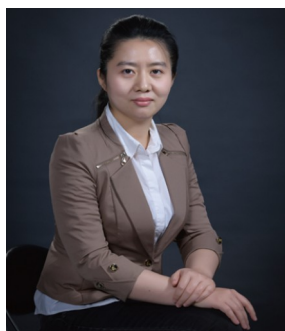


综述



任捷, 研究员、博士生导师, 2010年于美国纽约州立大学石溪分校获得博士学位, 随后在冷泉港实验室进行博士后研究, 2018年12月回国, 任中国科学院北京基因组所(国家生物信息中心)研究员。主要研究领域为染色质动态调控, 课题组以复制和转录在时空上相互干扰为切入点, 从遗传和表观遗传协同调控这一全新的角度研究其作用机制, 发现核酸修饰和高级结构调控染色质功能、维护基因组和表观遗传稳定性、从而延缓细胞衰老的新机制。系列原创性研究成果以第一或通讯(含共同)作者发表在*Cell*、*Nature*、*Cell Research*、*Cell Stem Cell*、*Nature Medicine*、*Nature Aging*、*Molecular Cell*等期刊上。目前主要开展干细胞稳态中的染色质调控及基因组稳定性、单分子测序技术开发的研究。

R-loop功能及应用前沿展望

苏思毅^{1,2#}, 徐莹莹^{1,2#}, 张丽媛^{1,2*}, 任捷^{1,2*}

(¹核糖核酸功能与应用重点实验室, 国家生物信息中心, 北京 100101;

²中国科学院北京基因组研究所, 北京 100101)

摘要: R-loop是一种特殊的三链结构, 由RNA侵入双链DNA形成, 包括RNA-DNA杂交链和单链DNA。R-loop广泛存在于原核与真核生物的基因组中, 参与调控基因转录、DNA复制和DNA损伤修复等多种生物学过程。然而, 异常的R-loop形成和积累可能导致基因组不稳定, 并与多种疾病发生相关。由于不同染色质环境下R-loop的差异性, 区分和研究R-loop的类别变得复杂且具有挑战性。尽管已有研究揭示了R-loop的多功能性, 但其形成机制和精细调控功能仍需进一步探讨。因此, 本文围绕R-loop的类别、研究方法、功能及其应用前沿展开综述, 旨在为该领域的研究提供新的见解和方向。

关键词: R-loop; RNA-DNA杂交链; DNA转录; DNA复制; 基因组不稳定性; DNA损伤修复; 疾病

Functions and application prospects of R-loops

SU Siyi^{1,2}, XU Yingying^{1,2}, ZHANG Liyuan^{1,2*}, REN Jie^{1,2*}

(¹Key Laboratory of RNA Innovation, Science and Engineering, China National Center for Bioinformation, Beijing 100101, China;

²Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 100101, China)

Abstract: R-loops, each composed of a DNA:RNA heteroduplex and a displaced single-stranded DNA, are prevalent genomic structures that play critical roles in a myriad of cellular processes, including transcription, DNA replication and epigenetic inheritance. However, these structures can also contribute to pathological states. Dysregulation of R-loops has been implicated in a range of disorders and diseases, including

收稿日期: 2024-06-28

基金项目: 中国科学院战略性先导B项目(XDB0570100); 国家重点研发计划项目(2019YFA0802202, 2020YFA0803401); 国家自然科学基金项目(32171446, 32121001); 中国科学院稳定支持基础研究领域青年团队项目(YSBR-076)

#共同第一作者: 苏思毅, E-mail: susiyi2019d@big.ac.cn; 徐莹莹, E-mail: xuyingying@big.ac.cn

*通信作者: 任捷, E-mail: renjie@big.ac.cn; 张丽媛, E-mail: zhangly@big.ac.cn

neurodegenerative disorders, cancer, and autoimmune diseases, emphasizing the need for their tight regulation. However, the precise mechanisms governing R-loop formation, stability, and resolution remain elusive, and the categorization of R-loop based on their chromatin context is yet to be fully established. Recent advancements have significantly expanded our knowledge of R-loops, revealing their diverse functions and the potential consequences of their dysregulation. In light of these advancements, this review aims to provide new insights and research directions by examining the categorization, methodologies, impacts of R-loops. By consolidating the current knowledge, this review seeks to enhance our understanding of R-loops and stimulate further investigations in this field.

Key Words: R-loop; RNA-DNA hybrid; DNA transcription; DNA replication; genome instability; DNA damage repair; disease

近年来,随着基因组和表观组测序技术的发展,我们对于大部分DNA在转录过程中形成相对稳定的转录本RNA有了更深入的认识。其中,RNA与其互补的DNA链通过序列特异性退火结合,形成RNA-DNA杂交链,进而以一种新颖的方式发挥功能。具体而言,当这种杂交链同时伴随产生一条DNA单链时,该三链结构被称为R-loop。R-loop这一术语源于D-loop的概念,后者描述了一条DNA链与双链DNA中的一条竞争性结合的情形^[1]。在广义上,即使只有RNA-DNA杂交链而没有形成DNA单链,这种情况也包括在R-loop的研究范围内。R-loop广泛存在于原核和真核生物的基因组中,其状态对调控细胞的多种生物学过程至关重要,如基因转录、DNA复制、表观遗传修饰以及DNA损伤修复等。异常的R-loop存在可能导致基因组不稳定和表观遗传记忆改变,从而引发肿瘤、神经系统疾病和炎症等多种疾病。

1 R-loop研究前沿

R-loop的生物学研究涵盖了从分子机制到疾病的广泛领域。研究前沿包括深入探究R-loop形成的分子机制,及其在基因转录和表观遗传调控中的复杂作用。此外,在肿瘤、神经退行性疾病等病理变化中异常的R-loop与病理机制之间的关联也受到关注。随着高通量测序和生物化学技术的进步,更加精准、高效的R-loop检测方法不断衍生,为深入理解其功能提供了重要工具。

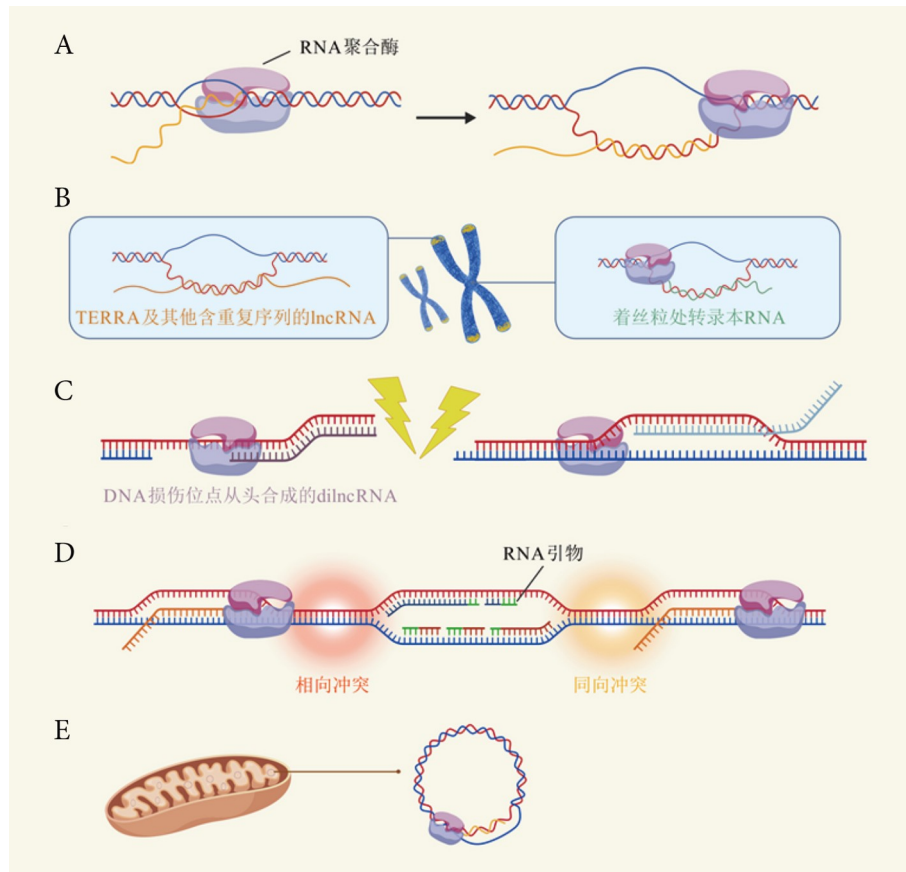
1.1 R-loop的形成与分布

目前,R-loop已成为研究的新热点,其功能受

多种因素影响,包括所处位置和染色质环境,但目前对R-loop的分类尚不完全清晰。根据RNA的来源和形成方式,本文将体内R-loop分为五种主要类型:转录伴生R-loop、重复序列R-loop、损伤相关R-loop、复制相关R-loop及线粒体R-loop(图1)。在这些分类中,转录生成的RNA通常是各类R-loop的主要来源,因此转录伴生R-loop与其他类型R-loop存在一定程度重叠,但由于它们分布于特殊位点并具有特殊特征,因此进行了单独区分。

1.1.1 转录伴生R-loop

在基因表达过程中,RNA聚合酶(RNA polymerase, RNAP)持续地沿着模板DNA链进行快速移动,伴随着RNA的不断合成,新生的RNA会在RNAP内部与其模板DNA进行短暂的互补配对以保证合成的准确性,形成一种瞬时性的R-loop结构^[2]。多种原因导致的RNAP暂停或速度失常也可能形成R-loop(图1A)。在基因内部,RNAP的转录水平会显著影响伴随转录的R-loop形成水平。转录水平的提高会大幅度提高转录产物的生成,提高的转录产物增加了RNA与DNA杂交的可能性。基因间区的RNAP转录也会形成R-loop,并影响基因内部的RNAP运行,进一步产生更多的R-loop。例如,RNAP I在核糖体DNA(ribosomal DNA, rDNA)上的高水平转录积累了大量R-loop,RNAP II在rDNA基因之间的间区转录会形成限制RNAP I转录的R-loop^[3]。此外,富含GC碱基的核酸序列更易发生碱基互补配对,表现为转录伴生R-loop热点区域具有GC偏好性。在控制抗体类别转换的基因中形成的R-loop分布于富含GC且高度重复的序



A: 伴随转录产生的R-loop; B: 端粒和着丝粒区域含高度重复序列的RNA形成R-loop; C: DNA损伤位点处及临近转录位点处形成R-loop; D: 复制-转录冲突处形成R-loop; E: 线粒体DNA转录产生R-loop

图1 五种R-loop类别

列中^[4]。

1.1.2 重复序列R-loop

端粒DNA是真核生物染色体末端的DNA序列, 其在哺乳动物中高度保守且具有高度重复的特点。RNAP II在端粒区域转录产生的长链非编码RNA被称为含有端粒重复的RNA(telomeric repeat-containing RNA, TERRA), 具有保护染色体末端不受破坏的功能。人类细胞中TERRA富含UUAGG重复序列, 具备能够与DNA链互补配对形成杂交链的优势。研究表明, 由重复序列组成的TERRA能够以非转录伴生的形式产生端粒R-loop, 该结构暴露的单链DNA增加了端粒受损伤的风险^[5]。在同样含有高度重复序列的着丝粒DNA上, 由转录形成的RNA产物也能形成R-loop(图1B)。在有丝分裂的染色体上形成的着丝粒R-loop能够招募ATR到着丝粒区域, 进而通过CHK1促进Aurora B激活, 确保染色体的准确分离和有丝分裂的正常进行^[6]。尽

管着丝粒R-loop在着丝粒功能和组装中起着重要作用, 但它可能会增加复制压力。在细胞DNA复制阶段, CENPA抑制着丝粒的转录和R-loop的形成, 从而保护着丝粒免受DNA损伤的影响^[7]。因此, R-loop对含有重复序列特征的端粒及着丝粒区域具有双重影响, 其受到细胞周期、细胞代谢状态和细胞寿命等多种因素的精密调控。此外, 其他DNA序列在基因组中的重复使得这些RNA能够与多个位点重新退火, 从而以异位形式形成R-loop。除了TERRA, 已有报道表明, 其他长链非编码RNA也可以通过形成R-loop发挥作用, 如lncRNA HOTTIP和APOLO^[8,9]。

1.1.3 损伤相关R-loop

基因突变、细胞死亡导致衰老并引发疾病。在人类细胞的DNA损伤修复过程中, 已经确定DNA双链断裂位点(double-strand breaks, DSBs)处存在RNA-DNA杂交链。杂交链中的RNA可能源自

RNAP在损伤位点处的单链DNA(ssDNA)上进行从头合成或临近损伤位点处RNAP的转录产物^[10,11](图1C)。DSBs上从头合成的RNA被称为DNA损伤诱导的lncRNAs(damage-induced long non-coding RNAs, dilncRNAs),其已经被证实,能够形成RNA-DNA杂交链,从而招募损伤修复复合物促进DNA恢复^[12]。位于损伤位点附近的RNAP转录活性也对DSBs的修复发挥作用,但这种周边性的RNA/R-loop如何响应与生成还不清楚。有证据表明,DNA损伤位点导致的RNAP暂停可能会形成R-loop,也可能是已经存在的R-loop中游离ssDNA作为模板进行从头合成而响应^[13]。总之,已有证据显示DNA损伤位点处RNA-DNA杂交链的存在,但仍不清楚它的形成可能受到哪些因素影响,目前认为可能的影响因素包括DNA损伤的类型、发生位点和时间等。

1.1.4 复制相关R-loop

DNA复制与转录是所有细胞中最重要的两个基本生命过程,但无论如何两种机制都无法避免发生于同一DNA分子上,此时就会导致复制-转录冲突的发生。DNA复制是从复制起点开始由DNA聚合酶沿着DNA链双向进行,转录则是由RNAP沿着模板DNA从基因上游向下游移动,因此产生了两种可能的相遇结果,即相向冲突和同向冲突(图1D)。同向冲突时,伴随RNAP转录产生的R-loop可以被前导链上的复制解旋酶处理,R-loop中的RNA最终会被移除或作为引物被复制体使用,从而有效解决冲突并顺利进行复制^[14,15]。相比较而言,发生在后随链上的相向冲突已经被证实具有更加严重的后果,会导致复制体停滞更久从而需要重启复制^[16,17]。相向冲突中的R-loop可能预先已经在RNAP中形成,但有证据表明相向冲突进一步增加了R-loop的形成和积累。出现在冲突发生区域的R-loop可能会导致冲突事件更加致命,一方面表现在R-loop可能会阻挡复制机器的继续进行,另一方面表现在R-loop可以引发更剧烈的DNA损伤^[15-18]。有研究表明,R-loop自身或许并不能完全阻断复制的进行,而R-loop与产生R-loop的RNAP组合在一起,或R-loop与其他一系列杂交链结合蛋白的组合,成为了复制不可逾越的障碍^[19]。一项通过电镜观察细胞中诱导产生的相向冲突研究表明,在

复制叉后方而不是冲突区域也发现了RNA-DNA杂交链的存在。杂交链中的RNA可能来源于复制叉后方DNA的转录产物,也可能是复制机制在解决掉相向的RNAP后残余的RNA重新退火至复制叉后方的ssDNA;但无论RNA来自哪里,如果这些RNA-DNA杂交链持续积累,复制体则无法顺利通过冲突区域^[20]。无论出现在前导链还是后随链,冲突区域上的RNA-DNA杂交链可以产生多种负面影响。需要更全面的工作来解释这些负面结果是如何产生的以及生命体如何有效解决和防范这些负面结果。

除了在复制-转录冲突区域发现的RNA-DNA杂交链外,与复制相关的杂交链也出现在正常的复制进程中。DNA Pol α 起始复合物在后随链上合成8~10个核苷酸长度的RNA引物用来起始DNA合成,这些短杂交链将在冈崎片段成熟过程中被替换^[21]。短RNA-DNA杂交链中的RNA也可能来源于复制过程随机掺入的核糖核酸,这种错误掺入通常会被细胞识别为基因组损伤,随后通过核糖核酸切除修复途径去除^[22]。

1.1.5 线粒体R-loop

真核细胞含有位于细胞核的核DNA与位于线粒体的线粒体DNA(mitochondria DNA, mtDNA)两套遗传物质。位于mtDNA的两条闭环DNA链呈不对称状态,其中含有28个基因编码框且富含G碱基的链被称为重链,另一条则被称为轻链,重链和轻链分别具有各自的复制起点H(origin-H, ori-H)和ori-L^[23]。研究表明,在mtDNA的非编码区域存在R-loop,其中RNA组分定位于ori-H区域,能够引导启动重链的合成,随后轻链一起配合完成mtDNA的复制(图1E)^[24]。此外,细胞中特异性消化RNA-DNA杂交链中RNA组分的RNase H1具有细胞核定位和线粒体定位两种形式,是mtDNA维持正常复制与转录表达的必需辅助因子^[25],表明mtDNA中R-loop受细胞严格调控且具有重要生理功能。

1.2 R-loop检测技术

最早检测R-loop的方法是通过电子显微镜观察体外分离的含R-loop的核酸样品^[1]。这种直接的方法至今仍被用于研究中,如通过电子显微镜观察到复制-转录冲突的复制叉后方存在杂交链^[20]。尽

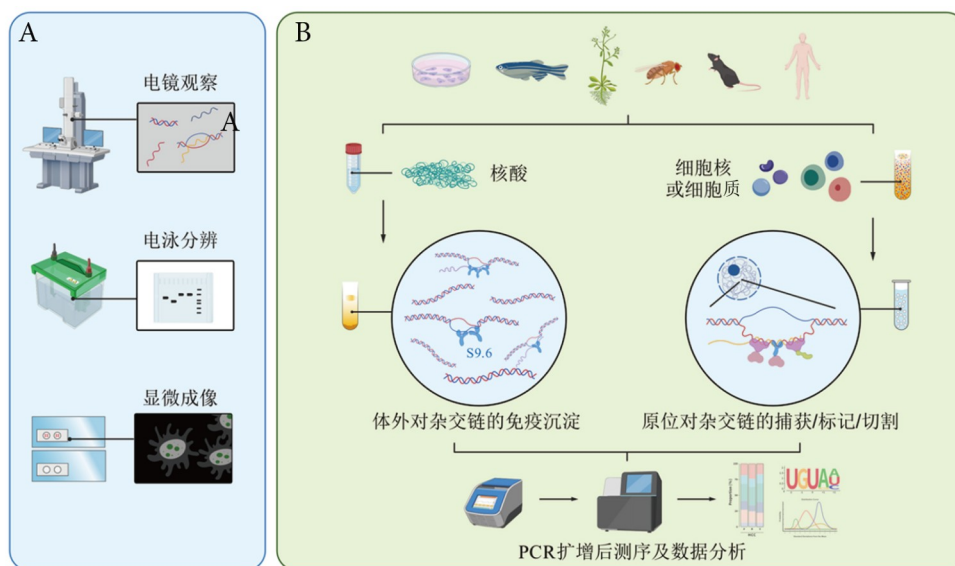
管已有研究团队详细介绍了如何进行这种检测及其统计学分析,但电子显微镜的使用和普及程度仍然存在潜在问题^[26]。也有利用重亚硫酸盐处理方法将R-loop中的ssDNA序列进行靶向,通过特异性引物扩增测序实现检测某一特定位置R-loop信息的目的^[4]。另一种常见的检测技术是利用凝胶电泳技术,借助R-loop结构在胶孔中迁移速率低于双链结构,分析核酸条带的迁移位置来评估R-loop含量(图2A)。

由于S9.6单克隆抗体的问世,R-loop检测技术得到了显著推动^[27]。S9.6抗体利用其对RNA-DNA杂交链的特异性识别与结合能力,已广泛应用于免疫荧光、免疫组化和斑点杂交等技术,适用于细胞系和组织样品(图2A)。然而,一些研究表明,S9.6抗体对双链RNA也有结合能力,这引发了对其特异性的质疑。通过分析S9.6抗体与RNA-DNA杂交链结合的晶体结构,证实了其对杂交链的高度特异性^[28]。优化样品固定方法也可以有效减少S9.6抗体在免疫荧光中对双链RNA的结合活性^[29]。此外,通过RNase H1处理可以减弱或消除杂交链信号,这是验证基于S9.6抗体检测方法特异性的一种常用策略^[30]。

随着高通量测序技术的兴起,利用S9.6抗体识别和结合RNA-DNA杂交链,并通过免疫沉淀富集

和测序捕获的核酸,已成为全基因组检测R-loop的主要方法^[31]。基于这一技术衍生出多种DNA-RNA免疫沉淀测序(DNA-RNA immunoprecipitation sequencing, DRIP-seq)方案:S9.6-ChIP-seq使用甲醛交联固定基因组DNA后超声打断,随后通过免疫沉淀来富集和测序DNA片段;bisDRIP-seq通过亚硫酸盐转化胞嘧啶后进行测序;DRIPc-seq是对S9.6沉淀RNA进行建库测序;S1-DRIP-seq通过S1核酸内切酶消化单链DNA后进行测序;RDIP-seq使用RNase I消化单链RNA后测序;ssDRIP-seq则利用限制性内切酶碎片化DNA后进行链特异性建库^[32-36]。此外,通过荧光定量PCR验证免疫沉淀前后样品中特定R-loop位点的含量,以检测和验证不同样品中的R-loop差异也是一种常用手段(图2B)。

尽管上述方法中使用RNase H1处理可以验证得到的信号的真实性,但由于这些步骤均涉及基因组核酸的提取和片段化,无法反映R-loop在细胞内天然的状态。因此,为了在保持细胞结构完整性的同时捕获R-loop,发展了几种新的检测方法。一种方法利用通过体内表达带有标签的酶活性突变型RNase H1来识别R-loop,并使用免疫沉淀和测序技术(R-ChIP)富集目标DNA片段。另外一种方法是将RNase H1与微球菌核酸内切酶MNase融合表达,用于在基因组中切割R-loop区域,随后对释放



A: 通过电子显微镜技术、凝胶电泳技术和免疫荧光等直接方法检测R-loop; B: 对各种来源的样品进行核酸提取后经杂交链的免疫沉淀进行R-loop测序分析,或者在维持细胞或细胞核结构下对杂交链进行原位的捕获/标记/切割后进行R-loop测序分析

图2 R-loop检测技术

的片段进行测序(MapR)^[37-39]。目前利用RNase H1蛋白中结合RNA-DNA杂交链的结构域HBD的方法也已出现,将串联的HBD与荧光蛋白融合表达,有效提升了免疫荧光检测R-loop的效果^[40,41]。串联的HBD与核酸切割酶Tn5的融合表达也已经被使用在基于CUT&Tag-seq的方法中,CUT&Tag技术的高效和便捷性使基于S9.6抗体的R-loop CUT&Tag-seq方法得到推广^[42,43](图2B)。

测序检测结果表明,R-loop不仅分布广泛且具有显著动态性。Crossley等^[44]利用定量DRIP-seq推测正常细胞中稳定存在的R-loop约为300个,其余的R-loop半衰期仅为约11分钟。随后,该团队采用自动化高通量电镜工作流程,进一步估算每个正常细胞平均含有614个R-loop,每Mb基因组DNA中平均含有0.19个R-loop结构^[20]。这些结果显示,R-loop的持续存在或快速消除预示着其功能上的巨大差异。近期研究集中在解析这些R-loop的生理和病理意义,以及它们的维持和机制。

1.3 R-loop的生理调控功能

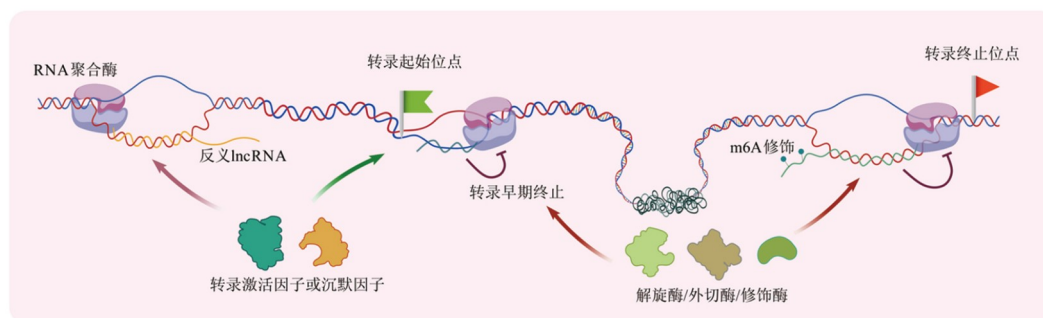
广泛分布的R-loop在细胞生理中具有重要作用,尤其是在基因表达调控方面展现出多层次、多方面的功能。通过与转录机制的直接相互作用,R-loop能够影响基因转录的起始、延伸和终止过程。此外,通过调节DNA甲基化、染色质的结构与开放状态等多种表观遗传信息,R-loop能够精确调控特异性基因表达模式。

1.3.1 R-loop调控基因表达

R-loop对基因表达的调控涉及促进或抑制基因转录以及对转录终止的调控(图3)。一些转录启动区域形成的R-loop与反义长非编码RNA(long

noncoding RNA, lncRNA)的转录合成密切相关。例如,编码中间丝波形蛋白的*VIM*基因上存在反义lncRNA的编码基因*VIM-AS1*,其转录启动导致R-loop的形成,招募核转录因子- κ B(nuclear factor kappa-B, NF- κ B)到*VIM*上促进基因转录^[45]。位于基因上游较远位置的反义lncRNA *GATA3-AS1*也能够形成R-loop,促进*GATA3*基因的转录激活^[46]。有趣的是,研究表明,在基因转录起始区域形成的R-loop可以提供单链DNA作为反义RNA转录的模板,这时R-loop本身成为“启动子”^[47]。这些证据提示,转录启动区域形成的R-loop与基因表达调控密切相关,既可以促进也可以抑制基因表达。然而,全基因组水平的功能研究以及造成R-loop调控机制差异的原因仍需进一步探索。

RNAP在基因末端的停止机制是一个复杂的过程,其正确终止对基因表达至关重要。若RNAP未能及时停止,则可能导致转录通读,进而干扰下游基因的表达甚至引发DNA损伤。全基因组研究表明,部分基因转录终止位点附近存在R-loop富集,与之对应的,呈高密度分布于染色体上的基因3'端富含R-loop,这些紧邻的基因往往受转录通读的影响更严重^[48]。终止位点附近的R-loop促进转录终止的功能已经被证实。具体来说,R-loop中RNA-DNA杂交链的形成使得新生转录本缠绕在RNAP II行进的后方,这种扭转应力使得聚合酶前进变缓^[49]。此外,终止区域的R-loop似乎能够与多种酶组分联合来调控转录停止。当细胞缺失解旋酶(如senataxin、DHX9、DDX5)或外切酶XRN2时,部分基因的转录无法正常终止,同时伴随R-loop水平的整体上升^[50-53]。然而,这些酶组分与转



转录起始位点上游反义lncRNA转录形成R-loop,招募转录激活因子或沉默因子操纵基因转录起始,转录早期伴随形成的R-loop调控转录活动早期终止,转录终止位点附近R-loop抑制转录继续进行,并通过招募解旋酶、外切酶、修饰酶促进转录终止

图3 R-loop调控基因表达

录通读是否完全依赖于R-loop尚需进一步研究。有研究显示, 解旋酶DDX21能够被R-loop招募到基因末端, 进而促进METTL3对终止位点附近RNA进行 m^6A 修饰, 随后引导XRN2参与转录终止^[54]。另外, 有研究提出, 识别 m^6A 修饰的YTHDF2蛋白对于维持R-loop稳态至关重要, m^6A 修饰对基因组稳定性的贡献部分依赖于R-loop^[55]。这表明部分R-loop所介导的转录终止功能由识别或响应其RNA修饰的蛋白质或酶组分决定, 但目前仍缺乏解释转录终止位点R-loop形成机制的研究。一项研究显示, RNAP II为转录起始后的短暂停状态形成的R-loop提供了一段游离ssDNA, SOSS-INTAC复合物结合该ssDNA并通过其核酸内切酶活性诱导转录的早期终止^[56]。这是R-loop调控终止的新示例, 即通过转录早期终止来保护基因组稳定性或调控基因表达。未来需要更多关于R-loop参与转录终止的研究, 以全面理解R-loop如何发挥该功能。

1.3.2 R-loop调控表观遗传

除了通过与转录机器和相关调控蛋白发生作

用, R-loop还可以对表观遗传的多个维度(即DNA修饰、RNA修饰、组蛋白修饰和染色质高级结构等)进行调控(图4)^[57]。DNA修饰维度的证据表明, 位于转录起始区域的R-loop能够通过保护该区域DNA免受甲基化修饰来促进基因转录。R-loop可以阻止DNA甲基转移酶3B1(DNA methyltransferase 3B1, DNMT3B1)对特定区域甲基化^[31]。此外, R-loop与G-四链体在转录起始区域重叠, 体外研究显示, R-loop先于G-四链体形成, G-四链体也能通过隔离DNMT1来保护DNA免受甲基化^[58-60]。生长阻滞与DNA损伤蛋白45A(growth arrest and DNA damage inducible alpha, GADD45A)能够与启动子区域的R-loop结合, 招募去甲基化酶TET1来介导DNA去甲基化^[61]。数千个R-loop位点与TET1结合位点重叠于启动子区域, 其中包括基因*Gadd45a*的启动子, 暗示R-loop-DNA去甲基化可能存在自我反馈调节。然而, 在转录起始区域, 5-羟甲基胞嘧啶(5-hydroxymethylcytosine, 5hmC)的低水平分布与R-loop的高水平分布不对应, 可能因为R-loop的

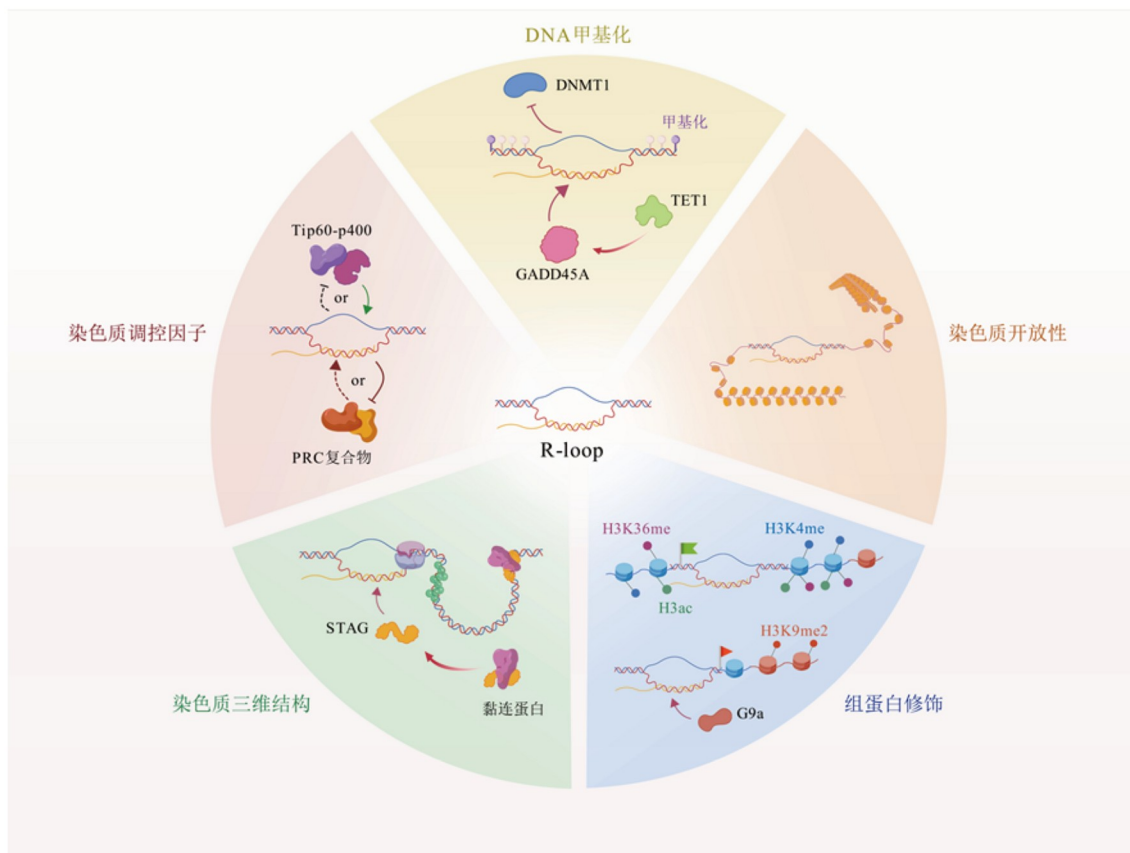


图4 R-loop调控表观遗传

形成晚于DNA 5mC向5hmC转化过程^[62]。在*DNMT3B*敲除或突变导致的功能缺失细胞中,着丝粒R-loop的稳态受损,引发基因组不稳定^[63]。这些发现强调了R-loop与DNA甲基化之间相互影响,需要进一步研究以阐明R-loop在哪些情景中起主导作用。

不仅作用于DNA修饰,R-loop对染色质状态的调控机制更为多样。例如,其独特的RNA-DNA双螺旋的A型构象直接干扰核小体在其上的组装,因而R-loop的存在本身可能就可以维持染色质的高开放性^[64,65]。全基因组R-loop测序结果显示,R-loop热点区域与染色质开放区域重合,这验证了该观点^[66]。细胞核内染色质相关RNA与开放、活跃的染色质状态相关,用核糖核酸酶或转录抑制剂处理会导致染色质结构坍塌,也支持该观点^[67]。全基因组测序分析结果显示,活性组蛋白H3第4位赖氨酸残基三甲基化修饰(H3K4me3)、H3K4me1、H3K36me3、H3乙酰化修饰与启动子处R-loop相关,而终止区域的R-loop与H3K4me1有关,提示H3K4me1可能是受R-loop影响最大的组蛋白修饰^[68]。此外,终止区域的R-loop与乙酰转移酶p300的招募增加相关,且能够招募组蛋白赖氨酸甲基转移酶G9a,其通过对组蛋白进行H3K9me2修饰促进异染色质形成^[69]。这些证据表明,转录终止区域R-loop通过将附近染色质转变为异染色质状态来发挥终止功能。

同时,R-loop也可以通过影响多种组蛋白修饰的写入、擦除和识别因子的结合来调控染色质状态。例如,R-loop既能够限制抑制性染色质修饰酶多梳抑制复合物2(polycomb repressive complex 2, PRC2)的亚基Suz12加载至启动子处,也能够招募具有激活性质的染色质重塑和乙酰化复合物Tip60-p400,从而影响染色质状态并调控干细胞分化潜能^[70]。该研究提出了一种可能的机制:当启动子处R-loop的形成还不稳定(低GC含量且基因低表达)时,激活或抑制的调控因子都可能被招募,而稳定的R-loop(高GC含量且基因高表达)则只招募能够增强或维持基因高表达的调控因子。另一项研究也支持部分富含CpG、多梳家族蛋白(polycomb-group protein, PcG)靶向的启动子形成的R-loop有助于PcG介导的基因沉默^[71]。他们提出了另一种可

能的机制:R-loop在转录活性基因中发挥激活因子的功能,而在PcG靶向的非活性基因中起到转录抑制因子的功能,R-loop中RNA的质量、所在基因的活性以及R-loop自身稳定性等多因素决定了其功能。这些假设表明,R-loop利用染色质调控因子发挥功能具有不确定性,但调控因子如何分辨不同的R-loop或者R-loop如何控制调控因子的来去仍然是未解之谜。

上述种种作用机制使得R-loop与染色质状态的关系错综复杂,同时因为组蛋白修饰改变往往伴随转录变化,其与组蛋白修饰的直接关联研究更具有挑战性。例如,H3K36me3由组蛋白甲基转移酶SETD2介导,其与RNAP的C末端结构域相互作用^[72],启动子上具有R-loop的基因比不含R-loop的基因能够富集更多的H3K36me3,并伴随RNAP的增多^[66]。这提示R-loop可能帮助RNAP在起始区域短暂停滞,从而招募修饰酶对周围组蛋白进行修饰,但也表明R-loop可能只是增强修饰强度的间接因素。尽管与RNA-DNA杂交链互作的修饰酶报道不多,但从ssDNA互作的修饰酶来看,R-loop提供的ssDNA或许能够结合具有SET结构域的H3K4甲基转移酶复合物^[72],这提供了R-loop可能直接促进修饰功能的证据,但仍需进一步验证。此外,Tip60-p400复合物的乙酰化修饰活性因RNase H1过表达显著下降,提示该复合物的某亚基可能直接互作于R-loop,这也支持R-loop直接影响修饰的观点^[70]。总之,进一步研究R-loop与染色质调控因子或组蛋白修饰因子的直接关联对于解析R-loop的调控功能至关重要。

此外,R-loop虽然是线性的核酸结构,它也对染色质三维结构具有重要的调控功能。染色质三维结构能保证遗传信息的可储存性以及基因表达的时空性。黏连蛋白复合物介导染色质环的形成,组成该复合物的STAG1和STAG2蛋白被证明能够结合R-loop结构且存在共定位现象,体内的干扰实验发现,STAG1能够结合R-loop并发挥抑制作用^[73,74]。另外的研究也证实了STAG2-黏连蛋白可能直接抑制R-loop的形成^[75]。这提示R-loop可能是STAG蛋白的聚集区域,与其他各种RNA结合蛋白一起支持黏连蛋白在染色质装载。有研究还指出,R-loop会阻挡体外由黏连蛋白介导的DNA压

缩, 具有界定染色质拓扑相关结构域边界的作用^[76]。体内由lncRNA HOTTIP形成的R-loop也被证明具有增强CTCF介导的边界作用, 进而促进染色质拓扑相关结构域的形成^[8]。此外, R-loop与G4的共存也能够促进CTCF的结合并增强染色质成环^[77]。尽管调控机制仍然不清晰, 但这些示例表明, 在研究R-loop调控功能时考虑其对染色质结构的影响同样重要, 同时也反映了R-loop在染色质动态调控中发挥多样化的重要功能。

1.4 R-loop指示并影响病理变化

R-loop的形成和消除受到多种途径和因子的调控, 包括mRNA处理、DNA超螺旋、单链DNA结合蛋白、RNA解旋酶等, 已有多篇优秀综述文章详细讨论了这些调控因子^[78-80]。当这些因子发生异常导致疾病时, 通常会伴随R-loop的显著变化, 下文将总结具有指示功能和影响病理进程的R-loop。

1.4.1 R-loop与肿瘤

癌症的发生发展通常以癌细胞的异常转录状态为标志, 活化的致癌基因表达产物和细菌性致癌物都会增加R-loop的形成, 从而增加复制压力和基因组不稳定性。Val-12突变的HRAS蛋白、肉瘤相关的融合蛋白EWS-FLI1、SS18-SSX等致癌蛋白表达的癌症中都表现出全基因组R-loop水平的增加, 随后出现依赖于R-loop的复制应激和DNA损伤^[81-83]。雌激素受体阳性乳腺癌中, R-loop高度富集于响应雌激素的基因组位点, 形成雌激素依赖性的、R-loop驱动DNA损伤。此外, 缺失BRCA1/SETX导致R-loop驱动DNA损伤, 缺失BRCA2也会导致异常的R-loop积累, 表明其他类型乳腺癌中R-loop的异常积累也是标志性事件且加剧DNA损伤^[84,85]。在多发性骨髓瘤和伯基特淋巴瘤中, B细胞中TDRD3的缺失导致c-MYC位点上R-loop积聚, 引发严重的DNA损伤和频繁的染色体易位^[86]。编码剪接因子SRSF2和U2小核RNA辅助因子1基因上的致癌突变在骨髓增生异常综合征中普遍存在, 其被证实会诱导R-loop积聚和复制应激^[87-89]。急性髓系白血病中lncRNA HOTTIP介导的R-loop形成也会驱动癌基因的表达从而加速白血病进程^[8]。幽门螺杆菌感染通过激活NF- κ B信号转导途径增加转录水平, 引发R-loop依赖性复制应激, 从而诱发胃癌发生^[90]。这些结果都强调了致癌因

素会产生异常的、高水平的R-loop, 引发DNA损伤和加剧基因组不稳定, 提示R-loop作为肿瘤发生发展标志物的可能性(图5)。

1.4.2 R-loop与神经系统疾病

人类基因组中存在大量的短串联重复序列, 其中一些与神经退行性或神经变性疾病有关, 如肌萎缩性脊髓侧索硬化症(C9ORF72和ATXN中GGGGCC重复扩展)、额颞叶痴呆(C9ORF72中GGGGCC重复扩展)、弗里德赖希共济失调(GAA或TTC重复扩展)、亨廷顿氏病(GAA或TTC重复扩展)、脆性X综合征(FMR1的CGG重复扩展)、强直性肌营养不良(CAG/CTG重复扩展)等^[91]。扩展的重复序列很容易形成异常的R-loop, 这种R-loop暴露的单链DNA由于形成二级结构而更加稳定^[92]。扩展重复序列上的R-loop会触发RNAP的转录受损和基因沉默, 缺失RNase H时还会导致重复序列继续扩展^[93,94]。因此, R-loop不仅作为神经系统疾病的表现形式存在, 而且可能是加重疾病进程的原因之一。有研究表明, 形成在三核苷酸重复序列上的R-loop与发生在该位置的碱基切除修复途径相互联系, 能够导致重复序列的缩短^[95]。这一发现提示, R-loop在重复序列上可能具有调控功能。有研究在FMR1基因CGG扩展上证实了这一点, R-loop能够促进FMR1的DNA去甲基化, 维持其活化状态, 并促进依赖错配修复途径的重复序列缩短, 恢复基因正常功能^[96]。因此, 进一步深入探索和分类扩展重复序列上位点特异性R-loop的功能, 对于揭示R-loop与神经系统疾病之间的关系至关重要。

R-loop调控因子的异常如同重复序列扩增一样, 也会导致神经系统疾病。例如, senataxin的突变会导致肌萎缩性脊髓侧索硬化症和2型动眼神经失用症共济失调, 这与R-loop的异常积累密切相关^[97,98]。因此, 神经系统中参与R-loop代谢的各种组分对于神经系统疾病的发展同样重要。

1.4.3 R-loop与炎症及免疫性疾病

Aicardi-Goutières综合征是一种由于大脑细胞中游离核酸片段的积累而引发的自身免疫性疾病, 参与核酸处理的蛋白质功能异常能够导致该病发生。这些蛋白质包括SAMHD1、RNase H2、TREX1和ADAR1等。它们的异常会增加R-loop水

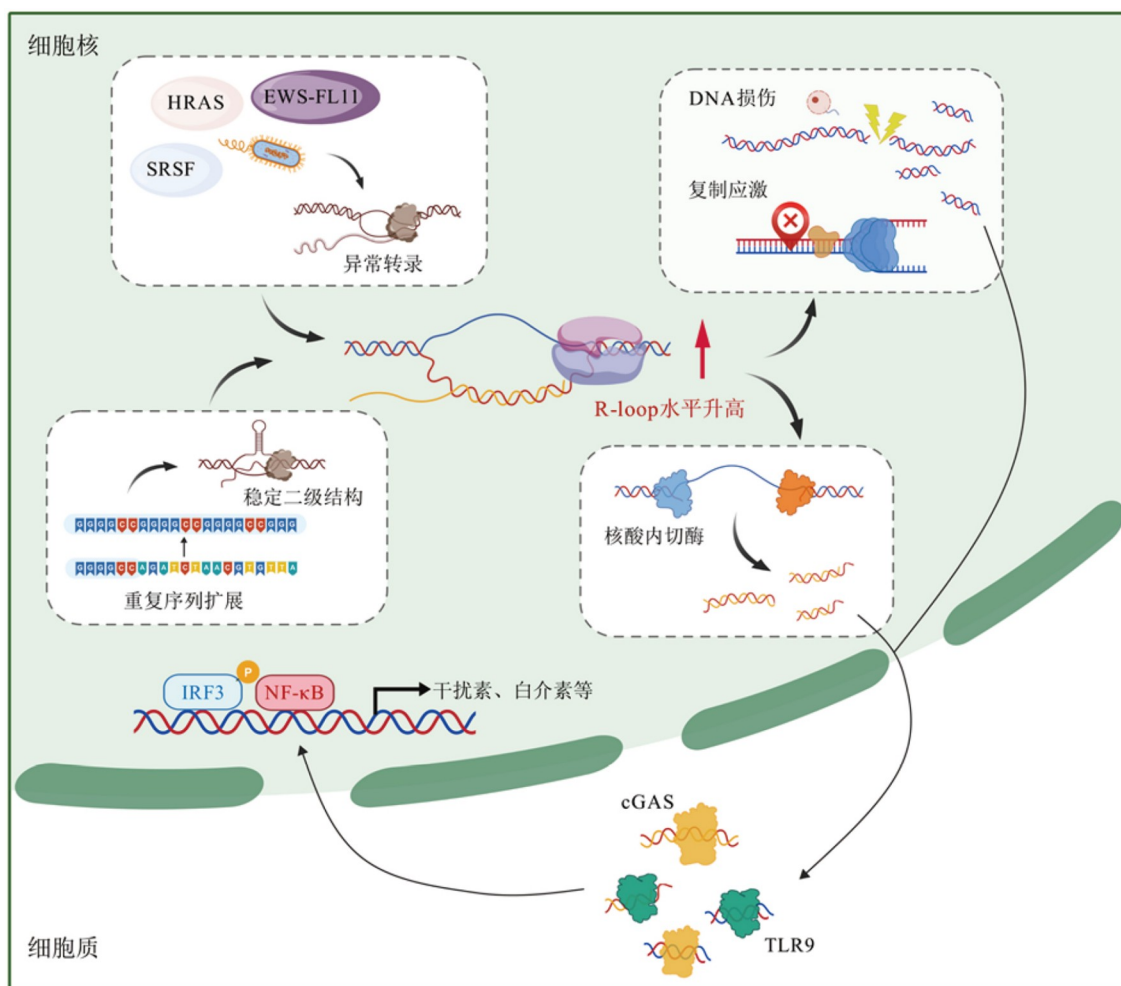


图5 R-loop指示并影响病理变化

平,与Aicardi-Goutières综合征中异常R-loop积累相一致^[99-101]。这些突变还可以导致先天免疫反应信号增加,其具体机制尚不完全清楚。研究证据表明,R-loop可能是增强先天免疫的重要原因。首先,R-loop可以引发DNA损伤和复制应激,诱导微核和游离ssDNA形成,其会被模式识别受体识别并激活免疫应答^[102,103]。其次,细胞质中高水平的RNA-DNA杂交链能够结合模式识别受体cGAS和TLR3,进而激活先天免疫反应^[104-106]。核酸内切酶XPG和XPF能够切割R-loop中的杂交链,释放到细胞质后直接激活免疫反应通路^[107]。这些发现解释了RNase H1过表达和DDX41缺失引起的炎症反应变化^[108,109]。盲鼯鼠作为寿命极长的啮齿类动物,能够利用逆转录转座子抵抗自发性和诱导性肿瘤;异常激活的逆转录转座子会导致胞质中RNA-DNA杂交链的增加,进而激活cGAS-STING-IFN通

路,诱导过度增殖细胞的死亡^[110];成人T细胞白血病中NF-κB的激活会导致XPF和XPG介导的依赖于R-loop的DNA损伤和衰老表型;幽门螺杆菌感染激活NF-κB介导的先天免疫信号会增大R-loop诱导的复制压力,提示炎症信号的增强也会进一步增加R-loop的危害性^[90-111]。R-loop与炎症反应之间的正反馈循环提示,未来操纵R-loop或外源免疫原性RNA-DNA杂交链以消除异常细胞在疾病中的应用潜力非常广泛。

2 实验室聚焦研究方向

本实验室致力于R-loop和染色质动态调控领域,围绕染色质相关RNA及其形成的R-loop结构,以R-loop的形成机制、功能表征和检测技术创新及应用为主要研究方向。通过技术改进、开发和推动分子机理的研究,以复制和转录时空互作为切

入点, 开辟了解析染色质上多维度分子事件协调机制的新思路。重要发现包括: 表观转录组与R-loop协同调控转录终止、维护基因组和异染色质结构稳定性从而延缓细胞衰老的新机制, 以及通过多种表观因子和转录因子激活成体组织干细胞来促进组织器官修复的新靶点。未来将在干细胞、衰老及再生的生理病理变化背景下, 利用新的R-loop检测技术及多维度组学信息的整合分析其基础规律和调控功能, 推动RNA介导的染色质动态调控在相关领域的应用。

3 R-loop研究的未来展望与尚需攻关的科学问题

R-loop曾经被认为是有害且罕见的转录副产物, 但随着检测方法的进步, 与R-loop相关的细胞分子机制及功能探索越来越多。R-loop的文献计量学分析表明, 自2010年以来, R-loop研究呈现爆发式增长, 研究热点逐渐从R-loop参与生物学过程的分子基础转向影响R-loop形成和稳定的潜在因素^[112]。全基因组水平检测R-loop的含量及特征揭示了其分布的广泛性, 因此了解细胞如何有效控制R-loop的出现及消除仍是一个关键问题。目前R-loop主要分为有益和有害两种类别, 即使两者的定义及分类标准还不够完善, 但该分类提示必然存在不同的分子机制调控不同种类R-loop。因此, 回答有益性或有害性R-loop如何维持或消除, 从而发挥哪些功能或造成哪些危害具有重要研究意义。

考虑到R-loop的高度动态性, 另一个研究热点集中在检测和分离体内R-loop结构的方法, 虽然1986年出现的S9.6抗体极大推动了基于二代测序的检测技术发展, 但现有检测技术难以对R-loop进行精准定位和连续性识别, 特别是针对R-loop的长度和边界定义仍然众说纷纭。此外, 不同时空状态的细胞内R-loop是否存在异质性仍有待研究。单细胞水平的R-loop检测技术将能够为这一问题的解决提供思路。

此外, R-loop和疾病之间的关联也成为研究的焦点之一, 主要研究方向集中在R-loop作为疾病的表征和R-loop加剧疾病发生发展进程两方面。有研究也提出了R-loop可能也具有提高肿瘤细胞免疫原性的新功能, 因此研究疾病发生发展中的R-loop特

征及其功能或危害性依然是该领域经久不衰的研究热点。建立新的关联知识将极大推动R-loop的深入研究。

近几年的研究为我们提供了大量的启发性内容, 对于未来详细定义有益性/有害性R-loop以及认知它们的维持/消除、功能/危害性提供了思路。R-loop检测技术的再次进步将会有效地提高功能认知, 可以预想, 未来有望融合人工智能辅助的蛋白分子设计方法, 依据需求设计出特异性更高、功能更集中的R-loop识别蛋白结构或分子。通过高特异性和高分辨率解析R-loop形成的基础规律, 将实现R-loop在更多环境下的高效率检测; 具备细胞类型和组织器官类型特异性的R-loop图谱有望被绘制出来, 辅助回答R-loop调控功能或危害性的异质性。此外, 新的R-loop与表观遗传现象之间的关联会被发现, 从而丰富R-loop功能及调控多样性。未来, R-loop有望作为一种诊断性的生物标志物用于临床分级和预后, R-loop的异质性可用于指导监测疾病的进程。以R-loop的病理学功能为指导, R-loop的调控因子及免疫原性RNA-DNA杂交链将有望成为新的疾病治疗靶点和工具。

参考文献

- [1] Thomas M, White RL, Davis RW. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc Natl Acad Sci USA*, 1976, 73(7): 2294-2298
- [2] Újvári A, Luse DS. RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat Struct Mol Biol*, 2006, 13(1): 49-54
- [3] Abraham KJ, Khosraviani N, Chan JNY, et al. Nucleolar RNA polymerase II drives ribosome biogenesis. *Nature*, 2020, 585(7824): 298-302
- [4] Yu K, Chedin F, Hsieh CL, et al. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol*, 2003, 4(5): 442-451
- [5] Arora R, Lee Y, Wischniewski H, et al. RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat Commun*, 2014, 5(1): 5220
- [6] Kabeche L, Nguyen HD, Buisson R, et al. A mitosis-specific and R loop-driven ATR pathway promotes faithful chromosome segregation. *Science*, 2018, 359(6371): 108-114

- [7] Giunta S, Hervé S, White RR, et al. CENP-A chromatin prevents replication stress at centromeres to avoid structural aneuploidy. *Proc Natl Acad Sci USA*, 2021, 118(10): e2015634118
- [8] Luo H, Zhu G, Eshelman MA, et al. HOTTIP-dependent R-loop formation regulates CTCF boundary activity and TAD integrity in leukemia. *Mol Cell*, 2022, 82(4): 833-851
- [9] Ariel F, Lucero L, Christ A, et al. R-Loop mediated trans action of the apolo long noncoding RNA. *Mol Cell*, 2020, 77(5): 1055-1065.e4
- [10] Sharma S, Anand R, Zhang X, et al. MRE11-RAD50-NBS1 complex is sufficient to promote transcription by RNA polymerase II at double-strand breaks by melting DNA ends. *Cell Rep*, 2021, 34(1): 108565
- [11] Liu S, Hua Y, Wang J, et al. RNA polymerase III is required for the repair of DNA double-strand breaks by homologous recombination. *Cell*, 2021, 184(5): 1314-1329
- [12] Pessina F, Giavazzi F, Yin Y, et al. Functional transcription promoters at DNA double-strand breaks mediate RNA-driven phase separation of damage-response factors. *Nat Cell Biol*, 2019, 21(10): 1286-1299
- [13] 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
- [14] Zardoni L, Nardini E, Brambati A, et al. Elongating RNA polymerase II and RNA: DNA hybrids hinder fork progression and gene expression at sites of head-on replication-transcription collisions. *Nucleic Acids Res*, 2021, 49(22): 12769-12784
- [15] 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
- [16] Chappidi N, Nascakova Z, Boleslavskaya B, et al. Fork cleavage-religation cycle and active transcription mediate replication restart after fork stalling at co-transcriptional R-Loops. *Mol Cell*, 2020, 77(3): 528-541
- [17] Pomerantz RT, O'Donnell M. on RNA polymerase. *Science*, 2010, 327(5965): 590-592
- [18] 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
- [19] Brüning JG, Marians KJ. Replisome bypass of transcription complexes and R-loops. *Nucleic Acids Res*, 2020, 48(18): 10353-10367
- [20] Stoy H, Zwicky K, Kuster D, et al. Direct visualization of transcription-replication conflicts reveals post-replicative DNA: RNA hybrids. *Nat Struct Mol Biol*, 2023, 30(3): 348-359
- [21] Gloor JW, Balakrishnan L, Campbell JL, et al. Biochemical analyses indicate that binding and cleavage specificities define the ordered processing of human Okazaki fragments by Dna2 and FEN1. *Nucleic Acids Res*, 2012, 40(14): 6774-6786
- [22] Reijns MAM, Rabe B, Rigby RE, et al. Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell*, 2012, 149(5): 1008-1022
- [23] Garone C, Minczuk M, Falkenberg M. Mitochondrial DNA replication in mammalian cells: Overview of the pathway. *Essays Biochem*, 2018, 62(3): 287-296
- [24] Holt IJ. The mitochondrial R-loop. *Nucleic Acids Res*, 2019, 47(11): 5480-5489
- [25] Lima WF, Murray HM, Damle SS, et al. Viable RNaseH1 knockout mice show RNaseH1 is essential for R loop processing, mitochondrial and liver function. *Nucleic Acids Res*, 2016, 44(11): 5299-5312
- [26] Stoy H, Luethi J, Roessler FK, et al. Direct R-loop visualization on genomic DNA by native automated electron microscopy. *Methods Mol Biol*, 2022, 2528: 1-20
- [27] Boguslawski SJ, Smith DE, Michalak MA, et al. Characterization of monoclonal antibody to DNA-RNA and its application to immunodetection of hybrids. *J Immunol Methods*, 1986, 89(1): 123-130
- [28] Bou-Nader C, Bothra A, Garboczi DN, et al. Structural basis of R-loop recognition by the S9.6 monoclonal antibody. *Nat Commun*, 2022, 13(1): 1641
- [29] Skourti-Stathaki K. Detection of R-Loop structures by immunofluorescence using the S9.6 monoclonal antibody. *Methods Mol Biol*, 2022, 2528: 21-29
- [30] Chédin F, Hartono SR, Sanz LA, et al. Best practices for the visualization, mapping, and manipulation of R-loops. *EMBO J*, 2021, 40(4): e106394
- [31] Ginno PA, Lott PL, Christensen HC, et al. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell*, 2012, 45(6): 814-825
- [32] El Hage A, Webb S, Kerr A, et al. Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet*, 2014, 10(10): e1004716
- [33] Dumelie JG, Jaffrey SR. Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. *Elife*, 2017, 6: e28306
- [34] Wahba L, Costantino L, Tan FJ, et al. S1-DRIP-seq

- identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev*, 2016, 30 (11): 1327-1338
- [35] Nadel J, Athanasiadou R, Lemetre C, et al. RNA: DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenet Chromatin*, 2015, 8(1): 46
- [36] Xu W, Xu H, Li K, et al. The R-loop is a common chromatin feature of the *Arabidopsis* genome. *Nat Plants*, 2017, 3(9): 704-714
- [37] Chen L, Chen JY, Zhang X, et al. R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. *Mol Cell*, 2017, 68(4): 745-757.e5
- [38] Yan Q, Shields EJ, Bonasio R, et al. Mapping native R-loops Genome-wide using a targeted nuclease approach. *Cell Rep*, 2019, 29(5): 1369-1380
- [39] Wulfridge P, Sarma K. A nuclease- and bisulfite-based strategy captures strand-specific R-loops genome-wide. *Elife*, 2021, 10: e65146
- [40] Stopar A, Nicholson AW. Multivalent forms of the ribonuclease^{HI} hybrid binding domain are high-affinity binders of RNA-DNA hybrids. *FEBS Lett*, 2023, 597(3): 472-482
- [41] Martin RM, de Almeida MR, Gameiro E, et al. Live-cell imaging unveils distinct R-loop populations with heterogeneous dynamics. *Nucleic Acids Res*, 2023, 51(20): 11010-11023
- [42] Wang K, Wang H, Li C, et al. Genomic profiling of native R loops with a DNA-RNA hybrid recognition sensor. *Sci Adv*, 2021, 7(8): eabe3516
- [43] Jiang Y, Huang F, Chen L, et al. Genome-wide map of R-loops reveals its interplay with transcription and genome integrity during germ cell meiosis. *J Adv Res*, 2023, 51: 45-57
- [44] Crossley MP, Bocek MJ, Hamperl S, et al. qDRIP: a method to quantitatively assess RNA-DNA hybrid formation genome-wide. *Nucleic Acids Res*, 2020, 48 (14): e84
- [45] Boque-Sastre R, Soler M, Oliveira-Mateos C, et al. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc Natl Acad Sci USA*, 2015, 112(18): 5785-5790
- [46] Gibbons HR, Shaginurova G, Kim LC, et al. Divergent lncRNA GATA3-AS1 regulates GATA3 transcription in T-Helper 2 cells. *Front Immunol*, 2018, 9: 2512
- [47] Tan-Wong SM, Dhir S, Proudfoot NJ. R-loops promote antisense transcription across the mammalian genome. *Mol Cell*, 2019, 76(4): 600-616
- [48] Ginno PA, Lim YW, Lott PL, et al. GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res*, 2013, 23(10): 1590-1600
- [49] Belotserkovskii BP, Liu R, Tornaletti S, et al. Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proc Natl Acad Sci USA*, 2010, 107(29): 12816-12821
- [50] Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell*, 2011, 42(6): 794-805
- [51] Cristini A, Groh M, Kristiansen MS, et al. RNA/DNA hybrid interactome identifies DXH9 as a molecular player in transcriptional termination and R-loop-associated DNA damage. *Cell Rep*, 2018, 23(6): 1891-1905
- [52] Morales JC, Richard P, Patidar PL, et al. XRN2 links transcription termination to DNA damage and replication stress. *PLoS Genet*, 2016, 12(7): e1006107
- [53] Mersaoui SY, Yu Z, Coulombe Y, et al. Arginine methylation of the DDX5 helicase RGG/RG motif by PRMT5 regulates resolution of RNA:DNA hybrids. *EMBO J*, 2019, 38(15): e100986
- [54] Hao JD, Liu QL, Liu MX, et al. DDX21 mediates cotranscriptional RNA m6A modification to promote transcription termination and genome stability. *Mol Cell*, 2024, 84(9): 1711-1726
- [55] Abakir A, Giles TC, Cristini A, et al. N6-methyladenosine regulates the stability of RNA:DNA hybrids in human cells. *Nat Genet*, 2020, 52(1): 48-55
- [56] Xu C, Li C, Chen J, et al. R-loop-dependent promoter-proximal termination ensures genome stability. *Nature*, 2023, 621(7979): 610-619
- [57] Yan P, Liu Z, Song M, et al. Genome-wide R-loop landscapes during cell differentiation and reprogramming. *Cell Rep*, 2020, 32(1): 107870
- [58] Miglietta G, Russo M, Capranico G. G-quadruplex-R-loop interactions and the mechanism of anticancer G-quadruplex binders. *Nucleic Acids Res*, 2020, 48(21): 11942-11957
- [59] Mao SQ, Ghanbarian AT, Spiegel J, et al. DNA G-quadruplex structures mold the DNA methylome. *Nat Struct Mol Biol*, 2018, 25(10): 951-957
- [60] Lim G, Hohng S. Single-molecule fluorescence studies on cotranscriptional G-quadruplex formation coupled with R-loop formation. *Nucleic Acids Res*, 2020, 48(16): 9195-9203
- [61] Arab K, Karaulanov E, Musheev M, et al. GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nat Genet*, 2019, 51(2): 217-223
- [62] Sabino JC, de Almeida MR, Abreu PL, et al. Epigenetic

- reprogramming by TET enzymes impacts co-transcriptional R-loops. *Elife*, 2022, 11: e69476
- [63] Shih HT, Chen WY, Wang HY, et al. DNMT3b protects centromere integrity by restricting R-loop-mediated DNA damage. *Cell Death Dis*, 2022, 13(6): 546
- [64] Noy A, Pérez A, Márquez M, et al. Structure, recognition properties, and flexibility of the DNA-RNA hybrid. *J Am Chem Soc*, 2005, 127(13): 4910-4920
- [65] Dunn K, Griffith JD. The presence of RNA in a double helix inhibits its interaction with histone protein. *Nucleic Acids Res*, 1980, 8(3): 555-566
- [66] Sanz LA, Hartono SR, Lim YW, et al. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol Cell*, 2016, 63(1): 167-178
- [67] Caudron-Herger M, Rippe K. Nuclear architecture by RNA. *Curr Opin Genet Dev*, 2012, 22(2): 179-187
- [68] Chédin F. Nascent connections: R-loops and chromatin patterning. *Trends Genet*, 2016, 32(12): 828-838
- [69] Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*, 2014, 516(7531): 436-439
- [70] Chen PB, Chen HV, Acharya D, et al. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat Struct Mol Biol*, 2015, 22(12): 999-1007
- [71] Skourti-Stathaki K, Torlai Triglia E, Warburton M, et al. R-Loops enhance polycomb repression at a subset of developmental regulator genes. *Mol Cell*, 2019, 73(5): 930-945
- [72] Kizer KO, Phatnani HP, Shibata Y, et al. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol Cell Biol*, 2005, 25(8): 3305-3316
- [73] Porter H, Li Y, Neguembor MV, et al. Cohesin-independent STAG proteins interact with RNA and R-loops and promote complex loading. *Elife*, 2023, 12: e79386
- [74] Pan H, Jin M, Ghadiyaram A, et al. Cohesin SA1 and SA2 are RNA binding proteins that localize to RNA containing regions on DNA. *Nucleic Acids Res*, 2020, 48(10): 5639-5655
- [75] Cheng N, Li G, Kanchwala M, et al. STAG2 promotes the myelination transcriptional program in oligodendrocytes. *Elife*, 2022, 11: e77848
- [76] Zhang H, Shi Z, Banigan EJ, et al. CTCF and R-loops are boundaries of cohesin-mediated DNA looping. *Mol Cell*, 2023, 83(16): 2856-2871
- [77] Wulfridge P, Yan Q, Rell N, et al. G-quadruplexes associated with R-loops promote CTCF binding. *Mol Cell*, 2023, 83(17): 3064-3079
- [78] Yang S, Winstone L, Mondal S, et al. Helicases in R-loop formation and resolution. *J Biol Chem*, 2023, 299(11): 105307
- [79] Petermann E, Lan L, Zou L. Sources, resolution and physiological relevance of R-loops and RNA-DNA hybrids. *Nat Rev Mol Cell Biol*, 2022, 23(8): 521-540
- [80] Brickner JR, Garzon JL, Cimprich KA. Walking a tightrope: the complex balancing act of R-loops in genome stability. *Mol Cell*, 2022, 82(12): 2267-2297
- [81] Kotsantis P, Silva LM, Irmscher S, et al. Increased global transcription activity as a mechanism of replication stress in cancer. *Nat Commun*, 2016, 7(1): 13087
- [82] Gorthi A, Romero JC, Loranc E, et al. EWS-FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma. *Nature*, 2018, 555(7696): 387-391
- [83] Jones SE, Fleuren EDG, Frankum J, et al. ATR is a therapeutic target in synovial sarcoma. *Cancer Res*, 2017, 77(24): 7014-7026
- [84] Hatchi E, Skourti-Stathaki K, Ventz S, et al. BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol Cell*, 2015, 57(4): 636-647
- [85] Bhatia V, Barroso SI, García-Rubio ML, et al. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature*, 2014, 511(7509): 362-365
- [86] Yang Y, McBride KM, Hensley S, et al. Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol Cell*, 2014, 53(3): 484-497
- [87] Chen L, Chen JY, Huang YJ, et al. The augmented R-loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations. *Mol Cell*, 2018, 69(3): 412-425
- [88] Nguyen HD, Yadav T, Giri S, et al. Functions of replication protein A as a sensor of R loops and a regulator of RNaseH1. *Mol Cell*, 2017, 65(5): 832-847
- [89] Nguyen HD, Leong WY, Li W, et al. Spliceosome mutations induce R-loop-associated sensitivity to ATR inhibition in myelodysplastic syndromes. *Cancer Res*, 2018, 78(18): 5363-5374
- [90] Bauer M, Nascakova Z, Mihai AI, et al. The ALPK1/TIFA/NF- κ B axis links a bacterial carcinogen to R-loop-induced replication stress. *Nat Commun*, 2020, 11(1): 5117
- [91] Malik I, Kelley CP, Wang ET, et al. Molecular mechanisms underlying nucleotide repeat expansion

- disorders. *Nat Rev Mol Cell Biol*, 2021, 22(9): 589-607
- [92] Loomis EW, Sanz LA, Chédin F, et al. Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. *PLoS Genet*, 2014, 10(4): e1004294
- [93] Groh M, Lufino MMP, Wade-Martins R, et al. R-loops associated with triplet repeat expansions promote gene silencing in friedreich ataxia and fragile X syndrome. *PLoS Genet*, 2014, 10(5): e1004318
- [94] Haeusler AR, Donnelly CJ, Periz G, et al. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, 2014, 507(7491): 195-200
- [95] Laverde EE, Lai Y, Leng F, et al. R-loops promote trinucleotide repeat deletion through DNA base excision repair enzymatic activities. *J Biol Chem*, 2020, 295(40): 13902-13913
- [96] Lee HG, Imaichi S, Kraeutler E, et al. Site-specific R-loops induce CGG repeat contraction and fragile X gene reactivation. *Cell*, 2023, 186(12): 2593-2609
- [97] Kanagaraj R, Mitter R, Kantidakis T, et al. Integrated genome and transcriptome analyses reveal the mechanism of genome instability in ataxia with oculomotor apraxia 2. *Proc Natl Acad Sci USA*, 2022, 119(4): e2114314119
- [98] Yeo AJ, Becherel OJ, Luff JE, et al. R-loops in proliferating cells but not in the brain: implications for AOA2 and other autosomal recessive ataxias. *PLoS One*, 2014, 9(3): e90219
- [99] Giordano AMS, Luciani M, Gatto F, et al. DNA damage contributes to neurotoxic inflammation in Aicardi-Goutières syndrome astrocytes. *J Exp Med*, 2022, 219(4): e20211121
- [100] Park K, Ryoo J, Jeong H, et al. Aicardi-Goutières syndrome-associated gene SAMHD1 preserves genome integrity by preventing R-loop formation at transcription-replication conflict regions. *PLoS Genet*, 2021, 17(4): e1009523
- [101] Coquel F, Silva MJ, Técher H, et al. SAMHD1 acts at stalled replication forks to prevent interferon induction. *Nature*, 2018, 557(7703): 57-61
- [102] Tigano M, Vargas DC, Tremblay-Belzile S, et al. Nuclear sensing of breaks in mitochondrial DNA enhances immune surveillance. *Nature*, 2021, 591(7850): 477-481
- [103] Mackenzie KJ, Carroll P, Martin CA, et al. cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature*, 2017, 548(7668): 461-465
- [104] Koo CX, Kobiyama K, Shen YJ, et al. RNA polymerase III regulates cytosolic RNA: DNA hybrids and intracellular microRNA expression. *J Biol Chem*, 2015, 290(12): 7463-7473
- [105] Mankan AK, Schmidt T, Chauhan D, et al. Cytosolic RNA:DNA hybrids activate the cGAS-STING axis. *EMBO J*, 2014, 33(24): 2937-2946
- [106] Rigby RE, Webb LM, Mackenzie KJ, et al. RNA: DNA hybrids are a novel molecular pattern sensed by TLR9. *EMBO J*, 2014, 33(6): 542-558
- [107] Sollier J, Stork CT, García-Rubio ML, et al. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell*, 2014, 56(6): 777-785
- [108] Weinreb JT, Ghazale N, Pradhan K, et al. Excessive R-loops trigger an inflammatory cascade leading to increased HSPC production. *Dev Cell*, 2021, 56(5): 627-640
- [109] Shen YJ, Le Bert N, Chitre AA, et al. Genome-derived cytosolic DNA mediates type I interferon-dependent rejection of B cell lymphoma cells. *Cell Rep*, 2015, 11(3): 460-473
- [110] Zhao Y, Oreskovic E, Zhang Q, et al. Transposon-triggered innate immune response confers cancer resistance to the blind mole rat. *Nat Immunol*, 2021, 22(10): 1219-1230
- [111] He Y, Pasupala N, Zhi H, et al. NF- κ B-induced R-loop accumulation and DNA damage select for nucleotide excision repair deficiencies in adult T cell leukemia. *Proc Natl Acad Sci USA*, 2021, 118(10): e2005568118
- [112] Li R, Liu B, Yuan X, et al. A bibliometric analysis of research on R-loop: landscapes, highlights and trending topics. *DNA Repair*, 2023, 127: 103502