

基因组印记中的长非编码 RNA*

杨文志¹⁾ 张 萃¹⁾ 王冠楠¹⁾ 李冬杰²⁾ 李世杰^{1)**}

(¹⁾河北农业大学生命科学学院, 保定 071000; ²⁾河北科技大学生物科学与工程学院, 石家庄 050018)

摘要 长非编码 RNA(lncRNA)是长度大于 200 bp 的一类非编码蛋白的 RNA, 因其在基因组中含量巨大以及重要的生物学功能引起了学术界的广泛关注. 基因组印记是一种表观遗传现象, lncRNAs 通过建立靶基因的印记而发挥重要的生物功能. 基因组印记可以用来研究 lncRNAs 在转录和转录后水平调控基因表达的分子机制. 本文选取 6 个印记机制研究比较透彻的印记区域, 包括 *Kcnq1/Cdkn1c*、*Igf2r/Airn*、*Prader-Willi (PWS)/Angelman (AS)*、*Snurf/Snrpn*、*Dlk1-Dio3* 和 *H19/Igf2*. 通过介绍包括基因间 lncRNAs(*H19*、*Ipw* 和 *Meg3*)、反义 lncRNAs(*Kcnq1ot1*、*Airn*、*Ube3a-ATS*)和增强子 lncRNAs(IG-DMR eRNAs)在内的 3 种类型 lncRNAs 在印记调控中的作用, 从而了解 lncRNAs 通过顺式或(和)反式作用多种机制调控亲本特异性靶基因的表达. 了解印记基因簇中 lncRNAs 的作用方式将有助于我们揭示 lncRNAs 在整个基因组中的作用机制.

关键词 长非编码 RNA, 基因组印记, 染色质, 表观遗传修饰

学科分类号 Q78

DOI: 10.16476/j.pibb.2015.0391

基因组印记(genomic imprinting)是一种表观遗传(epigenetic)现象, 是指在哺乳动物早期配子发生过程中接近 1% 的蛋白编码基因经过表观遗传标记, 从而使基因的表达具有亲本特异性. 基因组印记是胎盘类哺乳动物进行剂量补偿的生物学及生态学机制, 在妊娠过程中, 通过基因组印记对存在于亲本双方的利益对抗进行计量平衡(特别是生长相关基因)^[1-2]. 目前在鼠中被确定的印记基因有 150 多个.

印记基因通常成簇存在, 印记基因簇大小从几 kb 到 2~3 Mb. 目前所有已知的印记基因簇中都包括至少 1~2 个长非编码 RNA(long noncoding RNA, lncRNA), 其表达方式与其对应的蛋白编码基因的表达方式相反^[3-4]. 印记区域中 lncRNA 的活性转录通常会导致该区域中的其他基因(大多数为编码蛋白基因)的沉默^[5-7]. 长非编码 RNAs 的启动子位于印记区域的差异甲基化区(differentially methylated region, DMR), 在鼠中这部分缺失会导致蛋白编码基因印记的丢失. 由于 DMRs 控制临近基因的印记, 被称为印记调控区(imprinting control regions, ICRs), 其大小通常为 1~3 kb.

近几年的研究表明 lncRNAs 是基因亲本特异

性表达的主要调控力量. 长非编码 RNA 可以作为染色质隔离子, 通过调控增强子和基因启动子之间的相互作用长距离调控基因表达^[8], 这种机制以 *H19-Igf2* 印记基因簇的 *H19* 为例^[9-10]. 长非编码 RNA 还可以通过顺式作用抑制相邻基因的启动子, 这种机制以 *Kcnq1/Cdkn1c* 印记基因簇的 *Kcnq1ot1* 和 *Igf2r/Airn* 印记基因簇的 *Airn* 为例^[11]. 这两种机制都不能独立解释基因亲本特异性表达的原因, 但两种机制结合起来能更好地理解印记区域基于顺式作用的调控机制.

除了反义 lncRNAs, 最近研究还表明位于 2 个编码蛋白基因间的 lncRNAs 和来源于增强子区域的增强子 RNAs (enhancer RNAs, eRNAs)也调控基因的亲本特异性表达. 因此本文主要选取 6 个印记机制研究比较透彻的印记区域, 包括 *Kcnq1/Cdkn1c*、*Igf2r/Airn*、*Prader-Willi (PWS)/Angelman (AS)*、*Snurf/Snrpn*、*Dlk1-Dio3* 和 *H19/Igf2*, 介绍了

* 国家自然科学基金资助项目(31372312).

** 通讯联系人.

Tel: 0312-7528270, E-mail: lishijie20005@163.com

收稿日期: 2015-12-20, 接受日期: 2016-03-15

3 种长非编码 RNA (反义 lncRNAs、基因间 lncRNAs 和增强子 RNAs) 在基因亲本来源特异性表达中的作用机制。印记位点的紊乱会导致一些疾病, 包括 Prader-Willi 综合征、Angelman 综合征、Beckwith-Wiedemann 综合征、Silver-Russell 综合征^[12]和一些恶性肿瘤^[13], 因此本文还介绍了印记的长非编码 RNA 在治疗人类印记缺陷疾病中的靶点作用。尽管小非编码 RNA (small noncoding RNAs), 例如 snoRNAs、miRNAs 和 piRNAs 等, 在基因组印记中也起到重要的调控作用, 但不作为本文的分析内容。

1 反义长非编码 RNA (antisense lncRNAs)

反义长非编码 RNAs (antisense lncRNAs) 在基因组印记的发生和维持中有重要作用。Antisense lncRNAs 作为分子间相互作用的支架, 通过募集染色质修饰机构调控印记基因簇中多个基因的表达。*Kcnq1ot1*、*Airn*、*Nespas* 和 *Ube3a-ATS* 是一些研究比较深入的印记反义长非编码 RNA。Antisense lncRNAs 在基因组印记中的作用方式可以作为理解基因组上依赖 lncRNA 对基因表达进行调控的范例。

1.1 *Kcnq1*/*Cdkn1c* 印记区域的长非编码 RNA——*Kcnq1ot1*

Kcnq1ot1 基因, 又称 *Lit1*, 是由 RNA 聚合酶 II (RNAP II) 转录的大小为 91 kb 的长反义转录物^[14]。*Kcnq1ot1* 位于鼠 7 号染色体末端的 *Kcnq1*/*Cdkn1c* 印记区域, 其同源区域位于人 11p22 染色体。*Kcnq1*/*Cdkn1c* 印记区域在基因组上跨度达 1 Mb, 包括 10~12 基因, 其中母源染色体表达的编码蛋白基因包括 2 种表达方式, 一种是在各个组织中普遍表达, 一种是在胎盘中特异表达 (图 1)。而反义 *Kcnq1ot1* 长非编码 RNA 由于启动子区在母源染色体上发生甲基化, 所以仅限于父源染色体表达^[15]。*Kcnq1ot1* 的转录方向与其所在的宿主基因 *Kcnq1* 的转录方向相反 (图 1)。

用细胞培养和转基因鼠模型系统研究发现, *Kcnq1ot1* 长非编码 RNA 通过与包括组蛋白甲基转移酶 (EZH2 和 G9a) 和 DNA 甲基转移酶 (DNMT1) 在内的染色质修饰物相互作用, 并募集这些修饰物, 实现以顺式作用沉默印记基因^[14, 16-18] (图 1)。这种沉默途径被认为是建立了高度有序的染色质分区, 使染色质上富含抑制型染色质所具有的修饰 (如

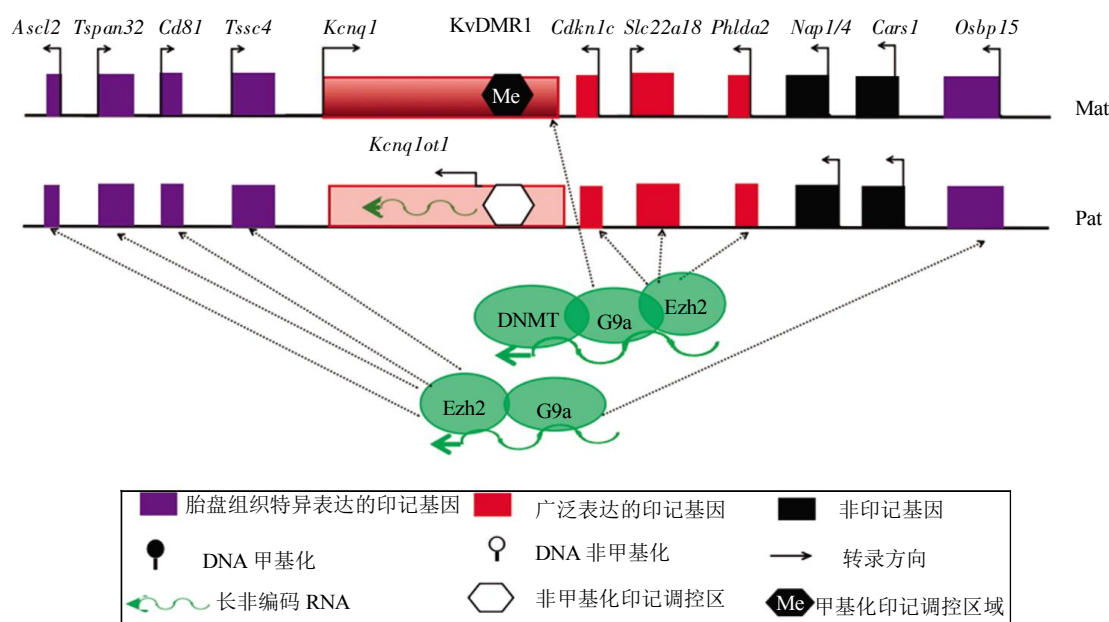


Fig. 1 *Kcnq1ot1* silences multiple genes by establishing higher-order-chromosomal interactions in cis in *Kcnq1*/*Cdkn1c* locus

图 1 *Kcnq1*/*Cdkn1c* 印记区域内的 *Kcnq1ot1* 通过建立高度有序的染色质相互作用顺式沉默多个基因的表达
Kcnq1ot1 的转录方向与其宿主基因 *Kcnq1* 的转录方向相反, *Kcnq1ot1* 启动子区在母源染色体上发生甲基化, 仅父源染色体表达。 *Kcnq1ot1* 通过募集组蛋白甲基转移酶 (EZH2 和 G9a) 和 DNA 甲基转移酶 (DNMT) 沉默基因表达。

H3K27me3、H3K9me3、H2AK119ub 和 DNA 甲基化), 而缺乏活跃型染色质所具有的标记物和 RNAP II^[16]. 鼠的 *Kcnq1ot1* RNA 在其 5' 端具有 890 bp 的沉默区域, 此区域中包括几个保守的重复序列^[19-21], 该重复序列可能通过将 *Kcnq1/Cdkn1c* 印记区域定位于富含沉默因子的核仁周边区域而实现沉默作用^[22]. 有实验证明, *Kcnq1ot1* RNA 上 890 bp 的沉默区域在染色质定位以及通过募集 DNA 甲基转移酶 DNMT1 而维持体细胞 DMRs 区 CpG 的甲基化中起重要作用^[17]. *Kcnq1ot1* 在广泛表达和胎盘特异表达的印记基因沉默的起始中都起作用, 但是 *Kcnq1ot1* 只参与维持广泛表达的印记基因的沉默, 而不参与维持胎盘特异表达的印记基因沉默^[23].

最近一篇研究报道了一种 RNA 介导的染色体构造捕获的方法(3C; R3C), 用此方法揭示了 *Kcnq1ot1* 通过顺式作用促进父源染色体上 *Kcnq1* 印记调控区(*Kcnq1* ICR, 又称 KvDMR1)和 *Kcnq1* 启动子内部染色质高度有序的相互作用, 而长距离调控包括 *Kcnq1* 基因在内的多个靶基因的沉默^[24]. 通过计算机模拟预测 *Kcnq1ot1* 基因的作用方式, 确定了 *Kcnq1ot1* 中的 DNA 结合结构域(TFO)以及 *Kcnq1ot1* 在靶基因上的结合位点(TTS), 发现 *Kcnq1ot1* 可能通过与靶基因启动子区以 Hoogsteen 氢键形成 RNA:DNA 三股螺旋而发挥作用^[25]. 通过计算机对基因组上的其他位点进行预测, 还发现依赖 lncRNA 的 DNA 识别模式不仅限于 *Kcnq1ot1* 基因, 也包括所有已知的通过表观修饰调控基因表达的 lncRNAs. 计算机预测的结果是否可以通过实验证实是非常令人关注的, 因为其可以帮助我们理解染色质调控 RNAs 的作用模式.

增强子的作用能够覆盖由 *Kcnq1ot1* 调控的印记基因的时空表达沉默. 一些受 *Kcnq1ot1* 调控的基因印记具有组织和发育阶段的特异性, 但其调控机制还没有被揭示. Korostowski 等^[26]的研究揭示了受 *Kcnq1ot1* 调控的 *Kcnq1* 基因如何逃脱由 *Kcnq1ot1* 调控的基因沉默. 在心脏发育过程中, *Kcnq1* 在 E13.5 中为母源等位基因表达, 而 E14.5 中转变为双等位基因表达. *Kcnq1* 基因由 E13.5 阶段单等位基因表达达到 E14.5 双等位基因表达的转变是由于一个上游心脏特异的增强子的激活导致的, 增强子随后与 *Kcnq1* 启动子通过形成高度有序的染色质环相互作用, 这个结论已经通过 3C 技术证

实. 这说明增强子调控的组织和发育阶段的特异性表达能够覆盖 lncRNAs 与基因启动子长距离相互作用导致的基因沉默.

1.2 *Igf2r/Airn* 印记区域的长非编码 RNA——*Airn*

长度约 500 kb 的 *Igf2r/Airn* 印记区域位于鼠 17 号染色体和人 6q26 染色体^[27], 该区域包含 3 个母源表达的蛋白编码基因 *Igf2r*、*Slc22a2* 和 *Slc22a3*, 以及一个父源表达的长非编码 RNA *Airn* (antisense to *Igf2r* RNA non-coding, *Airn*)(图 2). *Airn* 基因是其宿主基因 *Igf2r* 的反义转录物, 与 *Igf2r* 基因的启动子区重叠, 不与位于其上游的 *Slc22a2* 和 *Slc22a3* 基因重叠. *Airn* 转录产物约 118 kb, 其中 95% 不发生剪切^[28]. *Airn* 作为一个反式作用的沉默子, 沉默与其重叠的 *Igf2r* 和与其不重叠的 *Slc22a2* 与 *Slc22a3* 基因^[29]. 在人中, 除了在很少的胚胎组织和肾母细胞瘤(Wilm tumor)中 *Igf2r* 表现为单等位基因表达^[30], 而在绝大多数组织中, *Igf2r*、*Slc22a2* 和 *Slc22a3* 表现为双等位基因表达^[31]. Yotova 等^[32]鉴定了鼠 *Airn* 基因在人中的同源物(*Air*), 但其功能还没有被揭示.

研究证明 *Airn* 以不同于 *Kcnq1ot1* 的机制沉默 *Igf2r* 和 *Slc22a3* 基因, *Airn* 通过依赖转录(transcription-dependent)和依赖转录物(RNA-dependent)2 种方式顺式调控基因沉默. 在 ES 细胞中, 通过在距离 *Airn* 转录起始位点不同的位置插入转录终止信号, 发现导致 *Igf2r* 沉默的不是 *Airn* 基因的转录产物, 而是 *Airn* 基因的转录. 有趣的是, 依赖 *Airn* 转录的 *Igf2r* 的沉默是可逆的, 并且需要 *Airn* 的持续转录, 一旦 *Igf2r* 启动子区的 CpG 被甲基化修饰将导致 *Igf2r* 不可逆地沉默, *Igf2r* 的沉默也将不再依赖 *Airn* 转录^[33]. *Airn* 通过募集序列特异的染色质修饰物到非重叠基因的启动子上, 而沉默其表达^[34]. 因而 *Airn* 通过其转录物以及自身的转录来调控与其重叠以及不重叠基因(图 2). 这就提出一个问题, 导致目的基因沉默的是转录物还是转录? 是目的基因的启动子结构决定了 lncRNA 的作用方式吗? 仔细分析 *Igf2r* 基因的启动子结构对于揭示 lncRNA 的作用方式是有意义的. 还有一个值得注意的问题, DNMTs 是如何被募集到正在转录的 *Igf2r* 启动子上的? 这些问题的解决有助于我们理解依赖 lncRNA 转录的基因沉默.

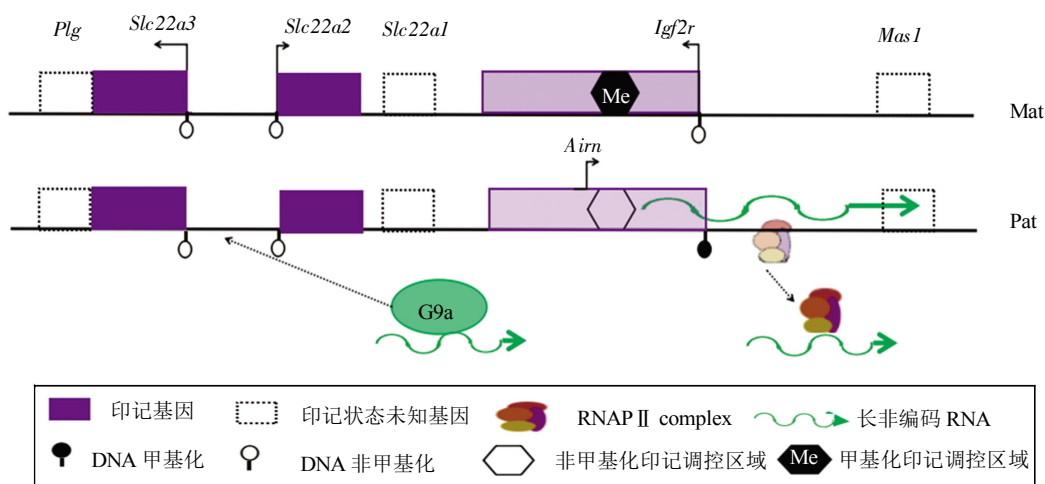


Fig. 2 *Airn* silences *Igf2r* and *Slc22a3* genes in cis in *Igf2r/Airn* locus

图2 *Igf2r/Airn* 印记区域的 *Airn* 顺式调控 *Igf2r* 和 *Slc22a3* 基因的沉默

位于 *Airn* 基因的启动子区 ICR 在母源染色体上甲基化, 父源染色体上非甲基化. *Airn* 通过阻止 *RNAP II* 复合物结合在与其重叠的 *Igf2r* 基因启动子区来沉默 *Igf2r* 的表达, *Airn* 通过募集组蛋白甲基转移酶 G9a 到非重叠基因 *Slc22a3* 的启动子上而沉默 *Slc22a3* 的表达.

1.3 Prader-Willi (PWS)/Angelman (AS) 印记区域的长非编码 RNA——*Ube3a-ATS*

长非编码 RNA *Ube3a-ATS* 位于与神经发育障碍相关的普拉德-威利/天使综合征(Prader-Willi/Angelman, PWS/AS)印记区域内, 该区域位于人的 15q11~q13 染色体, 跨越基因组大于 2 Mb 的区

域, 包括父源等位基因表达的 *Mkx3*、*Magel2*、*Ndn*、*Snrpn*、*Snord115*、*Snord116* 和一个母源等位基因表达的 *Ube3a*^[35].

PWS/AS 印记区域在鼠中是保守的, 位于 7 号染色体. PWS/AS 区域的印记中心 (imprinting center, IC) 由 PWS-IC 和 AS-IC 两部分组成(图 3),

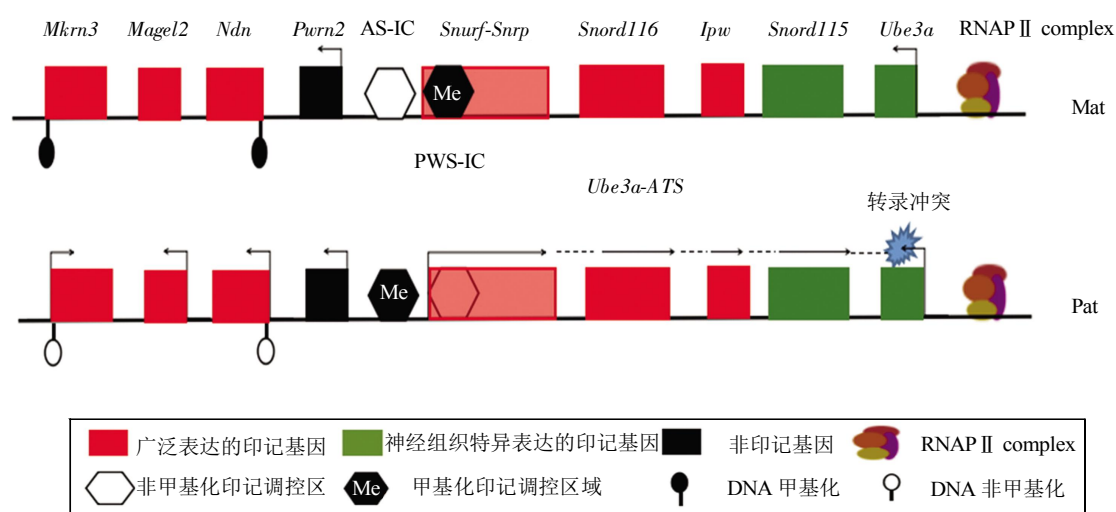


Fig. 3 Transcriptional collision of *Ube3a-ATS*-dependent in silencing of *Ube3a* in Prader-Willi (PWS)/Angelman (AS) locus

图3 Prader-Willi (PWS)/Angelman (AS) 印记区域的 *Ube3a-ATS* 依赖转录冲突沉默 *Ube3a* 基因的表达

在 *Ube3a* 的启动子区, 2 条亲本等位基因上都具有活跃启动子特异的组蛋白修饰(如 H3K4me3)以及 *RNAP II* 转录前起始复合物的形成. 在父源染色体上尽管 *Ube3a* 没有转录, 但其启动子仍维持活跃染色质状态, *Ube3a-ATS* 的转录通过了 *Ube3a* 的启动子, *Ube3a* 基因的沉默是通过与 *Ube3a-ATS* 的转录冲突实现的.

4.8 kb 的 PWS-IC 位于 *Snurf-Snrpn* 基因的启动子和外元 1, 通过激活其邻近的基因发挥作用, 而位于 *Snurf-Snrpn* 基因上游 35 kb 的 AS-IC(0.88 kb)的作用是抑制 PWS-IC, 进而在 PWS 基因上建立确切的 DNA 和组蛋白修饰. 父源染色体上 AS 基因 *Ube3a* 的活性受父源等位基因表达 *Ube3a-ATS* 的调控^[35-37].

选择性缺失 *Snrpn* 基因的启动子或用转录终止信号截短 *Ube3a-ATS* 能够使 *Ube3a-ATS* 的表达缺失并激活 *Ube3a* 的表达, 表明 *Ube3a-ATS* 和 *Ube3a* 之间具有相反的表达模式^[37-38]. *Ube3a-ATS* 是一种位于核中的转录物, RNA-DNA 原位杂交实验结果表明其成熟的转录物位于其转录单元的最近端. *Ube3a-ATS* 沉默 *Ube3a* 基因的分子机制并未表现出与 *Kcnq1ot1* 和 *Airn* 相同的在 RNA 水平上起作用. 在亲本等位基因上沉默的和活跃的 *Ube3a* 启动子都具有活跃启动子特异的组蛋白修饰(如 H3K4me3)并有 RNAP II 转录前起始复合物的形成, 而且来自双亲本等位基因上的 pre-mRNA 相同数量, 说明转录冲突是依赖 *Ube3a-ATS* 的 *Ube3a* 基因沉默的基础^[38](图 3). 然而在 AS 模型小鼠中, 用拓扑异构酶抑制剂和反义寡核苷酸(antisense oligonucleotides, ASO)处理可激活 *Ube3a* 的表达, 说明 *Ube3a-ATS* 在 *Ube3a* 基因的沉默中有一种可替代的模式. *Ube3a-ATS* 在 *Ube3a* 基因沉默中确切的作用机制还需进一步研究.

与 PWS 相关的 *Snord116* 位点目前备受关注. Powell 等^[39]研究了 *Snord116* 位点与 *Ube3a-ATS* 转录物之间的相互作用. 神经细胞特异表达的 *Ube3a-ATS* 基因的转录过程与父源 *Ube3a* 基因的沉默和染色质结构的去凝集相关. 在正常的神经元细胞中, 父源 *Snord116* 基因 G-skewed 位点 RNA 与 DNA 杂交时产生 R 环(R loops), R 环能够平衡 RNAP II 转录停止和转录延伸, 使得 *Ube3a-ATS* 基因转录和 *Ube3a* 基因沉默^[39]. *Snord116* 的宿主基因(*Snord116* host gene, 116HG)产生的 RNA 作为长非编码 RNA 云(lncRNA cloud)与代谢调控中的重要基因相互作用^[40].

2 基因间长非编码 RNA(intergenic lncRNAs)

有些印记区域中的长非编码 RNA 位于 2 个编码蛋白基因间, 被称为基因间长非编码 RNA (intergenic lncRNA), 在基因亲本来源特异性的表达

中起作用. *H19*、*Meg3* 和 *Ipw* 是研究得比较透彻的基因间 lncRNAs.

2.1 *H19/Igf2* 印记区域的长非编码 RNA——*H19*

H19 基因位于研究比较透彻的、约 80 kb 大小的 *H19/Igf2* 印记基因簇, 该基因簇定位于鼠的 7 号染色体和人的 11 号染色体, 包括非编码的 *H19* 基因和编码蛋白的 *Igf2* 基因^[41]. *H19* 基因编码约 2.3 kb 大小的 lncRNA, 是最先被鉴定的印记基因之一^[42-43]. *H19/Igf2* 印记基因簇在胚胎发育过程中起关键作用^[44-49], 由于父源等位基因启动子上 CpG 位点的甲基化, *H19* 只有来自于母源的等位基因表达. 应用转基因鼠模型研究 *H19* 基因在基因组印记中的功能, 发现敲除 *H19* 基因, 在内胚层组织中 *Igf2* 基因的印记不受影响, 而在中胚层组织中, *Igf2* 基因的印记部分消失, 这表明 *H19* 基因的功能具有细胞系依赖性^[50].

H19 在胚胎发育过程中表达量很高, 敲除 *H19* 基因不会导致胚胎致死但会在一定程度上影响生长^[51]. 然而在出生后, *H19* 除了在肌肉中维持高表达外, 在其他组织中表达量很低. 用未分化的多能间充质干细胞 C2C12 模型系统证实, *H19* 基因在肌源性分化中具有促进和抑制成肌分化两种相反的功能^[52-54], *H19* 通过两种机制抑制 C2C12 细胞肌源性的分化. Kallen 等^[52]研究表明 *H19* lncRNA 作为竞争性内源 RNA (competing endogenous RNA, CeRNA)控制 miRNA *let-7* 的水平, *H19* 的下调表达会导致 C2C12 细胞 *let-7* 依赖的成肌分化(图 4). Giovarelli 等^[54]发现, *H19* 充当 KH 型剪接调控蛋白(KH-type splicing regulatory protein, KSRP)的支架, 促进成肌素的降解, 进而抑制 C2C12 细胞分化成肌细胞(图 4). 然而与以上结论相反, *H19* 通过基因的外元 1 加工生成的 *miR-675-3p* 和 *miR-675-5p* 发挥促进肌源性分化的功能^[53](图 4).

H19 基因位于一个包括 16 个共表达印记基因的印记基因群(imprinted gene network, IGN)中, 此 IGN 包括众所周知的生长调控因子 *Igf2*、*Igf2r* 和 *Cdkn1c*^[55-56]. *H19* lncRNA 通过反式作用表观调控 IGN 中的成员从而调控胚胎生长. 最近有研究表明, *H19* 与甲基化 CpG 结合蛋白 MBD1(methyl-CpG-binding protein)相互作用, 将 H3K9 甲基转移酶募集到 IGN 中一些基因(包括 *Igf2*、*Slc38a4* 和 *Peg1* 基因)的 DMRs^[55].

此外, *H19* 还具有充当肿瘤抑制物^[57-58]和致

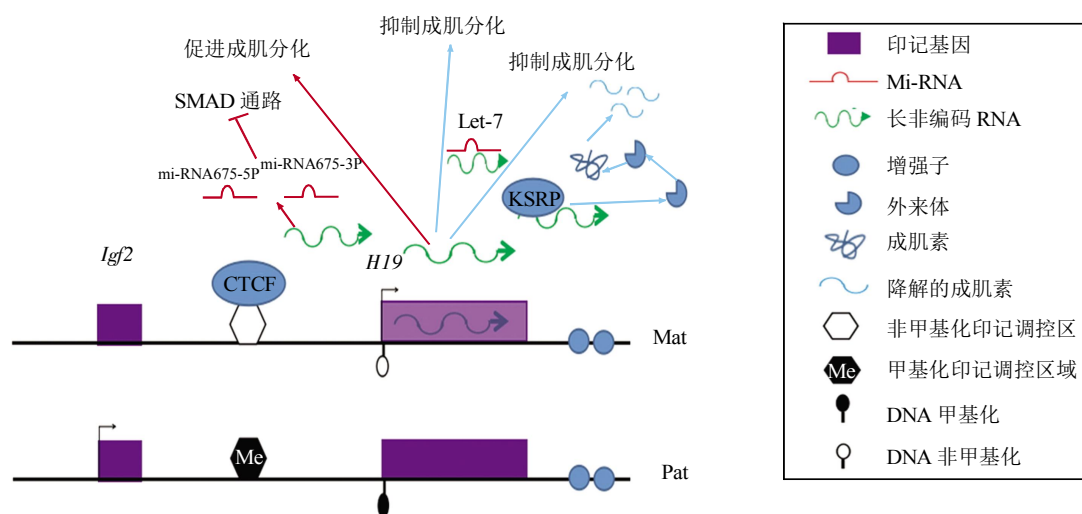


Fig. 4 The role of H19 in pro- and anti-myogenic functions in H19/Igf2 imprinted domain

图4 H19/Igf2 印记区域的长非编码 RNA H19 具有促进成肌分化和抑制成肌分化的功能

H19 lncRNA 作为 *let-7* miRNA 的竞争性内源 RNA 或充当 KSRP 蛋白的支架, 促进成肌素的降解, 从而抑制 C2C12 间质性干细胞分化成肌细胞。 *H19* 促进成肌分化的功能是通过 *H19* 转录物外元 1 加工生成的 *miRNA-675-3p* 和 *miRNA-675-5p* 实现的。

癌^[59-60]的功能。尽管 *H19* 在肿瘤发生与发展中的作用还存在争论, 但最近研究表明 *H19* 作为少数几种肿瘤的致癌 lncRNA, 其过表达能促进肿瘤的转移。综上所述, *H19* lncRNA 具有一些重要的生物学功能, 但这些功能是通过 *H19* 起的作用还是通过 *miR-675-3p* 和 *miR-675-5p* 起的作用, 或者是两者共同作用的结果还有待进一步研究。

2.2 Dlk1-Dio3 印记区域的长非编码 RNA——Meg3

母源等位基因表达的长非编码 RNA 基因 3 (maternally expressed gene 3, *Meg3*), 又称 *Gtl2* (gene trap locus2), 其所在的印记基因簇 *Dlk1-Dio3* 位于鼠的 12 号染色体和人的 14 号染色体上, 跨越基因组约 1 Mb, 该区域包含 3 个编码蛋白的父系表达基因 *Dlk1*、*Rtl1/Peg11* 和 *Dio3* 以及若干大小不同的母系表达非编码 RNA, 包括 snoRNA、miRNA 和长非编码 RNA *Meg3*^[61](图 5)。长非编码 RNA 可以通过不同剪接方式产生多种剪接体, 而不同剪接体的表达又具有组织特异性。据文献报道, 长非编码 RNA *Gtl2* 基因在人、小鼠和羊中分别有 12、7 和 5 种可变剪切体^[62-64], 在牛中, 我们研究了 *Gtl2* 基因的印记状态并发现了 6 种 *Gtl2* 基

因的可变剪切体^[65-66]。此外我们还分析了该区域中另外 2 个长非编码 RNA(*Meg8* 和 *Meg9*)的可变剪接和组织表达情况, 发现 *Meg8* 和 *Meg9* 基因在牛中分别具有 12 种和 3 种剪接体^[67-68]。然而关于这些长非编码 RNA 以及各种剪接体如何调控蛋白编码基因的表达还没有被揭示。

在 *Meg3* 基因上游有 3 个父源等位基因甲基化的 DMR 区, 3 个 DMR 区对该区域的印记影响程度不同^[61]。基因间的 (intergenic, IG)DMR 区是 *Dlk1-Dio3* 印记区域的印记调控区(ICR)。IG-DMR 调控 *Meg3* 的母源等位基因表达(图 5)。*Meg3* 和 *Dlk1-Dio3* 区域的其他非编码 RNAs 在鼠的成纤维细胞重编程为诱导多能干细胞(induced pluripotent stem cells, iPS)过程中起关键作用^[69]。*Meg3* 的表达可以作为 iPS 细胞处于具有支持胚胎发育的全能状态的标记。然而 *Meg3* 和 *Dlk1-Dio3* 区域的其他母源表达的印记基因如何促进 iPS 细胞具有完全的多能状态还不清楚。实验证据表明 *Meg3* 能够促进 PRC2 与 JARID2 的相互作用并使 PRC2/JARID2 复合物作用于基因组上的目的基因^[70]。

2.3 Snurf/Snrpn 印记区域的长非编码 RNA——Ipw

Ipw 是一个父源表达的 lncRNA, 位于最大的

印记基因簇 *Snurf/Snrpn* 基因簇(>2 Mb)中, 该印记区域位于鼠的 7 号染色体和人的 15 号染色体上, 在 70% 的普拉德 - 威利综合征 (Prader-Willi syndrome, PWS)病人中, 该染色体片段发生缺失^[71]. 鼠的 *Ipw* 基因只限制在大脑中表达, 而人的 *Ipw* 基因在所有组织中都表达. 人和鼠的 *Ipw* 基因的 5'端都有一个保守的 147 bp 的串联重复. *Ipw* 被认为是 1 Mb 的通读转录物加工产生的, 其包括 *Snord* 群和 *Ube3a* 基因^[72].

尽管人们对 *Snurf/Snrpn* 印记区域在 PWS/AS 综合征发病机理中的作用进行了比较深入的研究, 但 *Ipw* 基因的功能还没有被揭示. 近期有研究表明 *Ipw* 在精细调控人 14 号染色体上母源等位基因表达的基因中起重要作用. 通过比较来自正常人和

PWS 病人成纤维细胞的 iPS 细胞中的基因表达可以发现, 在来自 PWS 的 iPS 细胞中 *Ipw* 的表达水平非常低, 几个位于人 14 号染色体 *Dlk1-Dio3* 区域母源表达的基因 (maternally expressed genes, Megs), *Meg3*、*Mir370* 和 *Mir409* 的表达水平却显著上升, 而在来自正常人的 iPS 细胞中基因的表达没有发生变化^[73]. *Ipw* 是如何促进 *Megs* 的基因表达? *Ipw* 首先与组蛋白甲基转移酶 G9a 相互作用, 使 G9a 特异结合在 *Dlk1-Dio3* 区域的 IG-DMR 上, 使组蛋白 H3 的第 9 位赖氨酸(H3K9)发生甲基化, 促进 IG-DMR 区染色质结构改变, 形成异染色质 (图 5). 这是第一个长非编码 RNA 通过调控印记调控区的染色质结构而促进 2 个印记基因区域相互联系的例子.

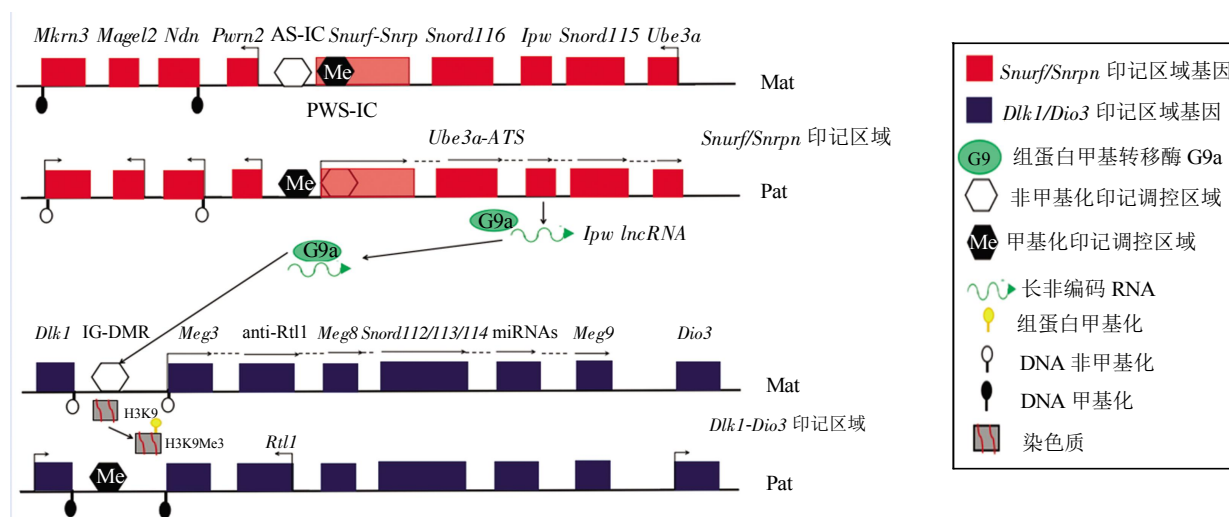


Fig. 5 *Ipw* lncRNA in *Snurf/Snrpn* imprinting domain promoting the repression of maternally expressed genes (Megs) in *Dlk1-Dio3* locus

图 5 *Snurf/Snrpn* 印记区域的长非编码 RNA *Ipw* 对 *Dlk1-Dio3* 印记区域母源等位基因的抑制

Dlk1-Dio3 印记区域 IG-DMR 区调控 *Meg3* 的母源等位基因表达. *Snurf/Snrpn* 印记区域的长非编码 RNA *Ipw* 与 G9a 甲基转移酶相互作用, 使 G9a 特异结合在 *Dlk1-Dio3* 印记区域 IG-DMR 上, 影响了 IG-DMR 促进 *Megs* 的基因表达.

3 增强子 RNA(eRNA)——IG-DMR RNA

增强子最初被认为是 DNA 上的一段序列, 能够远距离增强目的蛋白基因的表达. 增强子的基本特征包括富含 DNase- I 的高敏感性位点和 H3K4me¹、H3K4me² 和 H3K27ac 等组蛋白修饰. 大多数的增强子可进行双向转录, 但产生的转录物拷贝数很低, 5'端有帽子结构, 不发生剪接和多聚

腺苷酸化. 增强子 RNAs (enhancer RNAs, eRNA) 通过募集并(或)稳定基本转录复合物结合在目的基因的启动子区, 从而建立增强子和目的基因启动子区高度有序染色质作用来促进目的基因的表达^[74].

IG-DMR 是 *Dlk1-Dio3* 印记区域的印记调控区 (ICR). IG-DMR 的甲基化在亲本等位基因上存在差异, 父源染色体发生甲基化, 而母源染色体非甲基化. 非甲基化的 IG-DMR 被认为是具有增强子

特异的组蛋白修饰, 编码双向表达的非编码 RNAs. 非甲基化的 IG-DMR 对母源染色体上的非编码 RNAs 的表达起关键作用, 其中包括 *MEG3* 基因. 最近 Kota 等^[75]的研究发现从 IG-DMR 双向转录的产物调控 *Dlk1-Dio3* 区母源等位基因的表达. IG-DMR 编码的 eRNAs 表达紊乱会导致 IG-DMR 区的染色质上抑制基因表达的修饰物累积, 并使母源等位基因表达缺失. 在 ES 细胞中, IG-DMR 区母源染色体上双向表达的 eRNA 通过调控复制的起始时间和 *Dlk1-Dio3* 区的亚核定位来调控基因组印记. 母源染色体上 IG-DMR 编码的非编码 RNAs 的下调表达, 会使复制的起始时间从早到晚变化, 并使 *Dlk1-Dio3* 区向核边缘以及核内与基因沉默有关的子空间重新定位.

4 结语与展望

基因组印记是揭示依赖长非编码 RNA 进行表观遗传修饰现象的理想模型系统. 目前通过细胞培养结合转基因小鼠模型揭示了多个印记区域的 lncRNA 在哺乳动物发育过程中的功能. 研究比较透彻的印记区域中的 lncRNAs 在调控靶基因中的作用机制, 可以为揭示 lncRNAs 在其他生物环境中的作用模式提供基础. 大多数印记的长非编码 RNA 在功能和初始序列水平上相对高度保守, 因此可用于研究发育与疾病中的长非编码 RNA.

参 考 文 献

- [1] Bartolomei M S, Ferguson-Smith A C. Mammalian genomic imprinting. Cold Spring Harb Perspect Biol, 2011, **3**(7): a002592
- [2] Santoro F, Barlow D P. Developmental control of imprinted expression by macro non-coding RNAs. Semin Cell Dev Biol, 2011, **22**(4): 328–335
- [3] Mohammad F, Mondal T, Kanduri C. Epigenetics of imprinted long noncoding RNAs. Epigenetics, 2009, **4**(5): 277–286
- [4] Lee J T, Bartolomei M S. X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell, 2013, **152** (6): 1308–1323
- [5] Barlow D P. Genomic imprinting: A mammalian epigenetic discovery model. Annu Rev Genet, 2011, **45**: 379–403
- [6] Royo H, Cavaille J. Non-coding RNAs in imprinted gene clusters. Biol Cell, 2008, **100**(3): 149–166
- [7] Mohammad F, Mondal T, Kanduri C. Epigenetics of imprinted long noncoding RNAs. Epigenetics, 2009, **4**(5): 277–286
- [8] Sleutels F, Zwart R, Barlow D P. The non-coding air RNA is required for silencing autosomal imprinted genes. Nature, 2002, **415**(6873): 810–813
- [9] Hark A T, Schoenherr C J, Katz D J, et al. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature, 2000, **405**(6785): 486–489
- [10] Bell A C, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature, 2000, **405**(6785): 482–485
- [11] Kanduri C. Long noncoding RNA and epigenomics. Adv Exp Med Biol, 2011, **722**(5): 174–195
- [12] Li Y, Sasaki H. Genomic imprinting in mammals: Its life cycle, molecular mechanisms and reprogramming. Cell Res, 2011, **21**(3): 466–473
- [13] Uribe-Lewis S, Woodfine K, Stojic L, et al. Molecular mechanisms of genomic imprinting and clinical implications for cancer. Expert Rev Mol Med, 2011, **13**(67): 1–6
- [14] Pandey R R, Mondal T, Mohammad F, et al. Kcnq1 ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell, 2008, **32**(2): 232–246
- [15] Mancini-Dinardo D, Steele S J, Levorse J M, et al. Elongation of the Kcnq1 ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev, 2006, **20**(10): 1268–1282
- [16] Terranova R, Yokobayashi S, Stadler M B, et al. Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Dev Cell, 2008, **15**(5): 668–679
- [17] Mohammad F, Mondal T, Guseva N, et al. Kcnq1 ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. Development, 2010, **137**(15): 2493–2499
- [18] Redrup L, Branco M R, Perdeaux E R, et al. The long noncoding RNA Kcnq1 ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. Development, 2009, **136**(4): 525–530
- [19] Mohammad F, Pandey R R, Nagano T, et al. Kcnq1 ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. Mol Cell Biol, 2008, **28**(11): 3713–3728
- [20] Whitehead J, Pandey G K, Kanduri C. Regulation of the mammalian epigenome by long noncoding RNAs. Biochim Biophys Acta, 2009, **1790**(9): 936–947
- [21] Magistri M, Faghihi M A, St Laurent G, et al. Regulation of chromatin structure by long noncoding RNAs: Focus on natural antisense transcripts. Trends Gene, 2012, **28**(8): 389–396
- [22] Zakharova I S, Shevchenko A I, Zakian S M. Monoallelic gene expression in mammals. Chromosoma, 2009, **118**(3): 279–290
- [23] Mohammad F, Pandey G K, Mondal T, et al. Long noncoding RNA-mediated maintenance of DNA methylation and transcriptional gene silencing. Development, 2012, **139** (15): 2792–2803
- [24] Zhang H, Zeitz M J, Wang H, et al. Long noncoding RNA-mediated intrachromosomal interactions promote imprinting at the Kcnq1 locus. J Cell Biol, 2014, **204**(1): 61–75
- [25] He S, Zhang H, Liu H, et al. LongTarget: a tool to predict lncRNA DNA-binding motifs and binding sites via Hoogsteen base-pairing analysis. Bioinformatics, 2015, **31**(2): 178–186
- [26] Korostowski L, Sedlak N, Engel N. The Kcnq1 ot1 long non-coding RNA affects chromatin conformation and expression of Kcnq1, but

- does not regulate its imprinting in the developing heart. PLoS Genet, 2012, **8**(9): e1002956
- [27] Guenzl P M, Barlow D P. Macro lncRNAs: A new layer of cis-regulatory information in the mammalian genome. RNA Biol, 2012, **9**(6): 731–741
- [28] Barlow D P. Genomic imprinting: A mammalian epigenetic discovery model. Annu Rev Genet, 2011, **45**: 379–403
- [29] Koerner M V, Pauler F M, Huang R, *et al.* The function of non-coding RNAs in genomic imprinting. Development, 2009, **136**(11): 1771–1783
- [30] Oudejans C B, Westerman B, Wouters D, *et al.* Allelic *Igf2r* repression does not correlate with expression of antisense RNA in human extraembryonic tissues. Genomics, 2001, **73**(3): 331–337
- [31] Kalscheuer V M, Mariman E C, Schepens M T, *et al.* The insulin-like growth factor type-2 receptor gene is imprinted in the mouse but not in humans. Nat Genet, 1993, **5**(1): 74–78
- [32] Yotova I Y, Vlatkovic I M, Pauler F M, *et al.* Identification of the human homolog of the imprinted mouse Air non-coding RNA. Genomics, 2008, **92**(6): 464–473
- [33] Latos P A, Pauler F M, Koerner M V, *et al.* *Airn* transcriptional overlap, but not its lncRNA products, induces imprinted *Igf2r* silencing. Science, 2012, **338**(6113): 1469–1472
- [34] Nagano T, Mitchell J A, Sanz L A, *et al.* The air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science, 2008, **322**(5908): 1717–1720
- [35] Royo H, Cavaille J. Non-coding RNAs in imprinted gene clusters. Biol Cell, 2008, **100**(3): 149–166
- [36] Landers M, Calciano M A, Colosi D, *et al.* Maternal disruption of Ube3a leads to increased expression of Ube3a-ATS in trans. Nucleic Acids Res, 2005, **33**(13): 3976–3984
- [37] Meng L, Person R E, Beaudet A L. Ube3a-ATS is an atypical RNA polymerase II transcript that represses the paternal expression of Ube3a. Hum Mol Genet, 2012, **21**(13): 3001–3012
- [38] Meng L, Person R E, Huang W, *et al.* Truncation of Ube3a-ATS unsilences paternal Ube3a and ameliorates behavioral defects in the Angelman syndrome mouse model. PLoS Genet, 2013, **9** (12): e1004039
- [39] Powell W T, Coulson R L, Gonzales M L, *et al.* R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. Proc Natl Acad Sci USA, 2013, **110**(34): 13938–13943
- [40] Powell W T, Coulson R L, Crary F K, *et al.* A Prader-Willi locus lncRNA cloud modulates diurnal genes and energy expenditure. Hum Mol Genet, 2013, **22**(21): 4318–4328
- [41] Bartolomei M S, Ferguson-Smith A C. Mammalian genomic imprinting. Cold Spring Harb Perspect Biol, 2011, **3**(7): a002592
- [42] Bartolomei M, Zemel S, Tilghman S. Parental imprinting of the mouse *H19* gene. Nature, 1991, **351**(6322): 153–155
- [43] Matouk I, Raveh E, Ohana P, *et al.* The increasing complexity of the oncofetal *H19* gene locus: Functional dissection and therapeutic intervention. Mol Sci, 2013, **14**(2): 4298–4316
- [44] Kanduri C, Holmgren C, Franklin G, *et al.* The 5' -flank of themurine *H19* gene in an unusual chromatin conformation unidirectionally blocks enhancer-promoter communication. Curr Biol, 2000, **10**(8): 449–457
- [45] Thakur N, Tiwari V K, Thomassin H, *et al.* An antisense RNA regulates the bidirectional silencing property of the *Kcnq1* imprinting control region. Mol Cell Biol, 2004, **24**(18): 7855–7862
- [46] Sleutels F, Zwart R, Barlow D P. The non-coding air RNA is required for silencing autosomal imprinted genes. Nature, 2002, **415**(6873): 810–813
- [47] Hark A T, Schoenherr C J, Katz D J, *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. Nature, 2000, **405**(6785): 486–489
- [48] Bell A C, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. Nature, 2000, **405**(6785): 482–485
- [49] Williamson C M, Ball S T, Dawson C, *et al.* Uncoupling antisense-mediated silencing and DNA methylation in the imprinted *Gnas* cluster. PLoS Genet, 2011, **7**(3): e1001347
- [50] Bartolomei M, Zemel S, Tilghman S. Parental imprinting of the mouse *H19* gene. Nature, 1991, **351**(6322): 153–155
- [51] Barlow D P, Bartolomei M S. Genomic imprinting in mammals. Cold Spring Harb Biol, 2014, **6**(2): a018382
- [52] Kallen A N, Zhou X B, Xu J, *et al.* The imprinted *H19* lncRNA antagonizes let-7 microRNAs. Mol Cell, 2013, **52**(1): 101–112
- [53] Dey B K, Pfeifer K, Dutta A. The *H19* long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration. Genes Dev, 2014, **28**(5): 491–501
- [54] Giovarelli M, Bucci G, Ramos A, *et al.* *H19* long noncoding RNA controls the mRNA decay promoting function of KSRP. Proc Natl Acad Sci USA, 2014, **111**(47): 5023–5028
- [55] Monnier P, Martinet C, Pontis J, *et al.* *H19* lncRNA controls gene expression of the imprinted gene network by recruiting MBD1. Proc Natl Acad Sci, 2013, **110**(51): 20693–20698
- [56] Varrault A, Gueydan C, Delalbre A, *et al.* *Zac1* regulates an imprinted gene network critically involved in the control of embryonic growth. Dev Cell, 2006, **11** (5): 711–722
- [57] Yoshimizu T, Miroglio A, Ripoché M A, *et al.* The *H19* locus acts *in vivo* as a tumor suppressor. Proc Natl Acad Sci USA, 2008, **105**(34): 12417–12422
- [58] Hao Y, Crenshaw T, Moulton v, *et al.* Tumour-suppressor activity of *H19* RNA. Nature, 1993, **365**(6448): 764–767
- [59] Moulton T, Crenshaw T, Hao Y, *et al.* Epigenetic lesions at the *H19* locus in Wilms' tumour patients. Nat Genet, 1994, **7**(3): 440–447
- [60] Adriaenssens E, Dumont L, Lottin S, *et al.* *H19* overexpression in breast adenocarcinoma stromal cells is associated with tumor values and steroid receptor status but independent of p53 and Ki-67 expression. Am J Pathol, 1998, **153**(5): 1597–1607
- [61] McMurray E N, Schmidt J V. Identification of imprinting regulators at the *Meg3* differentially methylated region. Genomics, 2012, **100**(3): 184–194
- [62] Schuster-Gossler K, Bilinski P, Sado T, *et al.* The mouse *Gtl2* gene

- is differentially expressed during embryonic development, encodes multiple alternatively spliced transcripts, and may act as an RNA. *Dev Dyn*, 1998, **212**(2): 214–228
- [63] Zhang X, Gejman R, Mahta A, *et al.* Maternally expressed gene 3, an imprinted noncoding RNA gene, is associated with meningioma pathogenesis and progression. *Cancer Res*, 2010, **70**(6): 2350–2358
- [64] Bidwell C A, Shay T L, Georges M, *et al.* Differential expression of the *GTL2* gene within the callipyge region of ovine chromosome 18. *Anim Genet*, 2001, **32**(5): 248–256
- [65] 王建春, 李冬杰, 杜卫华, 等. *GTL2* 基因在体细胞核移植牛中的印记状态分析. *生物化学与生物物理进展*, 2008, **35**(10): 1183–1187
- Wang J C, Li D J, Du W H, *et al.* *Prog Biochem Biophys*, 2008, **35**(10): 1183–1187
- [66] Su H, Li D J, Hou X H, *et al.* Molecular structure of bovine *Gtl2* gene and DNA methylation status of *Dlk1-Gtl2* imprinted domain in cloned bovines. *Animal Reproduction Science*, 2011, **127**(1–2): 23–30
- [67] Hou X H, Li D J, Su H, *et al.* Molecular cloning, expression, and imprinting status of Maternally expressed gene 8 (Meg 8) in dairy cattle. *Russian Journal of Genetics*, 2011, **47**(8): 994–998
- [68] Zhang K, Li D J, Wang M N, *et al.* The differential expression of alternatively spliced transcripts and imprinting status of *MEG9* gene in cows. *Animal Genetics*, 2014, **45**(5): 660–664
- [69] Stadtfeld M, Apostolou E, Akutsu H, *et al.* Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature*, 2010, **465**(7295): 175–181
- [70] Kaneko S, Bonasio R, Saldana-Meyer R, *et al.* Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol Cell*, 2014, **53**(2): 290–300
- [71] Wevrick R, Kerns J A, Francke U. Identification of a novel paternally expressed gene in the Prader-Willi syndrome region. *Hum Mol Genet*, 1994, **3**(10): 1877–1882
- [72] Meng L, Person R E, Beaudet A L. Ube3a-ATS is an atypical RNA polymerase II transcript that represses the paternal expression of Ube3a. *Hum Mol Genet*, 2012, **21**(13): 3001–3012
- [73] Stelzer Y, Sagi I, Yanuka O, *et al.* The noncoding RNA IPW regulates the imprinted *DLK1-DIO3* locus in an induced pluripotent stem cell model of tPrader-Willi syndrome. *Nat Genet*, 2014, **46**(6): 551–557
- [74] Darrow E M, Chadwick B P. Boosting transcription by transcription: enhancer-associated transcripts. *Chromosome Res*, 2013, **21**(6–7): 713–724
- [75] Kota S K, Lleres D, Bouchet T, *et al.* ICR noncoding RNA expression controls imprinting and DNA replication at the *Dlk1-Dio3* domain. *Dev Cell*, 2014, **31**(1): 19–33

The Function of Long Non-coding RNAs in Genomic Imprinting*

YANG Wen-Zhi¹⁾, ZHANG Cui¹⁾, WANG Guan-Nan¹⁾, LI Dong-Jie²⁾, LI Shi-Jie^{1)**}

¹⁾ College of Life Science, Agricultural University of Hebei, Baoding 071001, China;

²⁾ Hebei University of Science and Technology, Shijiazhuang 050018, China)

Abstract Long noncoding RNAs (lncRNA) are non-protein coding transcripts longer than 200 nucleotides. The field of long noncoding RNA (lncRNA) research has been rapidly advancing in recent years as its large number and important biological functions. Genomic imprinting is an epigenetic phenomenon. The lncRNAs play a critical role in important biological functions by establishing genomic imprinting of target genes. Genomic imprinting has been a great resource for studying transcriptional and post-transcriptional-based gene regulation by lncRNAs. This review will focus on the mechanisms of imprinting in six of the most well-studied imprinted gene clusters: *Kcnq1/Cdkn1c*, *Igf2r/Airn*, Prader-Willi (PWS)/Angelman (AS), *Snurf/Snrpn*, *Dlk1-Dio3*, *H19/Igf2*, and overview the functional role of antisense lncRNAs (*Kcnq1ot1*, *Airn* and *Ube3a-ATS*), intergenic lncRNAs (*H19*, *IPW* and *MEG3*), and enhancer lncRNAs (IG-DMR eRNAs) to understand the diverse mechanisms being employed by them in *cis* and (/or) *trans* to regulate the parent-of-origin-specific expression of target genes. A better understanding of these downstream mechanisms will help to improve our general understanding of the function of ncRNAs throughout the genome.

Key words long noncoding RNA, genomic imprinting, chromatin, epigenetic modification

DOI: 10.16476/j.pibb.2015.0391

* This work was supported by a grant from The National Natural Science Foundation of China(31372312).

**Corresponding author.

Tel: 86-312-7528270, E-mail: lishijie20005@163.com

Received: December 20, 2015 Accepted: March 15, 2016