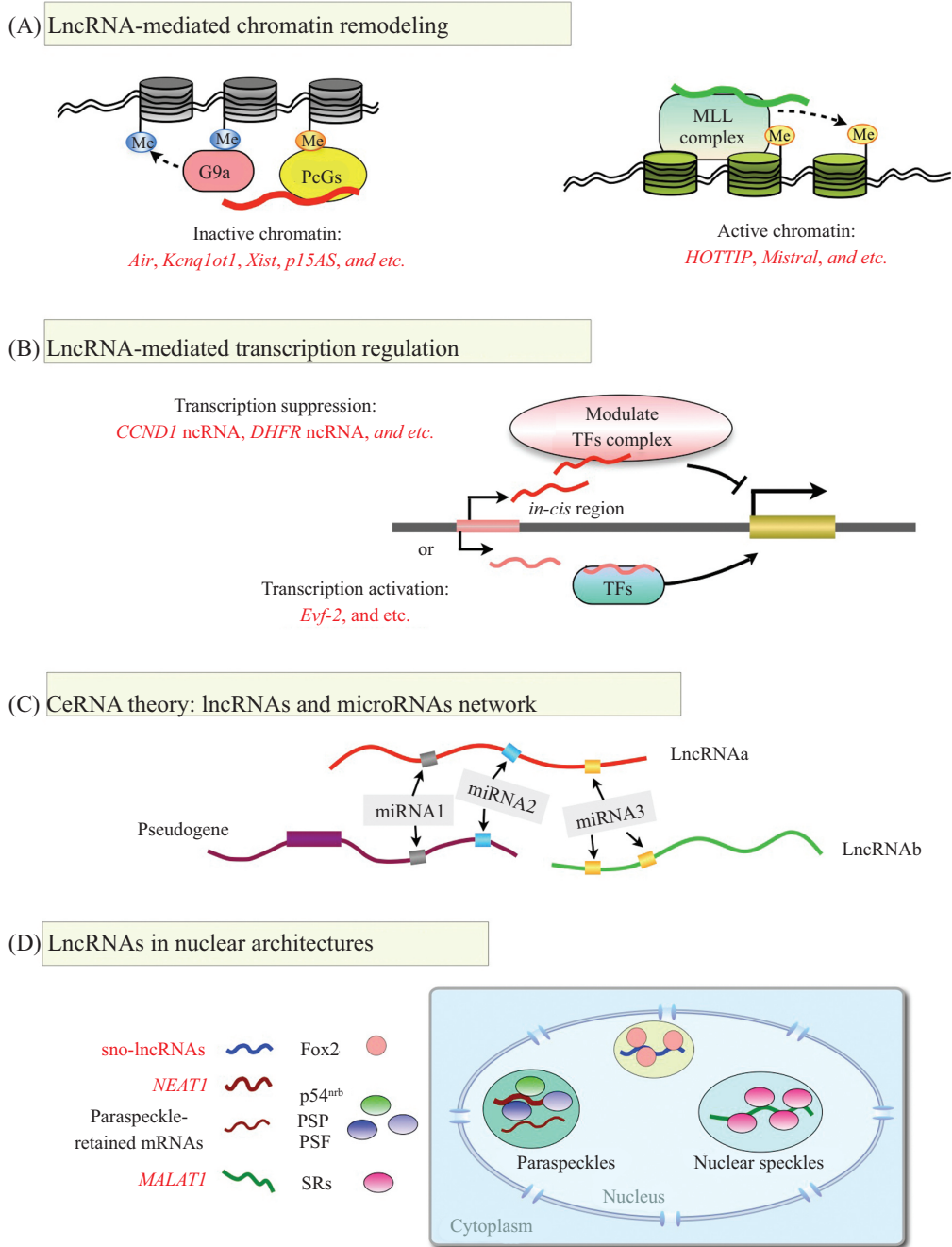
除了上述的功能, 长非编码RNA也参与多种细胞核亚结构的功能调控。已知细胞核是真核细胞内最重要和最复杂的细胞器。近年研究发现, 细胞核内复杂的高级结构远比人们所认识的更为复杂, 其中每条染色体不仅有自己的特定区域(chromosome territories), 还含有多种具有不同功能的亚核结构, 如核仁、splicing speckles、paraspeckles、PML小体(PML bodies)、Cajal小体(Cajal bodies)等。这些细胞核内的亚核结构具有不同的生物学功能, 如核仁参与rRNA的转录、加工成熟以及核糖体亚单位的组装; paraspeckles的形成依赖于长非编码RNA, 并且可以调控转录后mRNA的细胞核滞留; splicing speckles参与mRNA的可变剪接调控; Cajal小体参与了snRNA的转录后修饰以及snRNP和snoRNP的产生等<sup>[41]</sup>。最近研究发现, lncRNA对于细胞核亚结构的形成和功能具有十分重要的作用(图1D)。

### 5.1 *NEATI*和paraspeckles

细胞核亚结构paraspeckles于2002年被发现并定义, 其标志蛋白质有p54<sup>nrb</sup>、PSF和PSP1 $\alpha$ 等<sup>[42]</sup>。目前, 已知除人源胚胎干细胞(human embryonic stem cells, hESCs)外, paraspeckles在高等哺乳动物的多种细胞系和组织中都广泛存在, 不同细胞中含有5-20个paraspeckles, 且具有一定的组织特异性<sup>[19,43]</sup>。

早在2005年, Lamond等<sup>[44]</sup>研究表明, RNase处理可破坏细胞中paraspeckles结构, 这提示paraspeckles是一种RNA依赖的亚核结构。随后的研究证明, lncRNA *NEATI*(又称*Men  $\epsilon/\beta$* )参与了paraspeckles的组装<sup>[19,45-47]</sup>。人*NEATI*(nuclear enriched abundant transcript 1)转录自第11号染色体, 可以产生两条长非编码RNA的异构体, 长度分别约为3.7 Kb和20 Kb。敲除*NEATI*虽然不影响paraspeckles蛋白质组分(如p54<sup>nrb</sup>、PSF等)的表达, 却改变了这些蛋白质在细胞核中的分布以及paraspeckles的形成。这些实验结果表明*NEATI*在paraspeckles的形成过程中起到了关键作用。应用活细胞成像技术等直接证据, Mao等<sup>[48]</sup>进一步证实了不仅*NEATI*、而且*NEATI*的转录对于新生paraspeckles的组装至关重要。

有关paraspeckles的功能研究, Spector实验室和Carmichael实验室各自独立证明了paraspeckles参与了一些mRNA转录后的出核活动。哺乳动物细胞中部分mRNA 3'UTR中含有反向重复*Alu*元件(inverted repeated *Alus*, *IRAlus*), *IRAlus*可以形成分子内的双链RNA, 可以被ADAR酶编辑产生从腺氨酸(adenosine, A)到次黄嘌呤核氨酸(inosine, I)的修饰(A-to-I)。Chen等<sup>[49]</sup>研究发现, 这种在3'UTR的编辑直接导致编辑后的mRNA通过结合核蛋白质p54<sup>nrb</sup>复合体, 滞留在paraspeckles中。这是一个全新水平的调控, 即成熟的mRNA在转录和编辑之后不立即出核, 而是滞留在细胞核中, 不能进行翻译而导致相应蛋白质的低表达, 从而达到调控基因表达的目的。综上所述, *NEATI*不仅对细胞核亚结构paraspeckles的形成具有决定作用, 同时还调控一系列3'UTR具有*IRAlus*元件的mRNA在细胞核内的滞留<sup>[19]</sup>。更有意思的是, 虽然*NEATI*在人体的各种组织和细胞中广泛表达, 但是在未分化的人源胚胎干细胞中表达极低, 而定向分化可诱导*NEATI*表达量显著提高, paraspeckle结构重新形成。值得关注的是, 在未分化的人源胚胎干细胞中, 由于缺乏*NEATI*和paraspeckles, 一些对于干细胞多能性维持具有重要作用的mRNA, 例如



A: lncRNA参与表观遗传调控; B: lncRNA参与转录调控; C: lncRNA参与microRNA网络调控; D: lncRNA参与细胞核亚结构功能。

A: lncRNA-mediated heterochromatin silencing; B: lncRNA-mediated transcription regulation; C: lncRNA and microRNA regulation network; D: lncRNA in the regulation of nuclear architecture integrity and function.

图1 长非编码RNA参与细胞的重要功能调控

Fig.1 Long noncoding RNAs (lncRNAs) play important roles in mammalian cells

Lin28(3'UTR含有核滞留IRAlus序列)可以在转录之后快速出核,翻译产生干细胞所需要的LIN28蛋白质<sup>[9]</sup>,从而提示NEAT1不仅在核亚结构形成中具有决定作用,而且对于人源胚胎干细胞的命运决定具有重要的调控作用。

Paraspeckles介导的mRNA核滞留不仅在人源

细胞中存在,在小鼠细胞中也存在类似的调控机制。研究发现,小鼠mCAT2基因通过可变剪切可以产生两种具有不同长度3'UTR的异构体,其中mCAT2的3'UTR较短,而CTN-RNA(尽管具有与mCAT2一样的编码框)的3'UTR较长,并可以被高度A-to-I编辑修饰,导致其滞留在细胞核的paraspeckles上。当遇

到环境胁迫时, 积聚在paraspeckles上的CTN-RNA 3' UTR(含有高度的A-to-I编辑修饰)以未知的方式被除去, 与*mCAT2*一样的编码框并被转运到细胞浆中, 快速编码mCAT2蛋白质以响应环境胁迫<sup>[50]</sup>。

尽管在细胞水平上*NEAT1*和paraspeckle被证明参与基因表达调控, 然而*NEAT1*基因缺失的小鼠虽然不形成paraspeckle结构, 却也没有明显的表型<sup>[51]</sup>, 从而提示这些调控作用对于小鼠的正常生命活动可能不是必需的。由于*Alu*元件是灵长类哺乳动物所特有的序列, *NEAT1*和paraspeckle所介导的3'UTR含有*IRAlus*元件的mRNA核滞留在人源细胞中更为广泛, 因此*NEAT1*和paraspeckle在人类基因表达调控中的作用可能比小鼠中重要。除了上述机制外, *NEAT1*和paraspeckles是否还具有其它的功能还有待研究。进一步深入研究paraspeckles参与调控的靶基因, 对于了解*NEAT1*的功能具有十分重要的作用。

## 5.2 MALAT1和splicing speckles

*MALAT1*(metastasis-associated lung adenocarcinoma transcript 1), 又名*NEAT2*(nuclear enriched abundant transcript 2), 转录自*NEAT1*基因下游约58 Kb区域, 于2007年发现特异定位在splicing speckles上<sup>[52-53]</sup>。Splicing speckles又名nuclear speckles, 是细胞核中RNA剪接因子(RNA splicing machinery)和多种蛋白质编码基因的mRNA前体分子(pre-mRNA)聚集形成的一类亚核结构, 参与RNA的剪接调节<sup>[54]</sup>。与*NEAT1*不同的是, *MALAT1*的缺失并不影响splicing speckles的形成<sup>[55-56]</sup>。有实验表明, 定位在splicing speckles上的*MALAT1*可以结合磷酸化的剪接因子, 调节这些剪接因子在splicing speckles以及细胞核内的分布, 进而影响某些mRNA的可变剪接<sup>[55]</sup>。Guo等<sup>[53]</sup>在CaSki细胞中的研究表明, *MALAT1*也可能通过调控凋亡途径相关基因参与到肿瘤细胞生长、浸润等生物学过程中。在非小细胞肺癌的研究中也发现沉默*MALAT1*可明显抑制非小细胞肺癌肿瘤细胞的迁移和肿瘤形成, 而过量表达*MALAT1*则具有促进作用<sup>[57]</sup>。这些研究均证实*MALAT1*在肿瘤的发生和发展中具有重要功能。除在肿瘤中具有重要作用外, *MALAT1*在神经系统突触形成和维持过程中也有作用<sup>[58]</sup>。然而有意思的是, Zhang等<sup>[56]</sup>最近制备了*MALAT1*基因敲除的小鼠, 该小鼠能够正常生长和发育, 不具有明显的表型; 同时研究表明, *MALAT1*基因的缺失并不直接影响mRNA的可变剪接活动,

但是对于包括*NEAT1*在内的一些附近基因的表达具有微弱的顺式调控作用。这些初步研究均提示, *MALAT1*具有复杂的生物学功能, 但是对于具体的分子机制尚需进一步探讨。

## 5.3 其它细胞核亚结构相关的lncRNAs

正如蛋白质发挥功能与其在细胞内的定位密不可分一样, 很多lncRNA也需定位在细胞核中特定位置发挥其功能。除了上述*NEAT1*和*MALAT1*外, 如*Xist*、*Air*等在转录后也分别定位于失活的X染色体或*Igf2r*基因印迹区域附近发挥各自的功能<sup>[20]</sup>。哺乳动物染色质末端的端粒重复序列*TARRA*(telomeric repeat-containing RNA)转录后积聚在端粒附近, 与RNA质量监控因子(RNA surveillance factors)相互作用, 对于端粒复制和长度的维持非常关键<sup>[59]</sup>。在细胞应激状态下, 微卫星DNA III(satellite DNA III, *Sat III*)可以转录出*Sat III*, *Sat III*招募一些与转录和剪接相关的蛋白质因子形成细胞核应激小体(nuclear stress bodies); 沉默*Sat III*可以破坏细胞核应激小体结构, 提示*Sat III*在细胞核应激小体过程中发挥作用<sup>[60]</sup>。同样在应激情况下, 核仁基因间区转录产生一类定位在核仁上的lncRNA, 这些lncRNA可以选择性结合VHL、HSP70和MDM2/PML等含有NoDS结构域的蛋白质, 并将其锚定在核仁区域, 从而通过改变这些功能蛋白质在细胞中的定位实现对蛋白质功能的调控<sup>[61]</sup>。

这些结果表明, lncRNA既可以通过与DNA或者蛋白质等相互作用, 参与细胞核亚结构的形成, 又可以定位在特定的细胞核亚核结构中, 通过调节与之结合的蛋白质因子来实现功能。目前, 人们提出细胞核亚核结构的形成有两种模式, 即随机组装模式(random self-organization)和序列组装模式(ordered stepwise assembly)<sup>[62]</sup>。随机组装模式认为细胞核亚核结构的组成蛋白质在细胞核中随机碰撞并结合, 共同形成完整的亚核结构; 而序列组装模式认为, 细胞核亚核结构的组成成分严格按照特定的顺序依次结合, 共同组成完整的亚核结构。两种不同的形成模式均有实验证据支持, 而有关lncRNA参与细胞核亚核结构形成的研究更倾向于支持后者, 基因转录产生的lncRNA作为“种子(seed)”, 作为核心成分依次结合相应的蛋白质, 共同起始细胞核亚核结构的形成<sup>[63]</sup>。

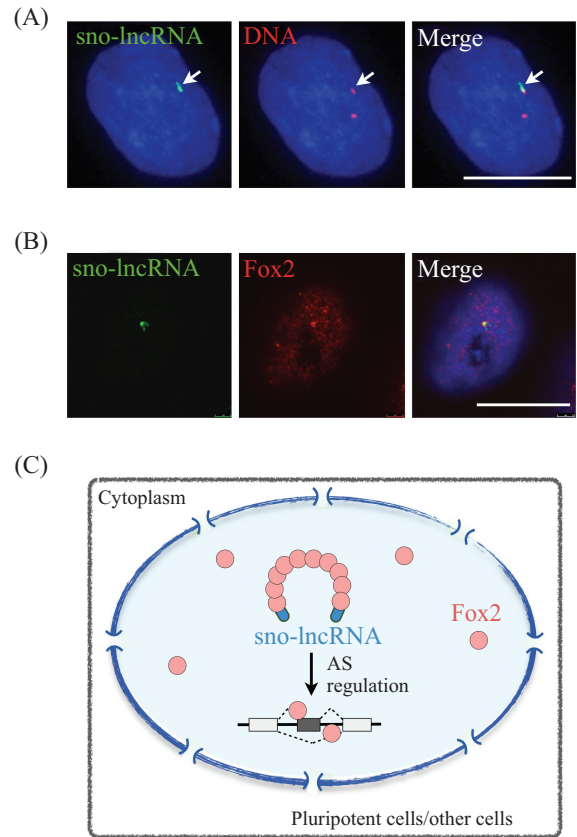
## 6 内含子来源的新类型长非编码RNA形

## 成细胞核亚定位, 并参与哺乳细胞内重要的功能调控

随着对基因组转录调控认识的不断深入, 人们发现大多数长非编码RNA可以像mRNA一样由RNA聚合酶II转录, 形成5'-帽子(cap)和3'-尾巴(polyA tail)的经典结构<sup>[12]</sup>; 而2012年9月发表的ENCODE数据表明, 人基因组已定义的9 640条lncRNAs中, 只有16.8%的lncRNAs含有polyA尾巴<sup>[13]</sup>。例如, 具有功能的NEAT1长异构体和MALAT1等就不具有典型的polyA结构, 而是由核酶RNase P剪切而来<sup>[37,45]</sup>。我们实验室研究发现, 内含子来源的非编码RNA序列既可以在剪接后稳定存在<sup>[11]</sup>, 又可以在细胞中发挥重要的调控作用<sup>[10]</sup>。几乎所有哺乳动物细胞的基因都由外显子和内含子组成。一般认为, 外显子片段通过转录剪接加工, 生成具有功能的成熟RNA, 而内含子序列在剪接后被核酸酶快速降解, 因此没有生物学功能。我们的最新研究表明, 一类双末端都含有小核仁RNA(snoRNA)的内含子序列在剪接发生过程中可以形成一类新型长非编码RNA, 命名为sno-lncRNA<sup>[10]</sup>。

在人类基因疾病小胖威利综合征(Prader-Willi syndrome, PWS)紧密关联区域存在五个sno-lncRNAs, 它们在人源胚胎干细胞中表达量极高, 有些sno-lncRNA的表达甚至与持家基因(如*Gapdh*)的RNA相似, 是目前已知的未分化人源胚胎干细胞中表达最为丰富的lncRNA, 这些sno-lncRNAs的成熟依赖于snoRNA的加工机制。功能研究表明, 尽管含有snoRNA的结构, sno-lncRNAs在细胞内并不与经典的snoRNA共定位, 而是在成熟后聚集于其转录位点附近, 形成一种全新的细胞核亚定位(图2A)。PWS综合征区域的sno-lncRNAs还含有多个剪接调控因子Fox2蛋白的特异结合位点, 从而可以像“海绵”一样吸附细胞核内的Fox2, 调节Fox2在细胞核内的局部浓度(图2B), 进而影响Fox2对特异mRNA底物的可变剪接调控(图2C)<sup>[10]</sup>。这些研究进一步表明, 与MALAT1<sup>[55]</sup>和*Sat III*<sup>[60]</sup>类似, sno-lncRNAs作为“分子海绵”可以结合特异蛋白质, 通过改变相应蛋白质的细胞定位调控其功能, 这也是lncRNA发挥功能的重要途径之一。

除了PWS区域的五个sno-lncRNAs之外, 我们发现基因组中的其它内含子区域也含有能够形成稳定的sno-lncRNAs, 或者能够形成稳定存在的, 但是不具有类似结构的非编码RNA(未发表数据)。除了



A: sno-lncRNAs形成特殊细胞核定位(绿色), 并积聚在其父系染色体基因转录位点附近(红色); B: 部分剪接调控因子Fox2蛋白(红色)积聚在sno-lncRNAs(绿色)附近; C: sno-lncRNAs参与可变剪接调控示意图。标尺=10 μm。

A: sno-lncRNAs (green) accumulate at a single chromosomal locus to its adjacent DNA region (red); B: sno-lncRNAs (green) localize to the enriched Fox2 accumulations (red) in the nucleus; C: a model for the role of sno-lncRNAs in normal cells. Bar=10 μm.

图2 Sno-lncRNAs在细胞核内的定位及功能(根据参考文献[10]修改)

Fig.2 Sno-lncRNAs nuclear localization and function(modified from reference [10])

以snoRNA结尾的长非编码RNA外, 最近研究还发现了稳定的环形非编码RNA结构<sup>[64]</sup>。这些不同分子结构形式的长非编码RNA分子具有显著的细胞和组织表达特异性, 提示其重要的生物学功能。然而, 内含子剪接后如何抵抗核酸酶降解而形成稳定的长非编码RNA, 以及它们在细胞内的定位和功能都有待于进一步研究。

## 7 长非编码RNA在生理和病理中的作用

目前已有的研究表明, lncRNA在正常的生命过程中发挥重要的作用, 包括干细胞多能性维持和重编程、细胞分化、以及生物体发育等; 同时也与一

些人类的重大疾病直接相关, 包括癌症、肌肉萎缩、PWS等。

例如, 雌性哺乳动物中的一条X染色体在胚胎发育的早期失活, lncRNA *RepA*和*Xist*均在X染色体失活的起始和维持中发挥重要作用<sup>[16]</sup>。除X染色体失活外, Rinn等<sup>[28]</sup>在4个*HOX*基因簇发现并确定了231个*HOX* ncRNA, 这些ncRNA的空间表达是按着胚胎发育的头尾轴分布的, 提示它们与胚胎发育的调控相关。Feng等<sup>[22]</sup>研究发现, lncRNA *Eyf-2*参与神经发育和模式形成的调控作用。部分lncRNA通过基因印记发挥发育调控作用, 如*Kcnq1ot1*<sup>[15,65]</sup>、*Air*<sup>[14]</sup>等。在哺乳动物胎盘组织中, *Kcnq1ot1*以种系特异的方式抑制*Kcnq1*基因簇的转录, 并维持*Kcnq1*基因簇的沉默; 父本来源的lncRNA *Air*以等位基因特异的方式顺式沉默老鼠胎盘中发育相关的*Slc22a3*、*Slc22a2*和*Igf2r*基因。这些lncRNA以调控表观遗传和转录等方式(见本文第2、3节)参与了这些重要的生长发育过程的调控。

除参与发育调控外, lncRNA在细胞干性维持和分化过程中也发挥着重要作用。位于*HOXA1*和*HOXA2*基因之间的lncRNA *HOTAIRMI*通过调节*HOXA*基因簇基因表达参与骨细胞生成<sup>[66]</sup>。Cesana等<sup>[40]</sup>发现了骨骼肌特异的lncRNA *linc-MD1*, 通过调节miR-133和miR-135的分布间接调节转录因子MAML1和MEF2C的表达, 从而参与骨骼肌特异基因的调控。Bertani等<sup>[23]</sup>发现的lncRNA *Mistral*可与MLL复合体作用激活附近*Hoxa6*和*Hoxa7*基因的转录, 进而调控在mESCs分化过程中胚层分化相关基因的表达。Dinger等<sup>[67]</sup>鉴定出在小鼠胚胎干细胞(mESCs)向拟胚体分化过程中, 一些lncRNAs与干细胞多能性或定向分化有关。Amaral等<sup>[68]</sup>发现lncRNA *Sox2ot*的转录区域与多能性关键因子*SOX2*的正义链有重叠。*Sox2ot*在mESCs中稳定表达, 而在分化过程中表达发生变化。Guttman等<sup>[24]</sup>通过RNAi技术大规模研究了mESCs的基因间lncRNAs(lincRNAs)的功能, 发现一些lincRNAs的敲除调控相关干性基因的表达, 进而干扰干细胞的多能性, 并认为这些lincRNA在小鼠干细胞多能性的维持方面发挥着指挥协调作用。

最近的研究证明, lncRNA对hESCs的干性维持和分化以及体细胞重编程等也具有重要的调控作用。例如, 2009年研究揭示了lncRNA *NEAT1*在细胞核亚结构paraspeckles形成以及mRNA核滞留

等途径的关键作用, 同时指出*NEAT1*和paraspeckles在hESCs中的缺失对干细胞多能性维持有调控作用<sup>[19]</sup>。2010年Daley和Rinn实验室合作发现很多lncRNAs在重编程细胞中特异高表达, 并提出lncRNAs(如*linc-RoR*等)可以通过抑制p53的激活等途径促进体细胞重编程的进程<sup>[26]</sup>。2011年, Ng等<sup>[25]</sup>进一步发现hESCs特异高表达的lncRNAs, 它们可能与染色体变构蛋白质相互作用, 影响hESCs的干性调控以及分化。

LncRNA不仅在正常的细胞分化和机体发育中具有重要的生理作用, 而且也在一些疾病中异常表达<sup>[69]</sup>。例如, *MALAT1*的高表达与非小细胞肺癌、子宫内膜基质瘤和肝癌正相关<sup>[70-72]</sup>。转录自*Igf2*位点的反义lncRNA *PEG8/IGF2AS*在Wilms肿瘤中高表达, 是正常肾组织表达量的十到上百倍<sup>[73]</sup>。p15是与白血病相关的周期蛋白质依赖激酶抑制因子, *p15*基因可以产生反义转录本*p15AS*。Yu等<sup>[74]</sup>发现, 在白血病中*p15AS*与*p15*表达呈现反相关性, 通过外源表达*p15AS*可以使染色质发生异染色质化, 沉默p15的S表达, 从而引起细胞的异常增殖, 形成肿瘤。

除了肿瘤, 在一些发育和神经性疾病中也发现了lncRNA的异常表达。Faghihi等<sup>[75]</sup>的研究表明, 在包括A $\beta$ 分泌肽在内的几种刺激下,  $\beta$ 分泌素1(beta secretase 1, BACE1)的自然反义链转录本*BACE1-AS*在Alzheimer病人的几个脑区中高表达, 并迅速可逆地上调BACE1, 增加A $\beta$ 分泌肽, 促进Alzheimer疾病的发生。lncRNA也被证明与面肩肱型肌营养不良(facioscapulohumeral muscular dystrophy, FSHD)的病理机制有关。Cabianca等<sup>[76]</sup>研究发现, 染色质相关的lncRNA *DEB-T*参与了FSHD疾病的发生。FSHD病人中4q35区域中D4Z4重复序列的缩短, 导致了4q35区域不能有效结合转录抑制复合物PcG, 进而引起*DBE-T*的转录, *DBE-T*又进一步募集ASHL蛋白质, 改变4q35区域的染色质构象, 最终引起4q35区域基因表达的改变。另外, PWS综合征是由第15号染色体长臂近中央关键区(15q11.2-12)微缺失引起的, 目前致病的具体分子机制还不清楚。我们实验室发现5个来源于内含子区域的sno-lncRNAs正好位于该PWS综合征染色体缺失区域。在正常人群中, 这些sno-lncRNAs在胚胎发育早期(人源胚胎干细胞)高表达, 因为含有Fox剪接因子的多个结合位点, 可以作为“分子海绵”结合大量Fox蛋白因子, 改变Fox的细胞核内定位和局部浓度, 调节Fox底物mRNA的可变剪接模式(图



2C); 而在PWS综合征病人的发育早期, 由于没有这些*sno-lncRNAs*分子, Fox蛋白在细胞核内的分布出现异常, Fox底物mRNA的可变剪接模式也因此出现异常。因此, 这些*sno-lncRNAs*在PWS综合征病人中的缺失或不表达, 有可能部分导致了PWS综合征的症状发生<sup>[10]</sup>。这些lncRNA的相关研究为人们进一步认识各种重要疾病提供了新的研究方向。

## 8 总结与展望

最新的ENCODE数据显示, 人类基因组中含有约9 640种lncRNAs<sup>[13]</sup>。虽然lncRNA不具有蛋白质编码的能力, 但越来越多的实验结果发现, lncRNA参与了哺乳动物细胞中许多重要且复杂的功能调控。尽管人们对lncRNA的研究已经取得了一定的进展, 但该领域仍然面临着一些基本问题亟待解答。

首先, lncRNA和蛋白质的相互作用关系。这包含两个方面: lncRNA与哪些特定的蛋白质相互作用及其分子基础, 以及lncRNA与蛋白质的相互作用到底具有何种功能并如何发挥功能。另外, 很多lncRNA可以介导其相互结合蛋白因子的正确定位(如特定基因调控区域), 或通过改变蛋白因子构象来增强或抑制其活性。因此, 研究lncRNA和蛋白因子的相互作用对深入认识lncRNA的功能非常重要。

其次, lncRNA参与染色质形成和染色体高级结构组装的分子机制。早期经典的细胞生物学研究发现RNA是染色质的组成成分, 然而由于技术手段的限制, 对于RNA参与染色体形成的研究一直未取得进展。最近的一些研究发现, lncRNA可改变通过表观遗传信息进而改变靶基因的表达; Blower等<sup>[77]</sup>研究也提示, RNA参与了细胞纺锤体组装等细胞周期活动。因此, 在这些方面的功能探索也将是lncRNA研究的重点之一。

第三, lncRNA与干细胞干性维持和分化的联系。研究表明, 在人和鼠的干细胞分化过程中lncRNA的表达谱都发生了变化, 分化过程中存在上调和下调的lncRNA, 然而它们的产生机制、亚细胞定位、生理功能等都还不清楚。因此, 研究lncRNA在细胞分化过程中扮演的角色, 以及其在发育过程中的作用将是lncRNA研究的重要方面。

第四, lncRNA与疾病的关系。目前, 在癌症、Alzheimer疾病、PWS综合征相关区域等多种疾病中都检测到了一些lncRNA的异常表达, 因此研究疾病发生过程中lncRNA的表达和变化对疾病研究

具有十分重要的作用, 并提示其可作为相关疾病的RNA分子标记物。但lncRNA在这些疾病产生过程中的分子机制仍有待研究, 目前仅有少数疾病相关的lncRNA在分子水平进行了研究, 考虑到每种疾病的差异性, 还需要更多的分子水平研究以解开谜团。

第五, 基因组中特殊来源lncRNA的发现和探究。目前, 人们发现的lncRNA的主要转录自基因间区和基因的反义转录本, 越来越多的研究发现细胞内还存在一些特殊来源的lncRNA, 如产生自基因内含子的lncRNA(如*sno-lncRNAs*)。然而, 由于分析的限制, 细胞内含有重复序列(如*Alu*元件)的RNA(包括mRNA和lncRNA)往往不能被人们发现, 因此应用新的技术手段和分析方法来研究这些含有重复序列的ncRNA基因也是未来的研究方向之一。

综上所述, lncRNA是一个全新的研究领域, 对其功能的研究更是领域内的热点, 有许多基本问题亟待解决, 这需要更多的科研投入和参与。更为重要的是, 对lncRNA在人类疾病中的调控作用研究, 将为相关疾病的深入认识提供新的RNA分子水平的基础, 进而为维护人类健康和治疗疾病提供新的思路和方法。

## 参考文献 (References)

- 1 Pennisi E. Shining a light on the genome's "dark matter". *Science* 2010; 330(6011): 1614.
- 2 Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 2007; 316(5830): 1484-8.
- 3 Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* 2010; 465(7295): 182-7.
- 4 Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 2009; 458(7235): 223-7.
- 5 Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA* 2009; 106(28): 11667-72.
- 6 Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, *et al.* Antisense transcription in the mammalian transcriptome. *Science* 2005; 309(5740): 1564-6.
- 7 Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, *et al.* The transcriptional landscape of the mammalian genome. *Science* 2005; 309(5740): 1559-63.
- 8 Rearick D, Prakash A, McSweeney A, Shepard SS, Fedorova L, Fedorov A. Critical association of ncRNA with introns. *Nucleic Acids Res* 2011; 39 (6): 2357-66.

- 9 Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* 2012; 7(2): e30733.
- 10 Yin QF, Yang L, Zhang Y, Xiang JF, Wu YW, Carmichael GG, *et al.* Long noncoding RNAs with snoRNA ends. *Mol Cell* 2012; 48(2): 219-30.
- 11 Yang L, Duff MO, Graveley BR, Carmichael GG, Chen LL. Genomewide characterization of non-polyadenylated RNAs. *Genome Biol* 2011; 12(2): R16.
- 12 Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* 2011; 25(18): 1915-27.
- 13 Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, *et al.* Landscape of transcription in human cells. *Nature* 2012; 489(7414): 101-8.
- 14 Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, *et al.* The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 2008; 322(5908): 1717-20.
- 15 Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, *et al.* Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 2008; 32(2): 232-46.
- 16 Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 2008; 322(5902): 750-6.
- 17 Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 2007; 445(7128): 666-70.
- 18 Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, *et al.* Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 2008; 454(7200): 126-30.
- 19 Chen LL, Carmichael GG. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: Functional role of a nuclear noncoding RNA. *Mol Cell* 2009; 35(4): 467-78.
- 20 Chen LL, Carmichael GG. Long noncoding RNAs in mammalian cells: What, where, and why? *Wiley Interdiscip Rev RNA* 2010; 1(1): 2-21.
- 21 Chen LL, Carmichael GG. Decoding the function of nuclear long non-coding RNAs. *Curr Opin Cell Biol* 2010; 22(3): 357-64.
- 22 Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD. The Evf-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.
- 23 Bertani S, Sauer S, Bolotin E, Sauer F. The noncoding RNA Mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin. *Mol Cell* 2011; 43(6): 1040-6.
- 24 Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 2011; 477(7364): 295-300.
- 25 Ng SY, Johnson R, Stanton LW. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J* 2012; 31(3): 522-33.
- 26 Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, *et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet* 2010; 42(12): 1113-7.
- 27 Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 2011; 472(7341): 120-4.
- 28 Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 2007; 129(7): 1311-23.
- 29 Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 2010; 142(3): 409-19.
- 30 Pontier DB, Gribnau J. *Xist* regulation and function explored. *Hum Genet* 2011; 130(2): 223-36.
- 31 Sleutels F, Zwart R, Barlow DP. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 2002; 415(6873): 810-3.
- 32 Mancini-Dinardo D, Steele SJ, Levorso JM, Ingram RS, Tilghman SM. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes Dev* 2006; 20(10): 1268-82.
- 33 Seidl CI, Stricker SH, Barlow DP. The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *EMBO J* 2006; 25 (15): 3565-75.
- 34 Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010; 464(7291): 1071-6.
- 35 Tsai MC, Manor O, Wan Y, Mosammamaparast N, Wang JK, Lan F, *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 2010; 329(5992): 689-93.
- 36 Yoon JH, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, *et al.* LincRNA-p21 suppresses target mRNA translation. *Mol Cell* 2012; 47(4): 648-55.
- 37 Wilusz JE, Freier SM, Spector DL. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* 2008; 135(5): 919-32.
- 38 Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the rosetta stone of a hidden RNA language? *Cell* 2011; 146(3): 353-8.
- 39 Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 2010; 465(7301): 1033-8.
- 40 Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, *et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 2011; 147(2): 358-69.
- 41 Spector DL. SnapShot: Cellular bodies. *Cell* 2006; 127(5): 1071.
- 42 Fox AH, Lam YW, Leung AK, Lyon CE, Andersen J, Mann M, *et al.* Paraspeckles: A novel nuclear domain. *Curr Biol* 2002; 12(1): 13-25.
- 43 Fox AH, Lamond AI. Paraspeckles. *Csh Perspect Biol* 2010; 2(7): a000687.
- 44 Lamond AI, Fox AH, Bond CS. P54nrB forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. *Mol Biol Cell* 2005; 16(11): 5304-15.
- 45 Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS,

- Spector DL. MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res* 2009; 19(3): 347-59.
- 46 Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, *et al.* An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell* 2009; 33(6): 717-26.
- 47 Sasaki YT, Ideue T, Sano M, Mituyama T, Hirose T. MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc Natl Acad Sci USA* 2009; 106(8): 2525-30.
- 48 Mao YS, Sunwoo H, Zhang B, Spector DL. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat Cell Biol* 2011; 13(1): 95-101.
- 49 Chen LL, DeCervo JN, Carmichael GG. Alu element-mediated gene silencing. *EMBO J* 2008; 27(12): 1694-705.
- 50 Prasanth KV, Prasanth SG, Xuan ZY, Hearn S, Freier SM, Bennett CF, *et al.* Regulating gene expression through RNA nuclear retention. *Cell* 2005; 123(2): 249-63.
- 51 Nakagawa S, Naganuma T, Shioi G, Hirose T. Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *J Cell Biol* 2011; 193(1): 31-9.
- 52 Hutchinson JN, Ensminger AW, Clemson CM, Lynch CR, Lawrence JB, Chess A. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* 2007; 8: 39.
- 53 Guo F, Li Y, Liu Y, Wang J, Li G. Inhibition of metastasis-associated lung adenocarcinoma transcript 1 in CaSki human cervical cancer cells suppresses cell proliferation and invasion. *Acta Biochim Biophys Sin (Shanghai)* 2010; 42(3): 224-9.
- 54 Lamond AI, Spector DL. Nuclear speckles: A model for nuclear organelles. *Nat Rev Mol Cell Biol* 2003; 4(8): 605-12.
- 55 Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* 2010; 39(6): 925-38.
- 56 Zhang B, Arun G, Mao YS, Lazar Z, Hung G, Bhattacharjee G, *et al.* The lncRNA malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep* 2012; 2(1): 111-23.
- 57 Schmidt LH, Spieker T, Koschmieder S, Humberg J, Jungen D, Bulk E, *et al.* The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. *J Thorac Oncol* 2011; 6(12): 1984-92.
- 58 Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, Xuan Z, *et al.* A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J* 2010; 29(18): 3082-93.
- 59 Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 2007; 318(5851): 798-801.
- 60 Valgardsdottir R, Chiodi I, Giordano M, Cobianchi F, Riva S, Biamonti G. Structural and functional characterization of noncoding repetitive RNAs transcribed in stressed human cells. *Mol Biol Cell* 2005; 16(6): 2597-604.
- 61 Yang LQ, Lin CR, Liu W, Zhang J, Ohgi KA, Grinstein JD, *et al.* ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* 2011; 147(4): 773-88.
- 62 Mao YS, Zhang B, Spector DL. Biogenesis and function of nuclear bodies. *Trends Genet* 2011; 27(8): 295-306.
- 63 Caudron-Herger M, Rippe K. Nuclear architecture by RNA. *Curr Opin Genet Dev* 2012; 22(2): 179-87.
- 64 Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genetics* 2010; 6(12): e1001233.
- 65 Terranova R, Yokobayashi S, Stadler MB, Otte AP, van Lohuizen M, Orkin SH, *et al.* Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev Cell* 2008; 15(5): 668-79.
- 66 Zhang X, Lian Z, Padden C, Gerstein MB, Rozowsky J, Snyder M, *et al.* A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* 2009; 113(11): 2526-34.
- 67 Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, *et al.* Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res* 2008; 18(9): 1433-45.
- 68 Amaral PP, Neyt C, Wilkins SJ, Askarian-Amiri ME, Sunkin SM, Perkins AC, *et al.* Complex architecture and regulated expression of the Sox2ot locus during vertebrate development. *RNA* 2009; 15(11): 2013-27.
- 69 Prasanth KV, Spector DL. Eukaryotic regulatory RNAs: An answer to the "genome complexity" conundrum. *Genes Dev* 2007; 21(1): 11-42.
- 70 Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, *et al.* MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 2003; 22(39): 8031-41.
- 71 Lin R, Maeda S, Liu C, Karin M, Edgington TS. A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. *Oncogene* 2007; 26(6): 851-8.
- 72 Yamada K, Kano J, Tsunoda H, Yoshikawa H, Okubo C, Ishiyama T, *et al.* Phenotypic characterization of endometrial stromal sarcoma of the uterus. *Cancer science* 2006; 97(2): 106-12.
- 73 Okutsu T, Kuroiwa Y, Kagitani F, Kai M, Aisaka K, Tsutsumi O, *et al.* Expression and imprinting status of human PEG8/IGF2AS, a paternally expressed antisense transcript from the IGF2 locus, in Wilms' tumors. *J Biochem* 2000; 127(3): 475-83.
- 74 Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, *et al.* Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 2008; 451(7175): 202-6.
- 75 Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, *et al.* Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med* 2008; 14(7): 723-30.
- 76 Cabianca DS, Casa V, Bodega B, Xynos A, Ginelli E, Tanaka Y, *et al.* A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* 2012; 149(4): 819-31.
- 77 Blower MD, Nachury M, Heald R, Weis K. A Rae1-containing ribonucleoprotein complex is required for mitotic spindle assembly. *Cell* 2005; 121(2): 223-34.