

长链非编码RNA与DNA形成三链结构发挥调控作用的研究进展

尹铮^{1#}, 都美阳^{2#}, 王仲璇³, 靳腾喻¹, 胡潇³, 刘悦林¹, 孙绍光^{2*}

(¹河北医科大学基础医学院临床医学系, 石家庄 050017; ²河北医科大学基础医学院生物化学与分子生物学教研室, 石家庄 050017; ³河北医科大学公共卫生学院预防医学系, 石家庄 050017)

摘要: 长链非编码RNA(lncRNA)是一类长度大于200个核苷酸的非编码RNA, 通常可以作为一种支架促进多种蛋白质、RNA和DNA相互作用。其中, 有些lncRNA可以通过Hoogsteen碱基配对直接与富含嘌呤的DNA结合, 形成RNA-DNA三链结构来发挥生物学功能, 如调控多种疾病的发生与发展。本文介绍了能够预测RNA-DNA三链结构的一些工具(如Long-Target、Triplexator、TDF、Triplex、Triplex-Inspector), 同时重点概述了lncRNA与DNA形成的RNA-DNA三链结构在疾病发生与发展过程中的具体作用机制, 以期多种疾病的研究与治疗提供新思路及见解。

关键词: 长链非编码RNA; 三链结构; 预测; 疾病

Research progress on regulation of long non-coding RNA by forming triple helix structures with DNA

YIN Zheng^{1#}, DU Meiyang^{2#}, WANG Zhongxuan³, JIN Tengyu¹, HU Xiao³,
LIU Yuelin¹, SUN Shaoguang^{2*}

(¹Department of Clinical Medicine, College of Basic Medicine, Hebei Medical University, Shijiazhuang 050017, China; ²Department of Biochemistry and Molecular Biology, College of Basic Medicine, Hebei Medical University, Shijiazhuang 050017, China; ³Department of Preventive Medicine, Hebei Medical University, Shijiazhuang 050017, China)

Abstract: Long non-coding RNA (lncRNA) is a class of non-coding RNA with a length of more than 200 nucleotides. It can usually be used as a scaffold to facilitate interactions between various proteins, RNA and DNA. Some lncRNAs can directly combine with purine-rich DNA through Hoogsteen base pairing to form RNA-DNA triple helix structure to perform biological functions, such as regulating the occurrence and development of various diseases. This review introduces some tools that can predict RNA-DNA triple helix structures in detail (such as Long-Target, Triplexator, TDF, Triplex, Triples-Inspector). Meanwhile, the specific mechanism of RNA-DNA triple helix structures formed by lncRNA and DNA in the occurrence and development of diseases was emphatically summarized, which is expected to provide new ideas and insights for the research and treatment of various diseases.

Key Words: long non-coding RNA; triple helix structure; prediction; disease

收稿日期: 2022-02-08

基金项目: 国家自然科学基金项目(82170439, 81670273, 81200215); 河北省自然科学基金项目(H2021206399, H2019206150); 大学生创新性实验计划项目(USIP2020040, USIP2020018)

#共同第一作者: 尹铮, E-mail: 1243625342@qq.com; 都美阳, E-mail: 18730281061@163.com

*通信作者: E-mail: sunshaoguang00@163.com

长链非编码RNA(long non-coding RNA, lncRNA)是长度大于200个核苷酸的非编码RNA,主要由RNA聚合酶Ⅱ转录,缺乏明显的开放阅读框架^[1-5]。lncRNA通过多种机制参与蛋白质编码和表观遗传基因的调控,如表观遗传修饰、选择性剪接、转录后和翻译调控^[6]。lncRNA可以跨越下游蛋白编码基因的启动子区域,干扰转录因子与DNA启动子的结合,从而抑制基因的表达。lncRNA可诱导DNA甲基化和异染色质形成,导致肿瘤抑制基因沉默^[7]。相反,lncRNA可以通过结合转录因子或激活辅助蛋白来促进靶基因的表达^[8],也能通过激活转录因子或诱导蛋白质复合体的形成来抑制基因表达^[9]。lncRNA不仅可以通过与输入蛋白质相互作用调节转录因子的亚细胞定位,也可以充当前体细胞小分子RNA或通过与小RNA结合来调节其活性^[10]。此外,lncRNA还可以通过影响染色质结构来控制mRNA功能和非编码转录,从而调节基因活性,如核小体位置和染色体循环^[11]。这些研究大都集中在lncRNA及其蛋白伴侣在基因表达调控中的相互作用,但是很少关注RNA-DNA三链结构的形成及其位点。lncRNA可识别并结合特定的双链DNA序列形成三链结构^[12]。近年来,越来越多的科研工作者开始深入探究三链结构的形成位点、调控功能及其在重大疾病中的作用机制,为疾病的诊断与治疗提供理论支持。

1 三链结构的形成

DNA双螺旋结构发现后不久,Felsenfeld等^[13]提出了DNA具有形成三链结构的可能性,即两条嘧啶链和一条嘌呤链将形成更复杂的结构。1986年,有人证明,短混合序列的三链形成寡核苷酸可以形成稳定的特异性三链DNA复合物^[14]。三链结构可以通过第三条链与Watson-Crick碱基对之间的Hoogsteen氢键或反向Hoogsteen氢键形成^[15](图1A)。可以基于第三链的方向形成两种类型的三链结构:平行结构和反平行结构。平行方向建立Hoogsteen相互作用,而反平行方向建立反向Hoogsteen相互作用。三链结构也可以标记为嘧啶或嘌呤基序,该命名法反映了第三条链中的主要碱基:嘧啶基序富含U-或C-,而嘌呤基序富含A-

或G-。一般来说,嘧啶基序倾向于平行取向,而嘌呤基序倾向于反平行取向^[16-18](图1B)。长期以来,分子间三链结构的形成被认为可能是控制细胞过程的重要机制,如在体内可以抑制蛋白质-DNA相互作用和功能过程,包括转录调节、染色质修饰;在体外可以影响转录后RNA加工和DNA修复^[19]。在生物体内有一类广泛存在的三链结构——R环(R-loop)。当某些基因转录形成mRNA分子难与模板链分离时,会形成RNA-DNA杂交体,这时非模板链与RNA-DNA杂交体共同构成R环结构。具体来说,R-环更多地形成于GC或富含嘌呤的区域,并且最常见于启动子附近或端粒基因末端^[20]。R环在许多生理过程中起重要作用,如线粒体复制启动^[21]、DNA修复^[22]和基因表达调控^[23-25]等。此外,研究发现,这些三链结构的形成与许多人类疾病如癌症、神经性疾病和心血管疾病等密切相关^[26]。

2 三链结构的预测

lncRNA在多种细胞过程中发挥重要作用,在医学和药物设计中具有重要意义^[26]。因此,分别用实验方法和计算方法来识别和预测这类分子的作用机制是非常重要的。然而,为了避免昂贵的实验方法,多种计算算法已经被开发出来以准确和快速地预测lncRNA与DNA形成的三链体结构。

2.1 三重域查找器

三重域查找器(triple domain finder, TDF)是一种预测RNA与DNA结合的软件。具体预测步骤是:首先在DNA区域发现lncRNA的DNA结合位点(DNA binding site, DBS)和RNA结合位点(RNA binding site, RBS)。接下来,通过发现具有重叠RBS的RNA中的区域来列出所有候选DNA结合结构域(DNA binding domain, DBD)。基本操作基于基因组区域,该区域是基因组中具有特定位置的线性区间。基因组区域测试评估具有预定DNA区域(目标区域)的lncRNA的三链结构结合潜力。由于缺少非目标区域,通过随机选择大小/长度与目标区域相同的基因组区域来生成多个非目标区域。对于给定的DBD,使用经验统计检验来评估具有至少一个DBS的目标区域的数量是否大于具有最少一个DBS的“随机”非目标区域的数量^[27]。

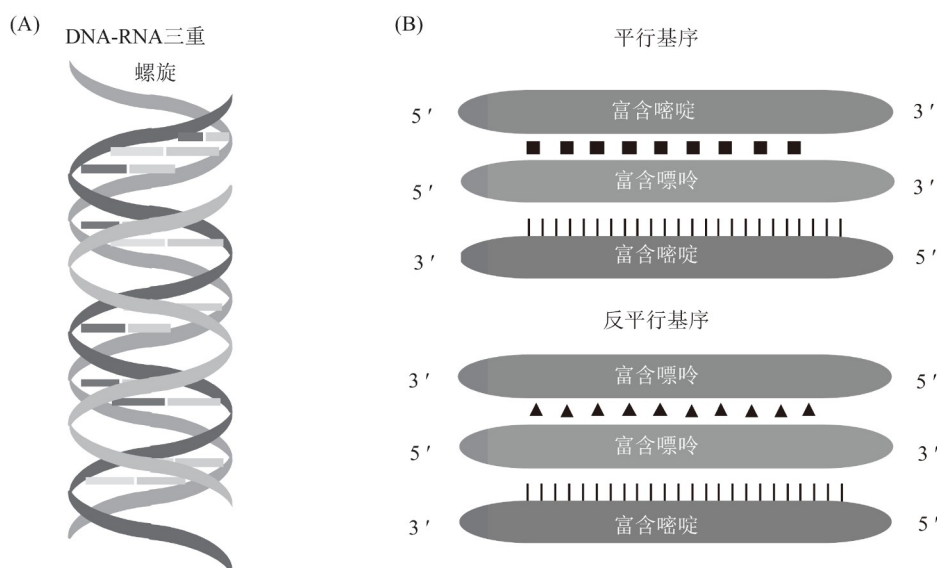


图1 RNA-DNA三链结构的形成

三重域查找器允许评估靶向一组常见DNA区域的多个lncRNA的三链结构。该方法可用于探索性分析,即评估在特定分化过程中差异表达的所有lncRNA。然后,在同一组靶区域(即差异表达基因的启动子区域或具有特定染色质标记的区域)中评估这些lncRNA。三重域查找器提供统计信息,如富集的DBD数量和对候选lncRNA进行排名的三链靶位点(triple helix target site, TTS)数量。由于lncRNA的大小可以变化(此处分析的数据为200~1 000个核苷酸),可以通过碱基数对统计数据进行了标准化,以避免较大的lncRNA产生偏差^[27]。

三重域查找器还提供了对目标区域和启动子进行排名的简单方法。基于DBS的两种统计数据为:与该区域关联的DBS的数量以及至少一个DBS覆盖的碱基对的百分比。三重域查找器还允许纳入实验证据,如基因表达倍数变化或RNA纯化法分离染色质测序(chromatin isolation by RNA purification-sequence, ChIRP-Seq)峰的得分作为排名的外部标准。对于启动子测试,三重域查找器还显示了与启动子相关的基因;对于基因组区域测试,在该区域的至少50 kb附近具有上游和下游基因^[5]。三重域查找器是检测和表征这些分析所预测的三链结构形成的RNA的独特工具,使我们能更好地剖析结合DNA的lncRNA的机制和功能。

2.2 Triplexator

Triplexator是一个用于查询三链结构形成的所

有三个“支柱”的计算工具包。即识别单链寡核苷酸中潜在的TFO(triplex-forming oligonucleotide);在双链核苷酸序列中鉴定合适的靶位点(TFO target sites, TTS);最重要的是根据上述规范三链结构形成规则评估潜在合作伙伴的兼容性。这种算法能够满足一组用户定义约束的所有可能的最大匹配。而且,它支持灵活的错误模型,其中允许非规范三链结构的容忍数量随着预测的三链结构的长度增加而增长。这样的错误率是由于观察到更长的三链结构通过更多的氢键而稳定的结果。Triplexator的范围和效率能够通过挖掘下一代测序数据来获取三链形成的足迹,从而解决体内三链结构形成的存在和功能这一基本生物学问题。Triplexator的灵活性还可以实现以生物技术或基因治疗为重点的应用程序驱动的基因组研究^[28]。

在可用性方面, Triplexator更加灵活:一些基本配对规则可以被用户排除;此外, Triplexator可以处理与末端不同的三链错配,这可能反映了结合的稳定性的。总之, Triplexator在预测精度和可用性方面明显优于其他计算方法^[29]。

2.3 Long-Target

Long-Target是一种通过Hoogsteen碱基配对分析预测lncRNA和DNA结合基序和结合位点的工具。对于目标基因组区域, Long-Target在每个碱基配对规则集上为其正负链构建两条RNA链。lncRNA可以与DNA双链体的任一链结合,因此,

预测的结合基序和它们的结合位点(即TFO和TTS)具有相同的序列。由于lncRNA可能包含数千个核苷酸,通常会报告多个TFO,尤其是针对多个规则集预测lncRNA时。通常,除TFO1外的其他TFO都是明显的假阳性,因为它们生成的三链结构少得多且重叠程度很差,因此不会影响预测结果^[30-32]。

Long-Target接受一对lncRNA和DNA序列作为输入,并输出9个文件,包括排序TFO及其三链结构的列表;报告排列测试结果的文件;报告TFO1、TFO2和TFO3的简单统计信息的文件;描述TTS序列的文件以及预测TFO1和TFO2生成的位置的文件。软件总共有八个参数:TT罚分、CC罚分、平均稳定性、比对同一性、最小TFO长度、序列随机混洗(非随机混洗)、随机混洗时间和偏移。前五个参数用于在用户定义的条件过滤三链结构;序列混洗和混洗时间用于置换测试;偏移量限制了三链结构中点之间的距离,以控制它们的重叠对三链结构进行排序^[31]。

Long-Target软件主要立足于三点:全面的Hoogsteen和Hoogsteen碱基配对规则、局部对比,以及对所有TFO/TTS预测的分析。目前,通过系统回顾已经整理出了24条碱基配对规则集,对于一段感兴趣的DNA区域,根据每一条碱基配对规则集重构四条RNA序列,并通过局部比对识别一个长链非编码RNA的DNA结合域和这段DNA区域中的长链非编码RNA结合位点。用置换检验来评估预测结果的敏感性与专一性。和Triplexator的预测结果相比,Long-Target预测出来的DNA结合位点位于目标基因的启动子区域、CpG岛和转座子区域,而且与ChIP-seq实验揭示的染色质组蛋白甲基化区域高度吻合。相比而言,Triplexator的许多预测结果偏离了启动子区域等重要已知调控元件^[31]。

2.4 Triplex-Inspector

Triplex-Inspector是用于序列特异性配体设计和适当目标选择的计算工具,可提高基于三链结构的基因组操作的准确性。该软件利用详尽的搜索算法,获取感兴趣的基因组基因座,并搜索所有推定的三链结构目标位点。随后,通过在基因组中寻找与它们具有高度序列相似性的任何基因座,来检查这些推定靶标中每个靶标的独

特性^[32]。

用户可以分别自定义目标站点搜索和评估使用的参数,而且还可以通过使用在目标站点搜索期间使用的相同参数设置来检测与目标站点具有高度相似性的离线目标。放宽脱靶搜索标准可以检测较少相似的脱靶位点,从而更全面地了解细胞毒性和其他意外后果的总体风险。Triplex-Inspector允许基于基因注释数据对相关目标位点进行交互式过滤,通过考虑染色质的组织,可以进一步优化目标配体设计。同时允许DNase I超敏性数据用于突出显示由于染色质可及性(或不可及性)而脱靶结合的可能性更大(或更少)的基因组区域。当前不存在用于任何给定应用程序中选择最佳三链目标靶点的单一标准,因此,手动、交互式数据检查至关重要。Triplex-Inspector将所有结果汇总到一个交互式Web浏览器应用程序中,从而可以在不同平台之间共享直观的数据操作,同时可以为每个目标站点簇生成一个图,从而使用户可以直观地评估其特性,一旦用户确定了具有可接受的脱靶风险的靶标,Triplex-Inspector将根据已有的模型输出匹配配体的序列^[33,34]。

3 lncRNA形成的RNA-DNA三链结构与疾病

上述预测方法证实了基因组中存在大量三链体形成基序,lncRNA通过形成RNA-DNA三链结构进行基因调控的作用位点正在逐渐浮现。研究表明,lncRNA形成的RNA-DNA三链结构参与了多种重要的细胞过程如表观遗传、转录调控和染色体失活等^[35],lncRNA在疾病发生过程中的重要作用正在被逐渐发现。因此,lncRNA的多种作用机制中可能隐含着各种疾病的潜在治疗靶点。

3.1 lncRNA PCAT92通过形成RNA-DNA三链体促进前列腺癌发展

现已发现许多lncRNA在前列腺癌中表达失调,这些lncRNA通过在染色质重塑、转录调控或转录后调控等过程中发挥功能作用而参与癌症发展^[36]。Bawa等^[37]发现,腺癌相关转录本92(PCAT92)位于第13号染色体上与前列腺癌有关的ABCC4基因附近。敲低PCAT92可抑制细胞增殖及ABCC4蛋白的表达,表明PCAT92通过调节ABCC4

在前列腺癌中发挥作用。LncRNA PCAT92和ABCC4之间的共享区域具有致癌转录因子(ZIC2)的结合位点。而ZIC2与预测的ABCC4启动子的结合已使用下拉分析得到证实。然而,在PCAT92敲低条件下,ZIC2与启动子区域的结合减少,而ZIC2转录水平没有任何变化。这表明PCAT92可能直接或间接在将ZIC2募集到ABCC4基因座中发挥作用。使用NanoITC可验证PCAT92转录物的预测区域与ABCC4启动子区域附近的预测染色体DNA之间的相互作用。据观察结果推测,PCAT92可能通过在ABCC4启动子附近形成RNA-DNA三链体来增加ZIC2的局部浓度以进行ABCC4的调节,从而在前列腺癌中发挥作用(图2)。

3.2 LncRNA MIR100HG通过与p27基因座形成RNA-DNA三链体促进三阴性乳腺癌细胞增殖

乳腺癌作为女性最常见的恶性肿瘤,可分为多种分子亚型,其中三阴性乳腺癌(TNBC)最具侵袭性,复发风险高。由于TNBC细胞缺乏孕激素受体(PR)、雌激素受体(ER)和表皮生长因子受体-2(HER-2),与其他类型的乳腺癌相比,其治疗方案的选择历来有限^[38]。因此,阐明与TNBC的发生与发展相关的lncRNA的调控机制并开发相应的靶向治疗药物显得尤为重要^[39]。

在之前的研究发现,MIR100HG与一些癌症的进展有关^[40-42],敲低MIR100HG可抑制细胞增殖并诱导G₁期细胞停滞,而MIR100HG的过表达则显著促进细胞增殖。Wang等^[43]发现,MIR100HG的过表达显著增加了TNBC细胞生长,其敲低则显著

干扰了肿瘤的生长。细胞周期相关基因在MIR100HG敲低组 and 对照组之间发生了显著变化,其中CDKN1B编码p27蛋白。p27蛋白是一种肿瘤抑制因子,可调节细胞周期进程^[44]。通过生物信息学预测,Wang等^[43]在MIR100HG中发现了三个与p27特异性结合形成RNA-DNA三链体结构的TFO。其中,MIR100HG的TFO1是p27基因座上的功能结合位点。MIR100HG通过TFO1序列与p27基因位点结合形成RNA-DNA三链体结构,可能由此募集表观遗传修饰的蛋白质直接结合并激活p27启动子来调节TNBC中的细胞增殖。

3.3 LncRNA REG1CP通过与REG3A启动子形成RNA-DNA三链体促进结直肠癌细胞增殖

再生胰岛衍生(REG)基因家族的产物是许多细胞生物学过程的重要调节因子,包括防止细胞死亡和促进细胞增殖^[45]。而该家族中的一种非蛋白质编码产物lncRNA REG1CP的功能仍需进一步研究。Yari等^[46]发现,REG1CP在结直肠癌细胞中经常上调,REG1CP促进REG3A的转录,并通过REG3A发挥促进细胞增殖的作用。生物信息学分析显示,REG3A基因包含转录起始位点上游的高嘌呤区域,而该区域恰与REG1CP中所含的TFO片段互补,其结合也在实验中得到证实。另外,RecQ解旋酶家族成员范可尼贫血组J蛋白(FANCI)可与REG1CP上的G-四链体结合。-2325/-2509片段可能是REG3A启动子的转录活性区,FANCI与此片段的结合需要REG1CP和REG3A启动子之间的关联,在此之后它可以解开REG3A启动子活性区

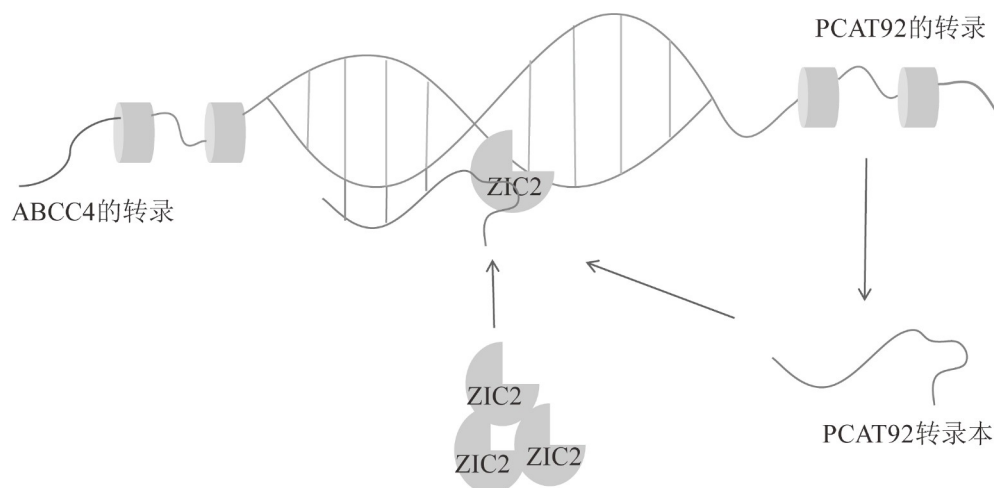


图2 PCAT92通过蛋白质ZIC2调节ABCC4的机制

的双链DNA。REG3A启动子的-2325/-2509片段存在糖皮质激素受体 α (GR α)的多个结合基序,并且GR α 的敲低降低了REG3A水平,表明GR α 对于REG3A的转录激活是必需的。以上研究结果表明,REG1CP通过在REG3A基因的远端启动子处与其形成RNA-DNA三链体,募集DNA解旋酶FANCI, FANCI通过REG1CP上的G-四链体与REG3A的核心启动子的2325/-2509片段相连,解开双链DNA并解除转录抑制,促进糖皮质激素受体 α 介导的REG3A转录,从而促进结直肠癌细胞增殖。

3.4 LncRNA CDKN2B-AS1通过形成RNA-DNA三链体促进动脉粥样硬化的进展

动脉粥样斑块的形成是动脉粥样硬化的重要过程,CDKN2B-AS1可促进动脉粥样硬化斑块的形成并抑制动脉粥样硬化中的巨噬细胞逆向胆固醇转运(mRCT),从而促进动脉粥样硬化的进程。研究表明,CDKN2B-AS1的高表达或CDKN2B的低表达标志着动脉粥样硬化。RT-qPCR结果表明,CTCF的过表达可以减少CDKN2B mRNA的转录,且CTCF和CDKN2B-AS1的联合过表达进一步抑制了CDKN2B mRNA的转录。因此,CTCF可能通过直接结合CDKN2B启动子区域抑制CDKN2B转录。CDKN2B-AS1和EZH2蛋白可以相互结合,EZH2可以被招募到CDKN2B的启动子区域,并且

可以受CTCF和CDKN2B-AS1的调控抑制CDKN2B的表达。EZH2通过在CTCF和CDKN2B-AS1的调控下催化H3K27三甲基化来抑制CDKN2B的表达。Long-Target网站(<http://lncrna.smu.edu.cn>)预测CDKN2B-AS1中发现的TFO与CDKN2B可以特异性结合形成RNA-DNA三链体结构。因此,CDKN2B-AS1通过与CDKN2B启动子区域形成RNA-DNA三链体来影响动脉粥样硬化(图3)。这种结构有助于在CDKN2B启动子区域募集EZH2和CTCF,从而抑制CDKN2B转录。CDKN2B-AS1促进CTCF与CDKN2B启动子结合,同时通过与CDKN2B启动子形成RNA-DNA三链体并募集EZH2与CDKN2B启动子结合,加速H3K27me3甲基化抑制CDKN2B表达^[47,48]。因此,CDKN2B-AS1形成的RNA-DNA三链体可以促进动脉粥样斑块的形成,可能成为动脉粥样硬化的潜在治疗靶点。

3.5 LncRNA DLX6-AS1通过形成RNA-DNA三链体介导子宫内膜癌细胞的增殖

子宫内膜癌是全世界最常见的妇科癌症之一,lncRNA与其发展有很强的关联性。Zhao等^[49]发现,lncRNA DLX6-AS1和DLX6之间的调控机制与子宫内膜癌的发展密切相关。LncRNA DLX6-AS1和DLX6在子宫内膜癌组织或细胞系中高表达,沉默lncRNA DLX6-AS1或DLX6可抑制肿瘤生长并促进细胞凋亡。p300是一种组蛋白乙酰转移

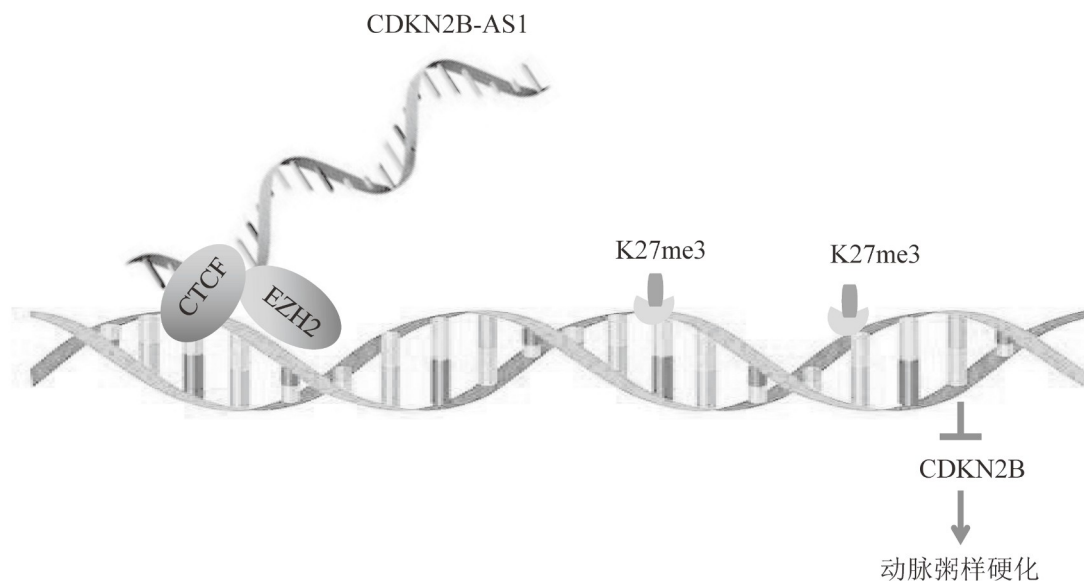


图3 CDKN2B-AS1通过形成RNA-DNA三链体调节动脉粥样硬化的机制

酶,可增加组蛋白的乙酰化和松散的染色质结构以促进转录的启动。通过ChIP可以检测出 $DLX6$ 启动子区域p300及转录因子E2F1的富集,并且通过JASPAR网站(<http://jaspar.genereg.net/>)预测了E2F1和 $DLX6$ 启动子之间的结合序列gcggcgggaag。根据Long-Target网站(<http://lncrna.smu.edu.cn/>)的预测, $DLX6$ -AS1和 $DLX6$ 可能通过Hoogsteen形成RNA-DNA三链体。在发现lncRNA $DLX6$ -AS1和 $DLX6$ 在子宫内膜癌组织或细胞系中均富集、沉默lncRNA $DLX6$ -AS1或 $DLX6$ 可抑制肿瘤生长并促进细胞凋亡的基础上,实验证明, $DLX6$ -AS1以RNA-DNA三链体的形式与 $DLX6$ 启动子结合,募集p300和E2F1形成RNP复合物,RNP复合物可以改变启动子的表观遗传状态并提高 $DLX6$ 的转录水平,从而促进子宫内膜癌细胞的增殖(图4)。

3.6 LncRNA LINC00525通过形成RNA-DNA三链体促进肺腺癌细胞增殖

肺腺癌(LUAD)是肺癌目前最普遍的组织学类型。实验结果显示,LINC00525在LUAD细胞和组织中高度表达,LINC00525过表达促进LUAD细胞的细胞周期进程和细胞增殖。而LINC00525的表达与 $p21$ mRNA的表达呈负相关,启动子荧光素酶测定显示 $p21$ (一种 G_1/S 状态相关基因启动子)在LINC00525敲低后的转录活性明显增加。因此,

LINC00525可能通过抑制 $p21$ 表达来促进细胞增殖。Long-Target程序预测了LINC00525中的两个潜在TFO和 $p21$ 启动子中相应的结合位点。并且,LINC00525与EZH2的结合通过RIP检测得到进一步证实^[50]。据报道,EZH2通过H3K27me3组蛋白修饰在表观遗传上调节靶基因表达^[51]。因此,LINC00525促进LUAD细胞的机制为,LINC00525通过与 $p21$ 形成RNA-DNA三链体,将EZH2引导至 $p21$ 启动子,从而在表观遗传上抑制 $p21$ 转录,继而调节LUAD细胞的细胞周期进程,从而促进LUAD的进展。

3.7 LncRNA KPRT4通过形成RNA-DNA三链体结构促进基底样乳腺癌细胞增殖

基底样乳腺癌(BLBC)是乳腺癌中最具侵袭性的亚型。有研究表明,KPRT4是BLBC进展的主要分子。Krüppel样因子5(KLF5)则是BLBC中的关键致癌转录因子。KLF5通过直接与KPRT4的启动子结合以促进KPRT4转录。相反,KPRT4又可使KLF5的转录增强。具体机制为,KPRT4在其5'结构域与转录因子YB-1相互作用,并在其3'结构域形成RNA-DNA三链体结构,将YB-1募集到KLF5的启动子,促进KLF5的转录,并最终建立前馈电路以促进细胞增殖^[52]。LncRNA在不同疾病中的作用及机制见表1。

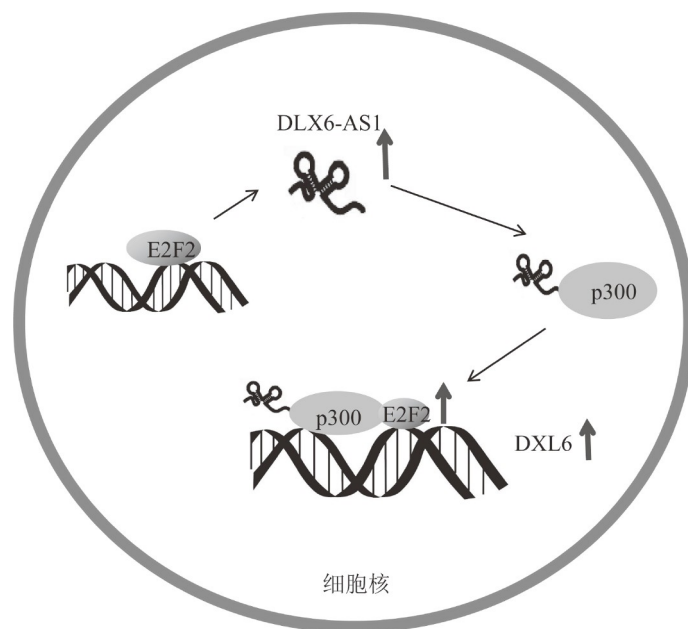


图4 $DLX6$ -AS1通过形成RNA-DNA三链体介导子宫内膜癌细胞增殖的机制

表1 RNA-DNA三链结构与疾病

LncRNA	DNA	作用
PCAT92	ABCC4	增加ZIC2的局部浓度, 促进前列腺癌发展 ^[37]
MIR100HG	p27	激活p27启动子, 促使TNBC细胞增殖 ^[43]
REG1CP	REG3A	募集DNA解旋酶FANCI, 促进糖皮质激素受体 α 介导的结直肠癌细胞增殖 ^[46]
CDKN2B-AS1	CDKN2B	抑制CDKN2B转录, 促进动脉粥样斑块形成 ^[48]
DLX6-AS1	DLX6	募集p300蛋白以促进DLX6的转录, 促进子宫内膜癌细胞的增殖 ^[49]
LINC00525	p21	抑制p21转录, 促进肺腺癌细胞增殖 ^[50]
KPRT4	KLF5启动子	将YB-1募集到KLF5的启动子, 促进KLF5的转录, 并最终建立前馈电路以促进细胞增殖 ^[52]

4 展望

LncRNA已被证明在表观遗传控制和转录、翻译、RNA代谢、干细胞维持和分化、细胞自噬和凋亡以及胚胎发育方面发挥重要作用^[53], 而LncRNA发挥作用的机制也成为备受关注的研究方向。其中, LncRNA与特定DNA形成的三链结构在染色质修饰、转录调控等方面发挥作用。通过本文所列的一系列检测三链体、统计表征DNA结合域的方法包括Long-Target、Triplexator、TDF、Triplex和Triplex-Inspector, 可以明确RNA-DNA三链体的作用方式。其中, Triplexator在目前研究中使用较多, 其范围更广、效率更高, 且更具灵活性, 所以越来越受到学者的青睐。当前预测三链结构形成的各种方法主要是根据基本配对规则, 依赖数学统计。但是这些方法存在一定的局限性: 一是它们虽能识别大量可形成三链结构的LncRNA, 但经过实验验证者数量有限; 其次, 预测仅考虑了理论关系, 而缺乏经过实验验证的数据支持, 表明实际并非所有预测出的LncRNA都能实际形成三链结构; 另外, 这些工具可以检测与特定LncRNA相互作用的DNA区域, 但并未揭示相互作用机制^[54]。2020年, 有研究者开发了一个名为TriplexFPP(三链结构形成潜力预测)的集成程序, 这是DNA-RNA三链结构预测中的第一个机器学习模型。TriplexFPP能够从所有具有确定的三链结构形成能力的LncRNA中预测最可能形成三链结构的LncRNA^[55]。它可能为探索LncRNA功能提供新的见解。

研究表明, LncRNA与肿瘤、神经性疾病、心血管等疾病密切相关^[56-58]。其中多项研究表明,

LncRNA对于癌症的发生与发展具有正向调控作用, 而目前对于LncRNA是否参与其他各类疾病的发病机制, 我们知之甚少, 仍待进一步证实。LncRNA在疾病治疗和预后中有巨大的应用潜力, LncRNA PCAT92、MIR100HG、REG1CP、CDKN2B-AS1、DLX6-AS1、LINC00525及KPRT4等均为相应疾病的治疗提供了潜在的、可发展的治疗靶点。

在以往研究中, LncRNA多通过与DNA的启动子区域形成三链体结构发挥基因表达调控作用。有研究表明, 某些LncRNA通过隔离它们的3'末端逃避细胞降解机制, 从而支持疾病进展。如卡波西肉瘤相关疱疹病毒(KSHV)多腺苷酸化核(PAN) LncRNA中的核表达元件(ENE)三链体, 这种分子机制同样也代表了治疗开发的方向^[59]。相对于LncRNA的其他作用机制, 三链结构的应用研究还处于初级阶段, 但相信随着LncRNA研究技术的改进及三链结构预测和检测方法的完善, 人们将会对LncRNA形成的三链结构在疾病中的调控机制有更深入的了解, 为疾病的诊断、治疗及预后开辟一条新的道路。

参考文献

- [1] Chen LL. Linking long noncoding RNA localization and function. *Trends Biochem Sci*, 2016, 41(9): 761-772
- [2] Goff LA, Rinn JL. Linking RNA biology to lncRNAs. *Genome Res*, 2015, 25(10): 1456-1465
- [3] Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*, 2012, 81(1): 145-166
- [4] Sentürk Cetin N, Kuo CC, Ribarska T, et al. Isolation and genome-wide characterization of cellular DNA:RNA

- triplex structures. *Nucleic Acids Res*, 2019, 47(5): 2306-2321
- [5] Huangfu N, Xu Z, Zheng W, et al. LncRNA MALAT1 regulates oxLDL-induced CD36 expression via activating β -catenin. *Biochem Biophys Res Commun*, 2018, 495(3): 2111-2117
- [6] Martens JA, Laprade L, Winston F. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature*, 2004, 429(6991): 571-574
- [7] Hirota K, Miyoshi T, Kugou K, et al. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature*, 2008, 456(7218): 130-134
- [8] Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell*, 2009, 136(4): 629-641
- [9] Feng J, Bi C, Clark BS, et al. The *Eyf-2* noncoding RNA is transcribed from the *Dlx-5/6* ultraconserved region and functions as a *Dlx-2* transcriptional coactivator. *Genes Dev*, 2009, 20(11): 1470-1484
- [10] 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): 1570-1573
- [11] Böhmendorfer G, Wierzbicki AT. Control of chromatin structure by long noncoding RNA. *Trends Cell Biol*, 2005, 25(10): 623-632
- [12] Pasquier C, Agnel S, Robichon A. The mapping of predicted triplex DNA: RNA in the *Drosophila* genome reveals a prominent location in development- and morphogenesis-related genes. *G3 Genes Genomes Genet*, 2017, 7(7): 2295-2304
- [13] Felsenfeld G, Rich A. Studies on the formation of two- and three-stranded polyribonucleotides. *Biochim Biophys Acta*, 1957, 26(3): 457-468
- [14] Dervan PB. Design of sequence-specific DNA-binding molecules. *Science*, 1986, 232(4749): 464-471
- [15] Brown JA. Unraveling the structure and biological functions of RNA triple helices. *WIREs RNA*, 2020, 11(6): e1598
- [16] Letai AG, Palladino MA, Fromm E, et al. Specificity in formation of triple-stranded nucleic acid helical complexes: studies with agarose-linked polyribonucleotide affinity columns. *Biochemistry*, 1988, 27(26): 9108-9112
- [17] Durland RH, Kessler DJ, Gunnell S, et al. Binding of triple helix forming oligonucleotides to sites in gene promoters. *Biochemistry*, 1991, 30(38): 9246-9255
- [18] Maldonado R, Längst G. Analyzing RNA-DNA triplex formation in chromatin. *Methods Mol Biol*, 2020, 2161: 247-254
- [19] Hartono SR, Malapert A, Legros P, et al. The affinity of the S9.6 antibody for double-stranded RNAs impacts the accurate mapping of r-loops in fission yeast. *J Mol Biol*, 2018, 430(3): 272-284
- [20] Posse V, Al-Behadili A, Uhler JP, et al. RNase H1 directs origin-specific initiation of DNA replication in human mitochondria. *PLoS Genet*, 2019, 15(1): e1007781
- [21] Teng Y, Yadav T, Duan M, et al. ROS-induced R loops trigger a transcription-coupled but BRCA1/2-independent homologous recombination pathway through CSB. *Nat Commun*, 2018, 9(1): 4115
- [22] Stolz R, Sulthana S, Hartono SR, et al. Interplay between DNA sequence and negative superhelicity drives R-loop structures. *Proc Natl Acad Sci USA*, 2019, 116(13): 6260-6269
- [23] Bonnet A, Grosso AR, Elkaoutari A, et al. Introns protect eukaryotic genomes from transcription-associated genetic instability. *Mol Cell*, 2017, 67(4): 608-621
- [24] Hamperl S, Bocek MJ, Saldivar JC, et al. Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell*, 2017, 170(4): 774-786
- [25] Lang KS, Hall AN, Merrih CN, et al. Replication-transcription conflicts generate R-loops that orchestrate bacterial stress survival and pathogenesis. *Cell*, 2017, 170(4): 787-799
- [26] Emamjomeh A, Zahiri J, Asadian M, et al. Identification, prediction and data analysis of noncoding RNAs: a review. *Mol Cell*, 2019, 15(3): 216-230
- [27] Kuo CC, Hänzelmann S, Sentürk Cetin N, et al. Detection of RNA-DNA binding sites in long noncoding RNAs. *Nucleic Acids Res*, 2019, 47(6): e32
- [28] Buske FA, Bauer DC, Mattick JS, et al. Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res*, 2012, 22(7): 1372-1381
- [29] Antonov IV, Mazurov E, Borodovsky M, et al. Prediction of lncRNAs and their interactions with nucleic acids: benchmarking bioinformatics tools. *Briefings BioInf*, 2019, 20(2): 551-564
- [30] Li W, Cowley A, Uludag M, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res*, 2015, 43(W1): W580-W584
- [31] 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
- [32] Goujon M, McWilliam H, Li W, et al. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res*, 2010, 38(Web Server): W695-W699
- [33] Buske FA, Bauer DC, Mattick JS, et al. Triplex-Inspector: an analysis tool for triplex-mediated targeting of genomic loci. *Bioinformatics*, 2013, 29(15): 1895-1897

- [34] Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*, 2009, 19(9): 1639-1645
- [35] Arun G, Diermeier SD, Spector DL. Therapeutic targeting of long non-coding RNAs in cancer. *Trends Mol Med*, 2018, 24(3): 257-277
- [36] Janaththani P, Srinivasan SL, Batra J. Long non-coding RNAs at the chromosomal risk loci identified by prostate and breast cancer GWAS. *Genes*, 2021, 12(12): 2028
- [37] Bawa PS, Ravi S, Paul S, et al. A novel molecular mechanism for a long non-coding RNA PCAT92 implicated in prostate cancer. *Oncotarget*, 2018, 9(65): 32419-32434
- [38] Bou Zerdan M, Ghorayeb T, Saliba F, et al. Triple negative breast cancer: updates on classification and treatment in 2021. *Cancers*, 2022, 14(5): 1253
- [39] Chen FY, Zhou ZY, Zhang KJ, et al. Long non-coding RNA MIR100HG promotes the migration, invasion and proliferation of triple-negative breast cancer cells by targeting the miR-5590-3p/OTX1 axis. *Cancer Cell Int*, 2020, 20(1): 508
- [40] Lu Y, Zhao X, Liu Q, et al. lncRNA MIR100HG-derived miR-100 and miR-125b mediate cetuximab resistance via Wnt/ β -catenin signaling. *Nat Med*, 2017, 23(11): 1331-1341
- [41] Emmrich S, Streltsov A, Schmidt F, et al. lncRNAs MONC and MIR100HG act as oncogenes in acute megakaryoblastic leukemia. *Mol Cancer*, 2014, 13(1): 171
- [42] Shang C, Zhu W, Liu T, et al. Characterization of long non-coding RNA expression profiles in lymph node metastasis of early-stage cervical cancer. *Oncol Rep*, 2016, 35(6): 3185-3197
- [43] Wang S, Ke H, Zhang H, et al. lncRNA MIR100HG promotes cell proliferation in triple-negative breast cancer through triplex formation with p27 loci. *Cell Death Dis*, 2018, 9(8): 805
- [44] Xie Q, Zhang R, Liu D, et al. Apigenin inhibits growth of melanoma by suppressing miR-512-3p and promoting the G₁ phase of cell cycle involving the p27 Kip1 protein. *Mol Cell Biochem*, 2022, 477(5): 1569-1582
- [45] Wang G, Cheng L, Chen M, et al. Comparative analysis of expression profiles of reg signaling pathways-related genes between AHF and HCC. *Biochem Genet*, 2019, 57(3): 382-402
- [46] Yari H, Jin L, Teng L, et al. lncRNA REG1CP promotes tumorigenesis through an enhancer complex to recruit FANCI helicase for REG3A transcription. *Nat Commun*, 2019, 10(1): 5334
- [47] Horswell SD, Fryer LGD, Hutchison CE, et al. CDKN2B expression in adipose tissue of familial combined hyperlipidemia patients. *J Lipid Res*, 2013, 54(12): 3491-3505
- [48] Ou M, Li X, Zhao S, et al. Long non-coding RNA CDKN2B-AS1 contributes to atherosclerotic plaque formation by forming RNA-DNA triplex in the CDKN2B promoter. *EBioMedicine*, 2020, 55: 102694
- [49] Zhao H, Xu Q. Long non-coding RNA DLX6-AS1 mediates proliferation, invasion and apoptosis of endometrial cancer cells by recruiting p300/E2F1 in DLX6 promoter region. *J Cell Mol Med*, 2020, 24(21): 12572-12584
- [50] Fang P, Chen H, Ma Z, et al. lncRNA LINC00525 suppresses p21 expression via mRNA decay and triplex-mediated changes in chromatin structure in lung adenocarcinoma. *Cancer Commun*, 2021, 41(7): 596-614
- [51] Laugesen A, Højfeldt JW, Helin K. Molecular mechanisms directing PRC2 recruitment and H3K27 methylation. *Mol Cell*, 2019, 74(1): 8-18
- [52] Du G, Sun J, Li Z, et al. A feedforward circuit between KLF5 and lncRNA KPRT4 contributes to basal-like breast cancer. *Cancer Lett*, 2022, 534: 215618
- [53] Ali T, Grote P. Beyond the RNA-dependent function of lncRNA genes. *Elife*, 2020, 9: e60583
- [54] Matveishina E, Antonov I, Medvedeva YA. Practical guidance in genome-wide RNA: DNA triple helix prediction. *Int J Mol Sci*, 2020, 21(3): 830
- [55] Zhang Y, Long Y, Kwok CK. Deep learning based DNA: RNA triplex forming potential prediction. *BMC Bioinf*, 2020, 21(1): 522
- [56] Fenoglio C, Ridolfi E, Galimberti D, et al. An emerging role for long non-coding RNA dysregulation in neurological disorders. *Int J Mol Sci*, 2013, 14(10): 20427-20442
- [57] Uchida S, Dimmeler S. Long noncoding RNAs in cardiovascular diseases. *Circ Res*, 2015, 116(4): 737-750
- [58] Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. *Cancer Cell*, 2016, 29(4): 452-463
- [59] Swain M, Ageeli AA, Kasprzak WK, et al. Dynamic bulge nucleotides in the KSHV PAN ENE triple helix provide a unique binding platform for small molecule ligands. *Nucleic Acids Res*, 2021, 49(22): 13179-13193