

非编码 RNA 与哺乳动物基因组印记的起源

张意军, 屈良鹄*

中山大学生命科学学院, 基因工程教育部重点实验室, 有害生物控制与资源利用国家重点实验室, 广州 510275

* 联系人, E-mail: lssqlh@mail.sysu.edu.cn

收稿日期: 2008-09-17; 接受日期: 2008-11-24

国家自然科学基金(批准号: 30830066)、国家重点基础研究发展计划(批准号: 2005CB724600)和教育部及广东省创新团队基金(批准号: IRT0447 和 NSF-05200303)资助项目

摘要 基因组印记是由亲本来源不同而导致等位基因表达差异的一种遗传现象, 主要发生在胎盘哺乳动物(真哺乳类)和显花植物中。大部分印记基因都分布在印记基因簇内, 其中包含大量的非编码 RNA 基因。印记基因的表达受印记控制区(ICRs)的顺式调控。基因组印记产生的原因及过程是现代遗传学研究的一个热点问题。分析印记同源区从非印记物种到印记物种的过渡, 为解决这一问题提供了重要启示。最近, 原始哺乳动物(有袋类和单孔类)模式物种全基因组测序的完成, 极大地促进了印记同源区的比较分析研究。本文对这些研究进行了回顾和分析, 发现非编码 RNA 与哺乳动物基因组印记获得关系密切。主要依据为: (1) 伴随着基因组印记的获得, 印记区有大量的非编码 RNA 新基因出现; (2) 与基因组印记相关的一些保守非编码 RNA 的表达发生了显著变化。此外, 对 15 种脊椎动物中印记 snoRNA 基因系统分析的结果表明: 印记 snoRNA 起源于真哺乳类与有袋类动物分化之后, 并且在真哺乳类辐射进化之前发生了迅速的扩张, 主要的基因家族在这一时期已经形成。这些结果进一步证明了非编码 RNA 与基因组印记获得的密切联系。非编码 RNA 可能主要通过调控印记表达和诱导染色体表观遗传修饰两种机制, 参与哺乳动物基因组印记的获得。

关键词

基因组印记
哺乳动物
非编码 RNA

基因组印记(genomic imprinting), 又称为遗传印记(genetic imprinting), 是近年来发现的一种不遵从孟德尔定律的依靠单亲传递某些遗传学性状的现象, 是指在配子或合子发生期间, 来自亲本的等位基因或染色体在发育过程中产生专一性的加工修饰, 导致后代体细胞中两个亲本来源的等位基因有不同的表达活性^[1], 具有这种现象的基因称为印记基因(imprinted gene)。目前主要在高等生物, 如显花植物和胎盘类哺乳动物中发现了基因组印记现象^[2,3]。

1984 年, McGrath 和 Solter^[4] 及 Surani 等人^[5] 在小鼠的核转移实验中发现仅包含母源基因组(孤雌生殖)或者仅包含父源基因组(雄核发育)的胚胎都在怀孕中

途就会死亡, 表明母源基因组和父源基因组对子代的正常发育都是必需的。1991 年, DeChiara 等人^[6] 在小鼠中通过基因剔除技术发现了第一个内源性印记基因, 胰岛素样生长因子 (IGF2)。若被剔除的等位基因源于父本则动物表现为侏儒, 相反如来自母本则无特殊表型, 实验表明在胚胎中 IGF2 基因被印记, 而且仅父源等位基因正常表达。但同时他们发现在脉络丛(choroid plexus)和柔脑膜(leptomeninges)中 IGF2 的两个等位基因是同时表达的。这是一个里程碑性的研究, 因为它不仅表明基因组印记可影响正常的内源性基因, 同时说明基因组印记具有组织特异性调节作用。目前在人类基因组中预测的印记基

因共有 205 个, 其中通过实验验证具有印记表达的有 53 个; 在小鼠中预测的印记基因共有 90 个, 其中 72 个经实验验证具有印记表达(美国杜克大学基因印记网站: <http://www.geneimprint.com>). 值得注意的是人类和小鼠的基因组印记状况存在很大的差异^[2]. 印记基因在基因组中通常是成簇出现, 这些印记基因簇所对应的染色体区域往往被称作印记区(imprinted region). 在人类基因组中, 有 7 个较大的印记基因簇, 分别位于 7q21, 7q32, 11p15, 14q32, 15q11-q13, 19q13 和 20q13 染色体区域^[2]. 在印记区中基因的表达或沉默是由印记控制区(imprinting control regions, ICRs, 又称作印记中心imprinting centers, ICs)通过顺式作用来调控的^[2], ICRs 的本质是一段差异甲基化区域(differentially methylated regions, DMRs), 它的调控作用可以覆盖很长的染色体区域(~Mb).

利用GO(Gene Ontology)分析工具GOSTat对 57 个印记蛋白编码基因的分析发现, 20%的印记基因与形态发生(morphogenesis)有关, 特别是参与到器官生成(organogenesis)的过程. 而这其中与神经形成(neurogenesis)的相关基因又占了最高的比例, 其次是细胞周期调节因子^[2]. 印记基因的缺陷可以导致一些严重的遗传疾病^[2], 如 Prader-Willi 综合征(PWS)、Angelman 综合征(AS)、Beckwith-Wiedemann 综合征(BWS)、假性甲状腺功能减退症(Pseudohypoparathyroidism)、Russell-Silver 综合征等.

哺乳动物基因组印记区中包含有大量非编码 RNA 基因, 据统计非编码RNA占到印记转录产物的 30%^[2], 并且每个印记区都至少包含一个非编码RNA 基因^[7]. 印记非编码RNA种类丰富, 包括反义RNA (antisense RNA)、核仁小分子RNA(snoRNA)、microRNA, 最近的报道^[8]表明基因组印记区可能存在piRNA(piwi-interacting RNA). 大部分的印记非编码RNA都是胎盘哺乳类中特有的, 在鱼类、鸟类、单孔类甚至有袋类中都找不到它们的同源分子. 同时, 印记非编码RNA在基因组织和表达方面也有特异之处: 例如在人类基因组的 *DLK1-DIO3* 和 *PWS* 印记区中各有两个印记snoRNA基因簇, 其中包含的 snoRNA 基因的总数量超过 100 个, 这样大的snoRNA 基因簇在基因组的其他区域还没有发现. 很多印记非编码RNA来自蛋白编码基因的反义转录产

物, 例如 *Air*, *Kcnq1ot1* 等, 这些反义转录产物很长, 其中往往包含有大量小分子非编码RNA. 有证据表明一些大分子非编码RNA参与了印记基因簇的调控, 它们的表达会导致附近蛋白编码基因的沉默. 目前认为, 印记小分子非编码RNA的主要功能是识别细胞内的RNA靶标并执行反式调节作用(*trans-acting*, 例如microRNA通过RNA干涉机制介导基因的转录后沉默), 而对其在基因组印记调控中扮演的角色还知之甚少^[8].

1 基因组印记的获得与非编码 RNA

众所周知, 二倍体生物的基因组中包含有两套完整的遗传信息, 一套来自父方, 一套来自母方, 并且这两套遗传信息中包含的等位基因在子代中可以对等地表达. 一般认为这种基因冗余是在自然选择压力下形成的一种保险机制, 即当其中一个基因由于突变等因素失去功能时, 另一个拷贝可以弥补这种缺失而不会损害生物的适应性. 基因组印记的出现打破了这种两个等位基因同时表达的“双保险”模式, 似乎为生物体的遗传和发育带来了新的风险, 这也可以解释为什么很多疾病与基因组印记相关. 但另一方面, 基因组印记现象只出现在胎盘哺乳类和显花植物这些进化的高等类群中, 表明这是一种被自然选择所偏爱的、先进的遗传方式. 因此, 为什么出现这种遗传方式及其起源成为学者广泛关注的问题. 研究人员试图利用印记物种(如真哺乳类)与非印记物种(如单孔类、有袋类)印记同源区的比较分析来寻找线索. 最近哺乳类中最古老的类群单孔类和有袋类中模式物种全基因组测序工作的完成, 使得这些研究取得了一些令人瞩目的成果. 通过回顾这些研究, 可以发现非编码 RNA 与哺乳动物基因组印记的出现有着密切的关联. 主要体现在伴随基因组印记的产生, 哺乳动物的非编码 RNA 有两方面显著的变化: 一是有大量的非编码 RNA 新基因在印记区出现, 二是非编码 RNA 的转录水平发生改变或者出现新的非编码 RNA 转录产物.

1.1 新非编码 RNA 基因的产生

(1) *BWS* 印记区的 *H19*. *BWS* 印记区是目前研究最为充分的印记区之一, 分别位于人类(*Homo sapiens*)基因组的 11p15.5 染色体区和小鼠(*Mus*

musculus)基因组 7 号染色体的末端, 第一个被发现的内源性印记基因 *IGF2* 就位于这个印记区中。在人类基因组中这个印记区全长约 800 kb, 可分为 *IGF2* 和 *CDKN1C* 两个印记亚区, 它们之间由 4 个非印记表达的基因(*TRPM5*, *TSSC4*, *CD81*, *PHEMX*)分隔(图 1(A))。Yokomine 等人^[9]对哺乳动物和鸟类的 *IGF2* 印记同源区进行了比较分析, 发现这个区域中的大部分蛋白编码基因在鸟类和哺乳类动物之间是保守的, 包括基因的转录方向以及内含子-外显子的结构特征, 这些基因在鸟类中没有印记表达。然而值得注意的是鸟类的基因组中缺失了一个非编码 RNA 基因——*H19* 以及位于该基因附近的印记中心(*H19* 印记中心)。最近, Smits 等人^[10]发现 *H19* 基因在有袋类中是存在的, 并且是印记表达的。这进一步证实了 *H19* 基因与印记表达的相关性。此外, *H19* 中包含的一个 microRNA 基因 *Mir-675* 在整个胎盘哺乳类中十分保守, 表明它可能行使重要的调控功能^[10]。

(2) *DLK1-DIO3* 印记区的 microRNA 和 snoRNA 基因簇。*DLK1-DIO3* 印记区位于人类 14 号染色体(14q32), 其中包括 3 个父源表达的蛋白编码基因: *Delta-like homologue 1(DLK1)*, 类反转座子基因 1 (*RTL1/MART1*), 3 型脱碘酶基因(*DIO3*); 以及众多母源表达的非编码 RNA 基因: *GTL2*, *RTL1* 反义转录产物(*RTL1AS*), C/D snoRNA 基因簇和 microRNA 基因簇。其中有 7 个 microRNA 位于 *RTL1AS* 基因中(图 1(B))。这些非编码 RNA 基因都是母源表达并且具有相同的转录方向, 被认为来自于同一个转录单元^[11,12]。*GTL2* 的 5' 端序列对这个非编码转录单元的转录活性具有重要影响^[13]。Edwards 等人^[13]通过对真兽亚纲、后兽亚纲、原兽亚纲以及鸟类和鱼类中 7 个物种的比较分析发现, *DLK1* 和 *DIO3* 这两个基因在 7 个物种中都存在。但在真兽亚纲以外没有发现 C/D snoRNA 基因簇和 microRNA 基因簇的同源序列, 表明这些基因是真兽类与有袋类分化之后才出现的。此外, 位于真兽亚纲 *RTL1AS* 中的 microRNA 基因在有袋类中也没有找到。研究表明这些 microRNA 可以通过类似 RNAi(RNA interfering) 的机制来切割 *Rtl1* mRNA^[14,15], 可能是针对 *RTL1* 基因反转座子的性质而发展出来的一种宿主防御机制。*RTL1* 基因获得新的功能之后, 这些 microRNA 又起到了稳定其表达的作用。距离

RTL1 基因较远的 microRNA 基因簇中的成员可能参与了神经与胎盘发育的过程^[16], 但不排除它们参与印记基因簇调控的可能。

(3) *PWS-AS* 印记区的 snoRNA 基因簇。*PWS-AS* 印记区位于人类染色体的 15q11~13 区域, 因这个区域与 Prader-Willi 综合征和 Angelman 综合征相关而得名。这个印记区包含 *PWS* 和 *AS* 两个亚区, 在 *AS* 亚区的远端分布有两个基因 *UBE3A* 和 *ATP10A*, 二者在神经组织中都只有母源表达(父源等位基因被沉默), 其中 *UBE3A* 被认为与 Angelman 综合征关系密切。*PWS* 亚区较大, 包含有 5 个父源表达的蛋白编码基因: *MKRN3*, *MAGE2*, *NDN*, *SNURF* 和 *SNRPN*。该亚区还包括两个父源表达的 snoRNA 基因簇, *HBII-85* 和 *HBII-52*, 它们从一个起始于 *SNURF-SNRPN* 上游的转录产物中加工而来(图 1(C))。最新的研究结果^[17] 表明, snoRNA 基因簇 *HBII-85* 与 Prader-Willi 综合征的关系最为密切。为了揭示该印记区的进化过程, 研究者比较了有袋类、单孔类和高等胎盘类动物的基因组相关区域和基因^[18]。在高等胎盘类之外没有发现 *MKRN3*, *MAGE2* 和 *NDN* 基因, 同时 *SNRPN-UBE3A* 这一基因串联结构也不存在, 取而代之的是 *UBE3A-CNGA3* 串联体, 在有袋类以下的物种中甚至不存在 *SNRPN* 基因, 这些表明真哺乳类中现存的基因串联结构是最近才构建起来的。有趣的是, 在高等胎盘哺乳类之外的同源区内也没有发现 snoRNA 基因簇。*HBII-52* 和 *HBII-85* 这两个 snoRNA 基因簇近来吸引了很多注意。*HBII-52* 被认为参与了 5-羟色胺 2C 受体(5-hydroxytryptamine 2C receptor, 5-HT_{2C}R)mRNA 前体的转录后调控, 包括 A-I 的 RNA 编辑^[19,20] 以及选择性剪接^[21]。越来越多的证据表明, *HBII-85* 基因簇与 Prader-Willi 综合征密切相关, 其作用的靶目标及分子机制已成为本领域最具挑战性的课题之一^[22]。

(4) X 染色体失活中心(XIC)的众多非编码 RNA。哺乳动物中的 X 染色体失活与基因组印记现象有着紧密的联系。在有袋类动物中仅父源的 X 染色体被失活^[23], 而在真哺乳类中 X 染色体的失活是随机的。但在啮齿类和牛(*Bos taurus*)的胚胎外组织中, 父源的 X 染色体也是被优先失活的^[24-26]。在人类和小鼠中, X 染色体的失活是通过非编码 RNA-X 染色体失活特

异转录物(X-inactive specific transcript, *XIST*)实现的^[27], 在*XIST*周围还分布着很多非编码RNA基因, 形成了X染色体失活中心(X-inactivation centre, XIC)(图 1 (D))^[28,29]. *XIST*有一个反义转录基因*TSIX*. 在小鼠中*Tsix* RNA与临近的另外一个非编码RNA *Xite*共同作用来确定X染色体的数目^[30]以及选择哪一条X染色体将被失活^[31,32]. 此外, 在*XIST*的上游还有两个大分子非编码RNA基因*JPX*和*FTX*. 这一区域由于组蛋白3的第9位赖氨酸(H3K9)被高度甲基化而成为“成核作用”中心, 其作用是使*XIST*介导的沉默作用传递到整

个X染色体^[33]. 有趣的是在*FTX*基因中还包含有一个microRNA基因簇^[34], 这些microRNA可能在X染色体失活的起始阶段介导染色体的重构. 在小鼠和牛的胚胎外组织中, *XIST*是印记表达的(父源基因表达)^[35,36], 并且调控这些组织中的X染色体失活. 从这个角度来讲, 可以将XIC看做是一个印记基因簇. 比较基因组分析表明真哺乳类以外的有袋类和单孔类中没有*XIST*基因, 也没有发现与*TSIX*, *XITE*, *FTX*, *JPX*以及*FTX*中包含的microRNAs相似的序列^[37]. 同源分析表明, *XIST*可能起源于一个蛋

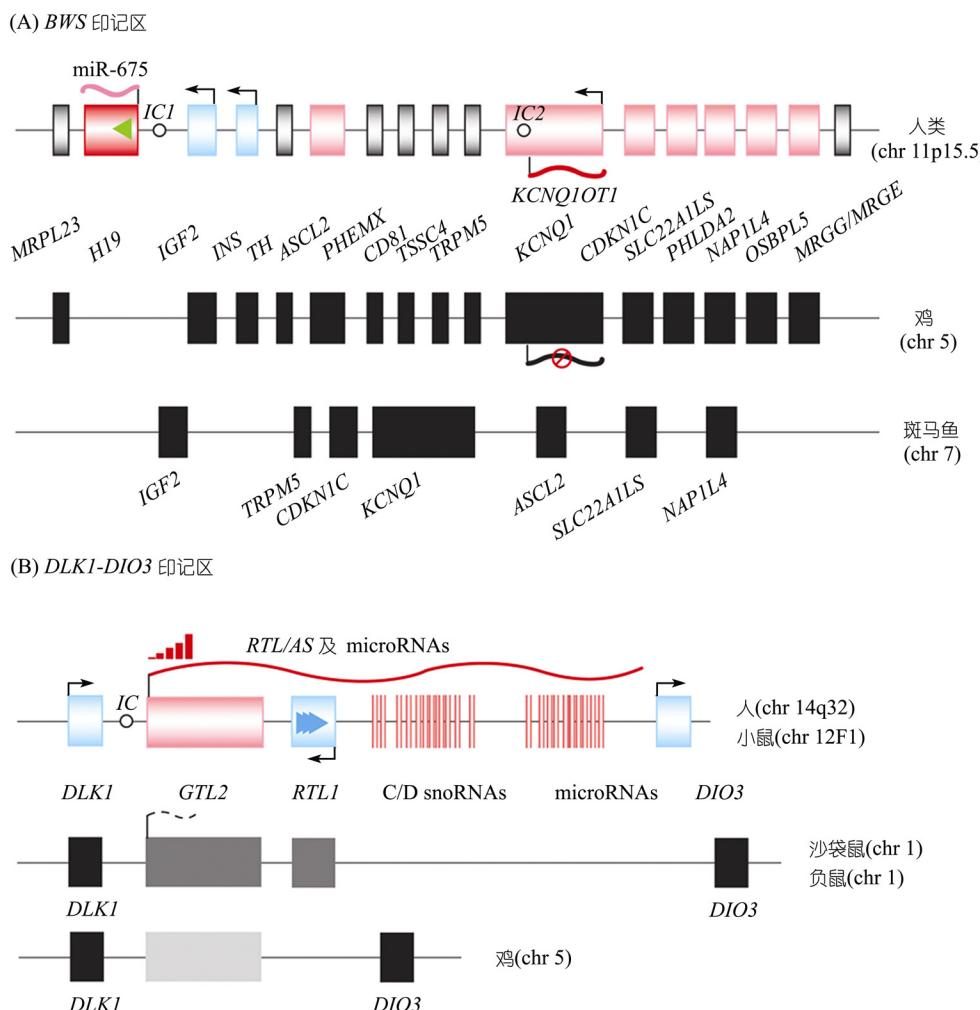


图 1 基因组印记同源区非编码 RNA 的进化

(A) BWS 印记区; (B) *DLKI-DIO3* 印记区; (C) *PWS-AS* 印记区; (D) X 染色体失活中心(XIC). 父源表达的印记基因由淡蓝色表示, 母源表达的印记基因由粉红色表示, 非印记表达的基因由灰色表示; 红色表示新产生的或者表达发生变化的非编码 RNA, 黑色代表非印记物种中的同源基因, 颜色深度减弱表明同源性降低. 箭头代表基因的转录方向, 波浪线代表非编码转录产物, 三角形代表分布在基因内部的 microRNA 基因, 红色叉形表示基因由于假基因化(pseudogenization)而衰亡, chr 代表染色体

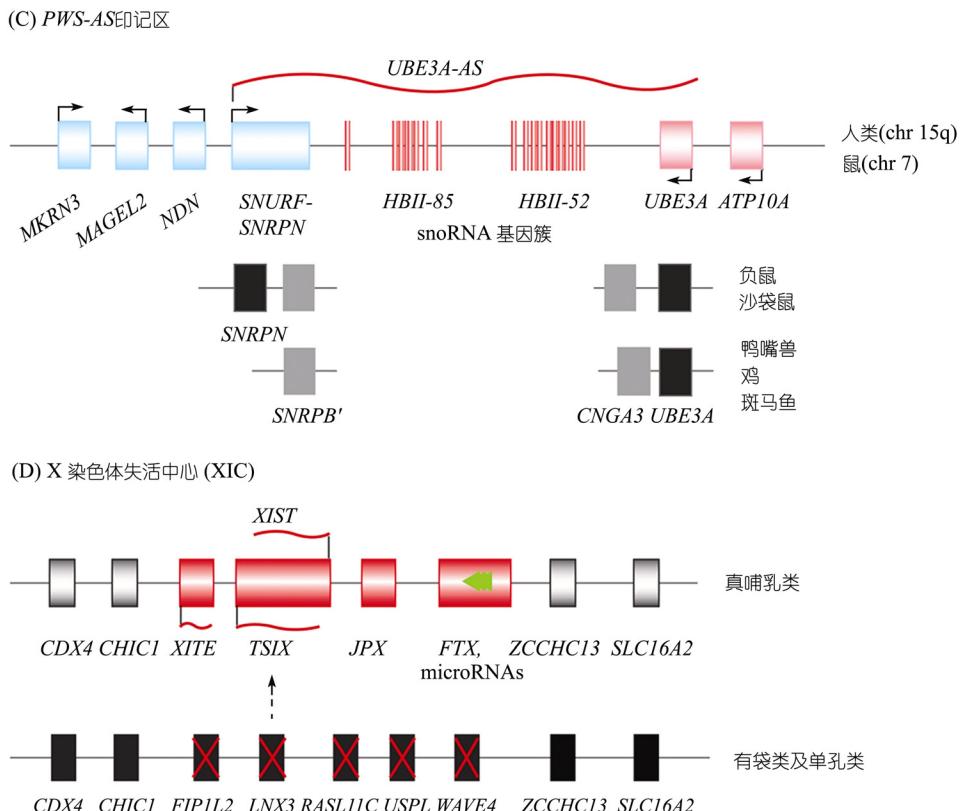


图 1(续)

白编码基因 $LNX3$ ，这个基因在除真哺乳类以外的羊膜动物中都存在。 $XIST$ 很可能是在真哺乳类进化早期 $LNX3$ 发生假基因化(pseudogenization)而产生的^[37]。

1.2 非编码 RNA 表达水平的变化

(1) $GTL2$ 基因表达上调。由于在其他类群中没有发现同源分子， $GTL2$ 曾被认为是通过转座事件插入到 $DLK1-DIO3$ 印迹区而产生的真哺乳类特异的非编码RNA基因^[38]。但最近的研究结果^[13]显示，这个基因在所有羊膜动物中都有表达，但是在真哺乳类中其表达水平较其他类群中高出很多(图 1(B))。有趣的是，在小鼠中这一印迹区内所有非编码RNA基因都来自同一转录单元，这些非编码RNA包括 $RTL1AS$ 以及为数众多的C/D snoRNA和microRNA，而这一转录单元的活性依赖于 $GTL2$ 基因的 5 端序列。因此 $GTL2$ 表达水平的升高对这些非编码RNA的表达具有重要影响。此外，这些印迹非编码RNA的起源也可能与 $GTL2$ 表达水平的变化有关， $GTL2$ 表达的增强为这些非编码RNA提供了一个理想的宿主^[13]。非编

码RNA 基因 $GTL2$ 在真哺乳类中表达显著上调，直接导致了这一区域的其他非编码RNA转录水平的提高，它们在 $DLK1-DIO3$ 印迹区获得印迹表达过程中可能扮演重要的角色。

(2) $KCNQ1OT1$ 表达启动。Paulsen等人^[39]利用人、鸡(*Gallus gallus*)、河豚(*Takifugu rubripes*)和斑马鱼(*Danio rerio*)的基因组数据对完整的 BWS 印迹区进行了比较分析。由于印迹中心在印迹表达调控中的重要作用，他们对 BWS 印迹区中两个印迹中心(IC1 和 IC2)进行了重点分析。IC1 即上文中提到的 $H19$ 印迹中心，IC2 位于一个母源表达基因 $KCNQ1$ 的内含子中，并且充当其反义RNA基因 $KCNQ1OT1$ ($LIT1$)启动子的角色(图 1(A))。他们的研究结果与Yokomine等人^[9]的吻合，在哺乳动物和鸟类之间， BWS 基因簇的基因共线性非常保守，在鱼类中这种共线性的保守程度有所下降。虽然 $KCNQ1$ 基因在脊椎动物中十分保守，目前还没有证据表明其反义转录物 $Kcnq1ot1$ RNA在鸟类中表达，说明这个非编码RNA的转录很可能 是从 哺乳类 中 才 开始 的。在 小 鼠

中, 这个非编码RNA对于维持其附近基因的印记表达具有重要作用, 并且这种调控作用与 *Kcnq1ot1* RNA的长度是正相关的^[40]。

以上研究发现, 印记非编码 RNA 的出现与哺乳动物基因组印记的获得有着密切的联系。 *BWS* 和 *DLK1-DIO3* 印记区中蛋白编码基因的共线性以及基因结构在真哺乳类和有袋类之间很保守, 一些基因甚至在整个脊椎类中保守。这种保守性表明在这些印记区中, 蛋白编码基因的变化并不是引发基因组印记产生的主要因素。相比而言, 在印记物种与非印记物种之间变化显著的是非编码 RNA 的出现、反转座基因的插入以及一些调控序列的产生。在这些可能导致基因组印记的因素中, 目前还无法判断哪个因素最为关键, 但非编码 RNA 的重要作用是十分明确的。同时, 以上的研究中尚存在着一些不足, 主要是研究中使用的物种较少, 往往根据一两个物种的分析结果来推测一大类群的情况, 因而无法反映出基因组印记的多样性与复杂性。例如, 长期以来对于基因组印记调控的研究主要集中在小鼠和人类上, 而最近的研究表明, 负鼠(*Monodelphis domestica*)和狗(*Canis familiaris*)中 *IGF2R* 基因的印记调控机制与前两者是不同的^[41,42]。这说明要全面地了解哺乳动物基因组印记现象, 需要对更多具有代表性的物种进行分析^[37], 越来越多脊椎动物全基因组数据的出现使得全面系统地分析印记非编码RNA的起源与进化成为可能。

2 哺乳动物印记 snoRNA 的进化

snoRNA 是一类富集于细胞核仁的小分子 RNA, 也是基因组印记区中最大的非编码 RNA 家族, 其结构与功能在生物进化中表现出高度多样性。近年来, 本研究小组采用计算机分析与实验相结合的方法, 在国际上率先完成了果蝇(*Drosophila melanogaster*)、水稻(*Oryza sativa*)、莱茵衣藻(*Chlamydomonas reinhardtii*)等真核生物 snoRNA 的系统分析和基因组注释, 揭示了大批新的 snoRNA 的结构与功能及其特有的基因组织和转录后加工机制。最近, 针对人和哺乳动物等模式生物基因组庞大的序列数据和如何发掘出新的 snoRNA 基因等难题, 本研究组自主开发出比较基因组学与智能模式识别扫描分析相结合的先进计算机 RNA 组学软件包——snoSeeker, 首次实现

了对人类向导和孤儿snoRNA的全基因组分析^[43]。为了研究印记非编码 RNA 基因的起源及进化过程, 对 15 个脊椎动物基因组中一类重要的印记非编码 RNA—snoRNA进行了系统鉴定和分析。除了鸟类、单孔类、有袋类的代表性物种基因组数据外, 还分析了真哺乳类中各主要类群的代表性物种, 如: 象(*Loxodonta africana*)、猬(*Echinops telfairi*)、犰狳(*Dasyurus novemcinctus*)、牛、狗、大鼠(*Rattus norvegicus*)、恒河猴(*Macaca mulatta*)和人等。目前发现的印记snoRNA基因都集中在 *DLK1-DIO3* 和 *PWS* 印记区, 以 snoSeeker为基础结合比较基因组和单基因组分析策略, 对这两个印记同源区中的snoRNA基因进行了系统鉴定并总结了哺乳动物中的分布情况。在真哺乳类以外的有袋类、单孔类和鸟类中都没有发现印记snoRNA基因, 表明这类印记非编码RNA起源于真哺乳类与有袋类的祖先分化之后(~180 Ma)。同时发现, 印记snoRNA的主要基因家族在哺乳动物早期分化谱系中已经存在, 表明印记snoRNA在真哺乳类祖先中的出现十分迅速, 可能经历了一次爆发式的扩张, 而非渐进式的进化过程。此外, 印记snoRNA成员的数量在哺乳类的主要类群间变化显著, 说明印记snoRNA基因在哺乳动物分化期间(或分化后)也发生过快速的扩张事件。这种快速的扩张使印记snoRNA基因(如 *HBII-52*)产生更高的基因多态性, 并为新功能的获得提供了基础(未发表资料)。印记snoRNA的爆发时间与哺乳类获得印记表达的时间相吻合, 暗示着两者之间可能存在密切联系。Glazov 等人^[16]对 *DLK1-DIO3* 印记区中 microRNA 基因簇的起源与进化进行了分析, 结果表明这个基因簇也是起源于真哺乳类。并且从其研究数据中可以发现, 这个 microRNA 基因簇的起源也可能经历了爆发式扩张, 这进一步印证了本研究组的推测。

3 非编码 RNA 在基因组印记获得中的作用

非编码 RNA 在印记获取过程中可能有两个方面的作用: 一是大分子非编码 RNA 通过顺式作用调控基因的印记表达; 二是小分子非编码 RNA 基因诱发了差异甲基化区域(DMR)的形成。非编码 RNA *Xist* 介导哺乳动物 X 染色体失活, 是非编码 RNA 参与基因组调控最典型的范例。而大分子非编码 RNA 的转录是印记基因簇调控印记表达的一种重要机

制^[44], 目前的研究显示, 大部分的印记基因簇的调控都是通过这一机制完成的^[45]. 其中代表性的基因包括AIR-IGF2R印记区的AIR基因和BWS印记区的KCNQ1OT1基因. 以KCNQ1OT1为例, KCNQ1OT1基因的启动子位于KCNQ1基因的第10内含子内^[46], 并且与一个母源高度甲基化的CpG岛(ICRs)重叠^[46,47]. 正常情况下这个基因为父源表达, 母源沉默. 而当父源染色体上的这个区域被缺失后, 这个印记基因簇的印记表达状态便丧失, 所有基因恢复双等位基因表达(bi-allelic expression)^[48,49]. 而通过插入转录终止信号将这个非编码RNA截短后, 该印记基因簇同样失去印记表达, 表明KCNQ1OT1转录产物本身直接参与了印记调控^[49]. 非编码RNA Air的情况与之十分相似^[8]. 因此, 很容易可以将大分子非编码RNA的出现和基因组印记表达的获得联系起来.

DNA甲基化在基因组印记的建立中扮演着重要的角色^[50], 而高度重复序列是DNA甲基化修饰的主要目标^[51~53]. 印记区中存在的大量以基因簇形式出现的小分子非编码RNA很可能充当了诱导甲基化的

信号. 以人类基因组为例, 在DLK1-DIO3印记区中包含一个有40拷贝的microRNA基因簇, 15q11~13印记区的两个snoRNA基因簇中的基因数量接近70个. 在这些基因簇中, 小分子非编码RNA基因间还分布着其他的重复序列和元件^[16,54], 使得整个重复序列的长度可以达到上百kb. 此外, snoRNA基因簇的G+C含量与附近序列表现出明显差异^[54]. 在植物中, 大量研究表明小分子非编码RNA介导的转录水平基因沉默与特异的甲基化修饰相关^[55~57]. 在裂殖酵母(*Schizosaccharomyces pombe*)中, 来自中心粒重复序列的siRNA经Dicer加工后进入RITS (RNA-induced transcriptional silencing complex)中, 可以在转录水平指导异染色质的形成和基因沉默^[58]. 在哺乳动物体系中, siRNA指导的DNA甲基化及组蛋白H3-K9甲基化也已经得到证实^[59,60]. 因此, 印记区中包含的大量小分子非编码RNA, 尤其是microRNA, 很有可能参与了与基因组印记相关的DNA甲基化修饰. 在图2中, 对非编码RNA在哺乳动物基因组印记获得过程中的作用和可能的机制进行了总结.

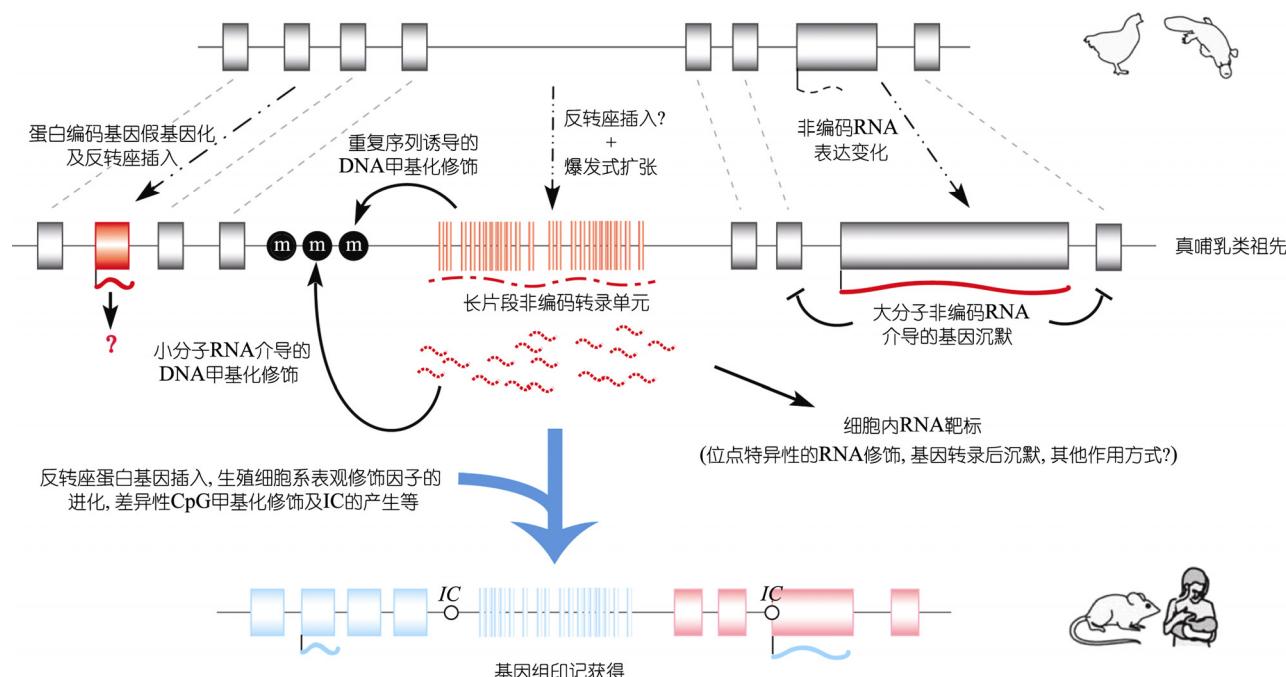


图2 非编码RNA在哺乳动物基因组印记获得过程中的作用和可能的机制

父源表达的印记基因由淡蓝色表示, 母源表达的印记基因由粉红色表示, 非印记表达的基因由灰色表示, 红色表示新产生的以及表达发生变化的非编码RNA; 波浪线代表非编码转录产物, 同源基因由虚线相连. 黑色实线箭头表示非编码RNA参与的调控过程, 黑色虚线箭头表示印记区非编码RNA的起源与进化过程

4 总结与展望

哺乳动物基因组印记起源及进化的过程是目前遗传和表观遗传学研究领域的热点问题。本文回顾了近期关于脊椎动物印记同源区比较分析的研究,发现非编码 RNA 与哺乳动物基因组印记的获得密

相关。对印记 snoRNA 基因在胎盘哺乳类中的进化过程和模式的研究进一步印证了这一观点。然而,迄今为止,人们对于非编码 RNA 在哺乳动物基因组印记获得过程中所扮演的角色及其作用机制的了解还十分有限,揭示印记非编码 RNA 的起源过程及其生物学功能将是解决这一问题的关键。

致谢 感谢本实验室董志威、张莹、杨博、马丽明及周鹤达提供相关资料,邵鹏、杨建华对本文提出建设性意见。

参考文献

- 1 詹晓渊,熊俊,胡以平.基因组印迹基因及其生物学意义.中国科学基金,2001,15(2): 68—72
- 2 Morison I M, Ramsay J P, Spencer H G. A census of mammalian imprinting. *Trends Genet*, 2005, 21(8): 457—465 [[DOI](#)]
- 3 Field L M, Lyko F, Mandrioli M, et al. DNA methylation in insects. *Insect Mol Biol*, 2004, 13(2): 109—115 [[DOI](#)]
- 4 McGrath J, Solter D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*, 1984, 37(1): 179—183 [[DOI](#)]
- 5 Surani M A, Barton S C, Norris M L. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature*, 1984, 308(5959): 548—550 [[DOI](#)]
- 6 DeChiara T M, Robertson E J, Efstratiadis A. Parental imprinting of the mouse insulin-like growth factor gene. *Cell*, 1991, 64(4): 849—859 [[DOI](#)]
- 7 Edwards C A, Ferguson-Smith A C. Mechanisms regulating imprinted genes in clusters. *Curr Opin Cell Biol*, 2007, 19(3): 281—289 [[DOI](#)]
- 8 Royo H, Cavaille J. Non-coding RNAs in imprinted gene clusters. *Biol Cell*, 2008, 100(3): 149—166 [[DOI](#)]
- 9 Yokomine T, Shirohzu H, Purbowasito W, et al. Structural and functional analysis of a 0.5-Mb chicken region orthologous to the imprinted mammalian Ascl2/Mash2-Igf2-H19 region. *Genome Res*, 2005, 15(1): 154—165 [[DOI](#)]
- 10 Smits G, Mungall A J, Griffiths-Jones S, et al. Conservation of the H19 noncoding RNA and H19-IGF2 imprinting mechanism in therians. *Nat Genet*, 2008, 40(8): 971—976 [[DOI](#)]
- 11 Lin S P, Youngson N, Takada S, et al. Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nat Genet*, 2003, 35(1): 97—102
- 12 Tierling S, Dalbert S, Schoppenhorst S, et al. High-resolution map and imprinting analysis of the Gtl2-Dnchc1 domain on mouse chromosome 12. *Genomics*, 2006, 87(2): 225—235 [[DOI](#)]
- 13 Edwards C A, Mungall A J, Matthews L, et al. The evolution of the DLK1-DIO3 imprinted domain in mammals. *PLoS Biol*, 2008, 6(6): e135 [[DOI](#)]
- 14 Seitz H, Youngson N, Lin S P, et al. Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat Genet*, 2003, 34(3): 261—262 [[DOI](#)]
- 15 Davis E, Caiment F, Tordoir X, et al. RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. *Curr Biol*, 2005, 15(8): 743—749 [[DOI](#)]
- 16 Glazov E A, McWilliam S, Barris W C, et al. Origin, evolution, and biological role of miRNA cluster in DLK-DIO3 genomic region in placental mammals. *Mol Biol Evol*, 2008, 25(5): 939—948 [[DOI](#)]
- 17 Sahoo T, del Gaudio D, German J R, et al. Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat Genet*, 2008, 40(6): 719—721 [[DOI](#)]
- 18 Rapkins R W, Hore T, Smithwick M, et al. Recent assembly of an imprinted domain from non-imprinted components. *PLoS Genet*, 2006, 2(10): e182 [[DOI](#)]
- 19 Cavaille J, Buiting K, Kiefmann M, et al. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an un-

- usual genomic organization. Proc Natl Acad Sci USA, 2000, 97(26): 14311—14316[DOI](#)
- 20 Vitali P, Basyuk E, Le Meur E, et al. ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. J Cell Biol, 2005, 169(5): 745—753[DOI](#)
- 21 Kishore S, Stamm S. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. Science, 2006, 311(5758): 230—232[DOI](#)
- 22 Peters J. Prader-Willi and snoRNAs. Nat Genet, 2008, 40(6): 688—689[DOI](#)
- 23 Richardson B J, Czuppon A B, Sharman G B. Inheritance of glucose-6-phosphate dehydrogenase variation in kangaroos. Nat New Biol, 1971, 230(13): 154—155
- 24 Takagi N, Sasaki M. Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. Nature, 1975, 256(5519): 640—642[DOI](#)
- 25 West J D, Frels W I, Chapman V M, et al. Preferential expression of the maternally derived X chromosome in the mouse yolk sac. Cell, 1977, 12(4): 873—882[DOI](#)
- 26 Xue F, Tian X C, Du F, et al. Aberrant patterns of X chromosome inactivation in bovine clones. Nat Genet, 2002, 31(2): 216—220
- 27 Heard E, Distech C M. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. Genes Dev, 2006, 20(14): 1848—1867[DOI](#)
- 28 Brown C J, Lafreniere R G, Powers V E, et al. Localization of the X inactivation centre on the human X chromosome in Xq13. Nature, 1991, 349(6304): 82—84[DOI](#)
- 29 Chureau C, Prissette M, Bourdet A, et al. Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. Genome Res, 2002, 12(6): 894—908
- 30 Lee J T. Regulation of X-chromosome counting by Tsix and Xite sequences. Science, 2005, 309(5735): 768—771[DOI](#)
- 31 Lee J T, Lu N. Targeted mutagenesis of Tsix leads to nonrandom X inactivation. Cell, 1999, 99(1): 47—57[DOI](#)
- 32 Ogawa Y, Lee J T. Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. Mol Cell, 2003, 11(3): 731—743[DOI](#)
- 33 Heard E, Rougeulle C, Arnaud D, et al. Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. Cell, 2001, 107(6): 727—738[DOI](#)
- 34 Hore T A, Koina E, Wakefield M J, et al. The region homologous to the X-chromosome inactivation centre has been disrupted in marsupial and monotreme mammals. Chromosome Res, 2007, 15(2): 147—161[DOI](#)
- 35 Kay G F, Penny G D, Patel D, et al. Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. Cell, 1993, 72(2): 171—182[DOI](#)
- 36 Dindot S V, Farin P W, Farin C E, et al. Epigenetic and genomic imprinting analysis in nuclear transfer derived Bos gaurus/Bos taurus hybrid fetuses. Biol Reprod, 2004, 71(2): 470—478[DOI](#)
- 37 Hore T A, Rapkins R W, Graves J A. Construction and evolution of imprinted loci in mammals. Trends Genet, 2007, 23(9): 440—448[DOI](#)
- 38 Weidman J R, Maloney K A, Jirtle R L. Comparative phylogenetic analysis reveals multiple non-imprinted isoforms of opossum Dlk1. Mamm Genome, 2006, 17(2): 157—167[DOI](#)
- 39 Paulsen M, Khare T, Burgard C, et al. Evolution of the Beckwith-Wiedemann syndrome region in vertebrates. Genome Res, 2005, 15(1): 146—153[DOI](#)
- 40 Kanduri C, Thakur N, Pandey R R. The length of the transcript encoded from the Kcnq1ot1 antisense promoter determines the degree of silencing. Embo J, 2006, 25(10): 2096—2106[DOI](#)
- 41 Weidman J R, Dolinoy D C, Maloney K A, et al. Imprinting of opossum Igf2r in the absence of differential methylation and air. Epigenetics, 2006, 1(1): 49—54
- 42 O'Sullivan F M, Murphy S K, Simel L R, et al. Imprinted expression of the canine IGF2R, in the absence of an anti-sense transcript or promoter methylation. Evol Dev, 2007, 9(6): 579—589
- 43 Yang J H, Zhang X C, Huang Z P, et al. snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome. Nucleic Acids Res, 2006, 34(18): 5112—5123[DOI](#)
- 44 Pauler F M, Koerner M V, Barlow D P. Silencing by imprinted noncoding RNAs: is transcription the answer? Trends Genet, 2007, 23(6): 284—292[DOI](#)
- 45 Thorvaldsen J L, Bartolomei M S. SnapShot: imprinted gene clusters. Cell, 2007, 130(5): 958

- 46 Smilinich N J, Day C D, Fitzpatrick G V, et al. A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proc Natl Acad Sci USA*, 1999, 96(14): 8064—8069 [[DOI](#)]
- 47 Mancini-DiNardo D, Steele S J, Ingram R S, et al. A differentially methylated region within the gene Kcnq1 functions as an imprinted promoter and silencer. *Hum Mol Genet*, 2003, 12(3): 283—294 [[DOI](#)]
- 48 Fitzpatrick G V, Soloway P D, Higgins M J. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat Genet*, 2002, 32(3): 426—431 [[DOI](#)]
- 49 Mancini-Dinardo D, Steele S J, Levorse J M, et al. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes Dev*, 2006, 20(10): 1268—1282 [[DOI](#)]
- 50 Wilkins J F. Genomic imprinting and methylation: epigenetic canalization and conflict. *Trends Genet*, 2005, 21(6): 356—365 [[DOI](#)]
- 51 Bender J. DNA methylation and epigenetics. *Annu Rev Plant Biol*, 2004, 55: 41—68
- 52 Royo H, Bortolin M L, Seitz H, et al. Small non-coding RNAs and genomic imprinting. *Cytogenet Genome Res*, 2006, 113(1-4): 99—108 [[DOI](#)]
- 53 Walter J, Hutter B, Khare T, et al. Repetitive elements in imprinted genes. *Cytogenet Genome Res*, 2006, 113(1-4): 109—115 [[DOI](#)]
- 54 Cavaille J, Seitz H, Paulsen M, et al. Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum Mol Genet*, 2002, 11(13): 1527—1538 [[DOI](#)]
- 55 Mette M F, Aufsatz W, van der Winden J, et al. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J*, 2000, 19(19): 5194—5201 [[DOI](#)]
- 56 Jones L, Ratcliff F, Baulcombe D C. RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. *Curr Biol*, 2001, 11(10): 747—757 [[DOI](#)]
- 57 Sijen T, Vijn I, Rebocho A, et al. Transcriptional and posttranscriptional gene silencing are mechanistically related. *Curr Biol*, 2001, 11(6): 436—440 [[DOI](#)]
- 58 Matzke M A and Birchler J A. RNAi-mediated pathways in the nucleus. *Nat Rev Genet*, 2005, 6(1): 24—35 [[DOI](#)]
- 59 Kawasaki H, Taira K. Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature*, 2004, 431(7005): 211—217 [[DOI](#)]
- 60 Morris K V, Chan S W, Jacobsen S E, et al. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science*, 2004, 305(5688): 1289—1292 [[DOI](#)]