



三链染色质结构R-loop的研究进展: 从检测方法到生物学功能

周劲聪^{1,2†}, 王文杰^{1,2†}, 孙前文^{1,2*}

1. 清华大学生命科学学院植物生物学中心, 北京 100084;

2. 清华大学-北京大学生命科学联合中心, 北京 100084

† 同等贡献

* 联系人, E-mail: sunqianwen@tsinghua.edu.cn

收稿日期: 2022-05-30; 接受日期: 2022-06-28; 网络版发表日期: 2022-09-08

国家自然科学基金(批准号: 32170321, 31822028, 91940306)资助

摘要 R-loop是由一条单链DNA和另一条RNA:DNA杂合链组成的三链染色质结构,在众多生物体中发挥着重要的生物学功能. 生物体内有众多调控因子通过调节R-loop的稳态来影响各水平的基因组调控事件,如转录、复制、DNA损伤及修复等. 近年来,随着R-loop的检测技术日益成熟,越来越多的R-loop生物学新功能正在被发现. 本文归纳比较了不同的R-loop检测方法,并对植物R-loop生物学功能研究进展进行了重点阐述.

关键词 RNA:DNA杂合链, R-loops, 基因组调控

R-loop是由一条单链DNA和另一条RNA:DNA杂合链混合组成的三链染色质结构, 目前已知在细菌、酵母及其他高等真核生物的基因组中均有大量分布,并在不同的生理过程中发挥非常重要的功能^[1,2]. 一般认为,新生RNA链可以与其模板DNA进行碱基配对形成RNA:DNA杂合链(RNA:DNA hybrids, RDH),同时非模板链成为单链DNA(single-stranded DNA, ssDNA). 由于这种三链结构就像一个含有RNA链的D-loop结构,因此早期被Thomas等人^[3]命名为R-loop. 一般来讲, RNA:DNA杂合链既可以在转录过程中瞬时产生(约8 bp)^[4],也可以在DNA复制过程中产生(约10~12 bp)^[5],然而这两种RNA:DNA杂合链的长度均比较短,与长度较长且较稳定的R-loop结构有所区分^[6].

科学界对R-loop的认识是一个循序渐进的过程. 1976年, Thomas等人^[3]第一次在电子显微镜下观察到R-loop结构. 在此项研究中,他们发现RNA:DNA在热力学上比双链DNA更稳定. 最新的微秒级全原子分子动力学实验(nanosecond-timescale all-atom molecular dynamics)证实, RDH双链表现出介于双链RNA的A-型(A-form)和双链DNA的B-型(B-form)之间的中间态,但可能由于RNA链的刚性强于DNA链,从拉伸模量(stretch modulus)、扭转伸缩耦合比(twist-stretch coupling)等宏观弹性参量,或从倾斜角(inclination)、H-rise和H-twist等螺旋结构参数方面比较, RNA:DNA更类似于双链RNA形成的小沟宽而浅、大沟深而窄的A-型结构^[7](图1),在溶液中也与双链RNA相似,不易

引用格式: 周劲聪, 王文杰, 孙前文. 三链染色质结构R-loop的研究进展: 从检测方法到生物学功能. 中国科学: 生命科学, 2023, 53: 289–303
Zhou J C, Wang W J, Sun Q W. Recent progress on R-loop biology: from detection methods to biological functions (in Chinese). Sci Sin Vitae, 2023, 53: 289–303, doi: [10.1360/SSV-2021-0422](https://doi.org/10.1360/SSV-2021-0422)

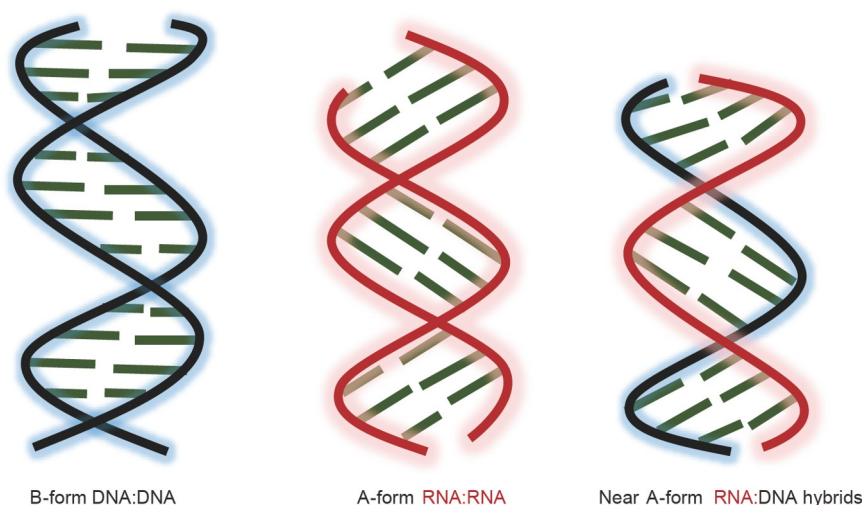


图 1 三种双链核酸结构示意图, 绘制数据来自文献[7]。从左往右依次是B-型双链DNA、A-型双链RNA、接近A-型的RNA:DNA杂合链的双螺旋结构示意图

Figure 1 Models for three kinds of double-stranded nucleic acid, modified from Ref. [7]. Double helix structures for B-form double-stranded DNA (left), A-form double-stranded RNA (middle), near A-form RNA:DNA hybrids (right)

凝聚^[8]。

近年来R-loop检测技术也从细胞水平检测发展到由非变性重亚硫酸盐测序为代表的单个位点/单基因检测, 进一步发展以S9.6抗体免疫沉淀为代表的高通量全基因组测序。不同物种全基因组R-loop分布以及动态变化图谱的绘制将为科学家们系统探究R-loop的生物学功能提供重要指导。有关R-loop在非植物系统中的功能总结在多篇综述文章中都有介绍^[1,2], 本文不再赘述。植物R-loop的生物学功能研究起始于发现R-loop可影响拟南芥开花抑制基因*FLC*位点的反义长链非编码RNA(long noncoding RNAs, lncRNAs)*COOL-AIR*转录^[9], 随后植物细胞核以及半自主细胞器基因组中R-loop的一系列生物学功能被逐一解析^[10~18]。本文将主要总结并比较目前已开发的R-loop检测技术, 同时针对植物R-loop的生物学功能研究进展进行着重综述。

1 R-loop检测技术的介绍

R-loop检测技术可分为两大类。一是从宏观层面的细胞水平检测: 例如在HeLa细胞系中, 将核酸酶RNH1中识别RNA:DNA的结构域HB与GFP荧光蛋白融合表达, 或将识别RNA:DNA的抗体S9.6与细胞孵育, 通过免疫荧光的方式表征细胞中R-loop的整体水

平^[19]; dot/slot-blot则需经过基因组DNA提取, 利用免疫印迹的手段检测基因组中R-loop的整体含量^[20]。然而这些宏观层面的检测方法都无法获取R-loop在基因组中所处的具体位置和序列信息。二是通过基因组单位点或全基因组测序方式检测R-loop, 从不同层次检测R-loop在基因组中的位置、序列等具体信息。

1.1 基因组单位点R-loop检测方法

目前已经开发出多种方案检测体内R-loop, 按检测规模可分为单个位点检测和全基因组高通量检测。其中, 体内R-loop的单位点检测主要借助重亚硫酸盐(sodium bisulfite)或限制性内切酶处理基因组, 再结合PCR技术, 可特异表征某一位点是否具有R-loop结构。例如, R-loop footprinting方法是在非变性条件下用重亚硫酸盐处理基因组DNA, R-loop结构中的单链DNA上的胞嘧啶(cytosine, C)被转化为尿嘧啶(uracil, U), 而RNA:DNA杂交链中以及双链DNA中的C受到保护而不被重亚硫酸盐转化。随后通过设计链特异性引物进行PCR扩增, 根据单克隆测序结果中由C转换为胸腺嘧啶(thymine, T)的碱基位置判定单链DNA的位置, 进而推测R-loop结构的大概边界(图2A)^[21]。此外, 通过诱导表达人源脱氨酶AID(activation-induced cytidine deaminase), 可使R-loop中单链DNA上的WRC/GYW(W代表腺嘌呤A和T; R代表A和鸟嘌呤G; Y代表C和

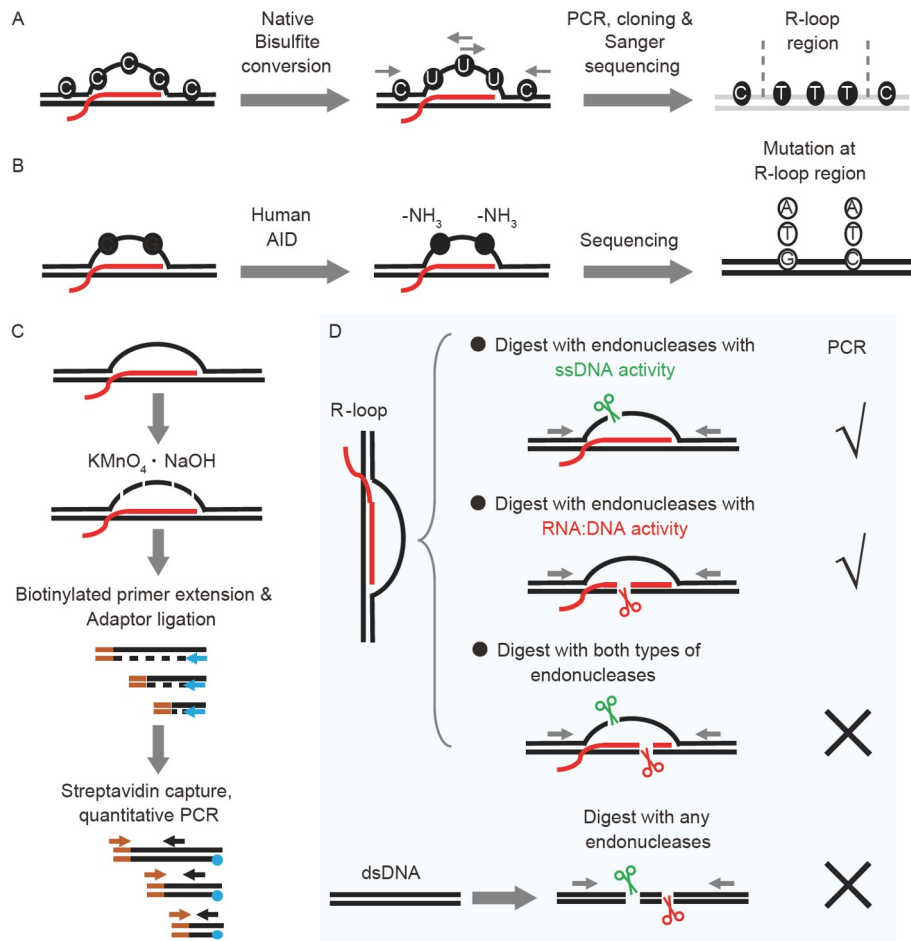


图2 单位点检测R-loop方法. A: R-loop footprinting检测方法; B: AID转换检测法; C: LM-PCR检测方法; D: DSR检测方法

Figure 2 R-loop detection methods of single locus. A: R-loop footprinting method; B: AID conversion method; C: LM-PCR method; D: DSR method

T)序列里的C和G碱基脱氨基化, 从而产生碱基突变, 因此可用碱基突变率来反映R-loop水平高低(图2B)^[22]. LM-PCR(ligation-mediated PCR)利用R-loop结构中的单链DNA易被氧化而产生断裂这一特性, 通过生物素标记的寡聚核苷酸富集断裂的单链DNA, 进一步利用PCR检测特定位点单链DNA的断裂程度, 从而间接反映此位点处单链DNA含量, 即R-loop含量(图2C)^[10,23]. 另一种鉴定单个位点R-loop水平的方法DSR(digestion-sensitive R-loop)则根据限制性内切酶对不同类型核酸的切割活性不同(例如, 除切割双链DNA外, HaeIII还对单链DNA有切割活性, AvaII或BamI还可以切割RNA:DNA杂交链中的DNA. 各种商业化限制性内切酶对不同类型核酸的切割活性可以在网站<http://rebase.neb.com/cgi-bin/hybcmbolist>中查询), 通过选取

不同类型的内切酶组合处理基因组DNA后, 结合酶切位点两侧的引物进行PCR扩增, 通过分析不同酶切组合后的PCR结果可反映该酶切位点附近的ssDNA和RNA:DNA含量^[12].

1.2 高通量测序检测技术进展

由于单位点R-loop检测方法通量低, 无法满足全基因组范围内R-loop鉴定的科研需要. 因此, 结合第二代高通量测序技术, 科学家们开发了多种多样的检测方法. 目前R-loop的全基因组高通量检测主要利用识别RNA:DNA杂交链的蛋白或抗体对RNA:DNA杂交链进行富集, 主要有以下2种策略.

(1) 通过识别RNA:DNA杂交链的蛋白进行富集(图3). RNase H1是一类保守的核糖核酸酶, 负责水解

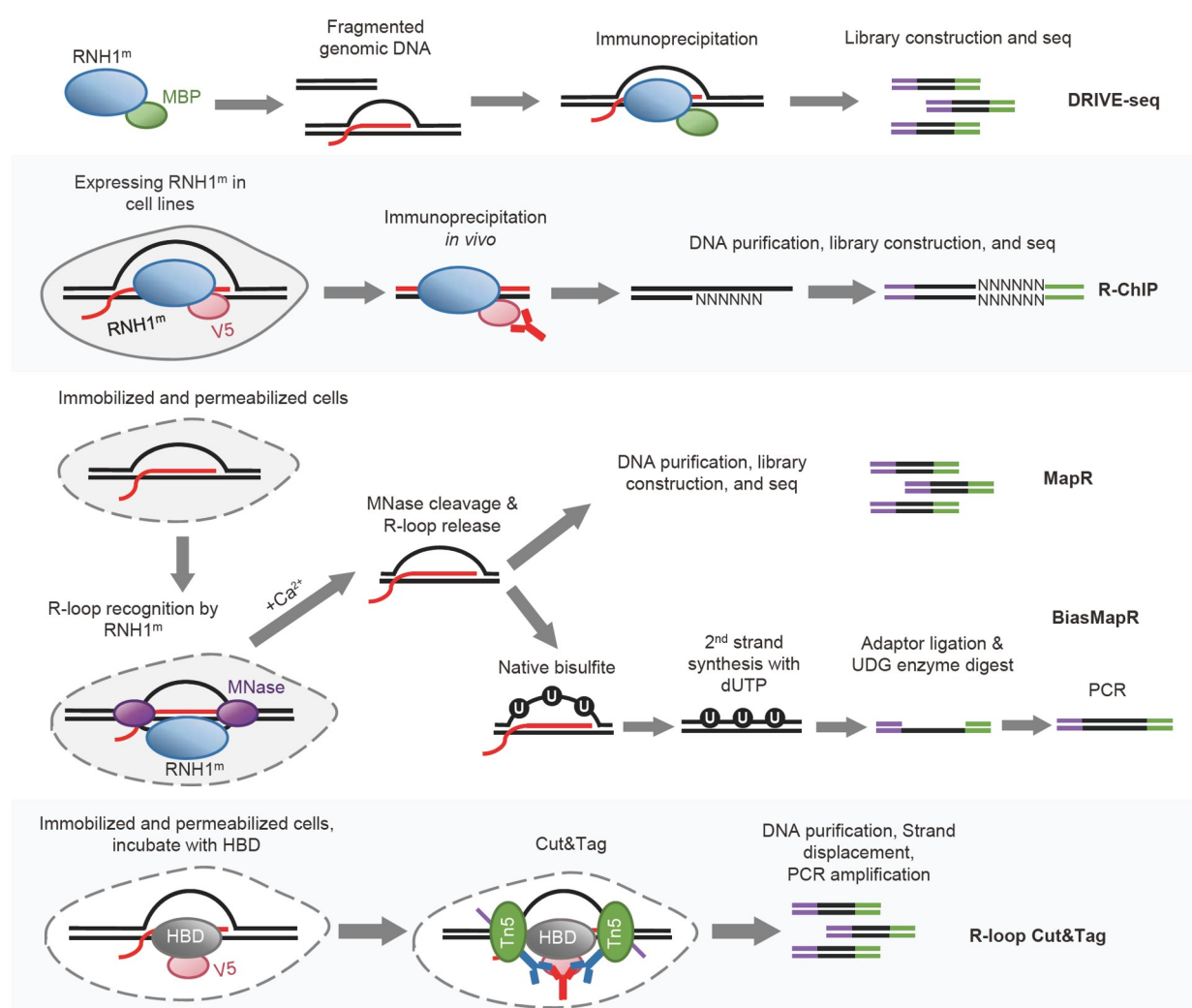


图3 基于RNA:DNA结合蛋白的R-loop高通量测序技术

Figure 3 RNA:DNA-binding-protein-based high-throughput strategies for detecting R-loops

RNA:DNA杂交链中的RNA^[24]。动物细胞的RNase H1蛋白D210N突变使得RNase H1蛋白失去水解RNA的活性,但保留了结合RNA:DNA的活性^[25]。据此,研究者们开发了利用突变型RNase H1蛋白富集RNA:DNA的高通量测序方法。DRIVE-seq使用体外表达纯化的突变型RNase H1蛋白与片段化的基因组DNA进行孵育,从中富集RNA:DNA^[26]。而R-ChIP是利用稳定表达突变型RNase H1(D210N)蛋白(带有V5标签)的HEK293T(或K562)细胞系,使用V5抗体进行免疫沉淀,进而富集RNA:DNA杂交链,然后通过DNA文库构建及测序分析鉴定哺乳动物细胞中R-loop位点^[27]。相

较于DRIVE-seq用酶切打断基因组DNA的方法^[26],R-ChIP在免疫沉淀前使用超声打断基因组,使得富集的R-loop信号具有更高的分辨率,但此方法需要构建稳定表达突变型RNase H1蛋白的实验材料,其使用范围受到了限制^[27]。

随后基于突变型RNase H1蛋白识别RNA:DNA杂交链的特点,研究者们开发出一系列优化的R-loop测序方法。MapR是将突变型RNase H1蛋白与微球菌核酸酶MNase(micrococcal nuclease)融合表达,并与固定后的细胞进行孵育,突变型RNase H1蛋白识别RNA:DNA杂交链,同时融合表达的MNase,利用其核酸切

割活性将R-loop结构从基因组中切割并解离出来,再从游离的RNA:DNA杂交链中提取DNA进行建库测序^[28].相比于前两种方法,MapR省去了免疫沉淀的步骤.相较于R-ChIP,MapR不需要稳定表达突变型RNase H1蛋白的细胞系,可广泛应用于各种遗传背景的细胞系,并且需要的细胞量较少,起始量可低至 10^5 个细胞,不足之处是获得的R-loop信号不具有链特异性信息^[28].随后,同一课题组开发了BisMapR测序方法,即在MapR基础上将解离下来的R-loop于非变性条件下用重亚硫酸盐处理,再经过RNase H消化、掺入dUTP的二链合成、末端修复、接头连接以及UDG酶切处理等步骤后构建文库,相比于MapR测序,BisMapR获得的R-loop测序结果保留了链特异性信息^[29].

随着CUT&Tag技术的广泛应用,研究者们也利用该技术开发出了R-loop CUT&Tag,用于原位R-loop检测^[30].在此方法中,RNase H1蛋白中识别RNA:DNA的结构域HBD与GST或His6标签融合表达,随后进入被concanavalin A磁珠固定的细胞中识别R-loop结构,再通过GST或His6标签抗体将转座酶招募至R-loop位点,通过施加 Mg^{2+} 离子激活转座酶活性并将文库接头插入R-loop结构区域,不再需要富集R-loop的步骤,最后直接通过PCR完成文库构建,比以上方法更加快捷.与MapR或BisMapR方法类似,R-loop CUT&Tag也不需要稳定表达突变型RNase H1蛋白的细胞系^[30].

除DRIVE-seq以外,上述方法识别R-loop的过程都是在体内发生,虽然可以在一定程度上表征细胞中全基因组R-loop的原位分布,但染色质的开放状态、体内R-loop自身结合蛋白的存在都会影响上述方法中RNA:DNA结合蛋白识别R-loop的效率,并不能完整反映R-loop在基因组中的分布.

(2) 通过识别RNA:DNA杂交链的抗体S9.6进行富集(图4).S9.6是利用ΦX174噬菌体DNA体外转录产生的RNA:DNA杂交链作为抗原免疫小鼠后,经过多轮杂交瘤细胞筛选而产生的单克隆抗体,对RNA:DNA杂合链具有高度的特异性和亲和力^[31].研究人员利用S9.6抗体结合DNA微阵列技术检测大肠杆菌、酵母等物种转录组,展现出高灵敏、无序列偏好性等特征^[32,33].因此,利用S9.6抗体,通过免疫沉淀从片段化的基因组中富集R-loop是一种常见的全基因组R-loop的检测方式,即DRIP-seq(DNA-RNA immunoprecipitation sequencing)^[26].相比于前文提及的R-loop富集方法,此类

方法最大的优势是S9.6抗体已经实现了商业化生产,可以在任何遗传材料中使用,不再需要表达纯化识别RNA:DNA的蛋白或构建稳定表达突变型RNase H的细胞株系,从而可避免细胞内可能的互作蛋白对稳定且过量表达蛋白的功能干扰.

根据S9.6富集前样品的处理方式以及富集后的建库方式不同,多种基于S9.6免疫沉淀的测序方法从DRIP-seq衍生出来(图4). (i) S9.6-ChIP-seq: 通过甲醛固定染色质后利用超声打断DNA,采用类似ChIP-seq的方式使用S9.6抗体对R-loop区域进行富集,随后通过建库测序获得基因组中的R-loop分布信息^[34]. (ii) bisDRIP-seq: 在含有重亚硫酸盐的条件下裂解细胞,接着使用S9.6抗体富集含有胞嘧啶转化为尿嘧啶的R-loop片段.此方法附加了单链中胞嘧啶被转换的信息,可以进一步精准定位R-loop区域^[35]. (iii) DRIPc-seq: 从S9.6富集的RNA:DNA杂交链中提取RNA,然后测序获得具有链特异性的R-loop分布信息^[36]. (iv) S1-DRIP-seq: 用核酸酶S1消化R-loop结构中的单链DNA,再从超声打断的基因组中用S9.6抗体富集RNA:DNA杂交链,提取DNA建库测序从而得到R-loop的分布信息^[37].此方法在超声打断基因组之前用核酸酶S1消化R-loop结构中的单链DNA,可保护大量的RNA:DNA杂交链不被超声处理破坏,进一步提高R-loop的信噪比. (v) RDIP-seq: 先用RNase I消化游离的单链RNA,再用S9.6抗体富集RNA:DNA杂交链,提取RNA建库测序,获得的R-loop分布信息背景低且具有链特异性^[38]. (vi) 在前文提到的R-loop CUT&Tag方法中,也可以用S9.6替代HBD来识别R-loop结构^[30].

ssDRIP-seq(single-strand DNA ligation-based library construction after DNA:RNA hybrid immunoprecipitation and sequencing)是一种应用较广泛的全基因组R-loop测序方法^[11].该方法将提取的基因组DNA用核酸内切酶尽量片段化(可选取多种不切割RNA:DNA杂交链的内切酶组合),随后使用S9.6抗体富集RNA:DNA,并提取RNA:DNA杂交链中的DNA进行单链DNA文库构建.在建库过程中,首先对单链DNA的3'末端进行接头连接,后续经过延伸、5'接头连接以及index PCR等步骤获得DNA文库^[11].此方法综合了上述方法的多种优点,主要包括: (i) 可能由于RNA:DNA和dsRNA在结构上有较高相似度(图1)^[7],也有研

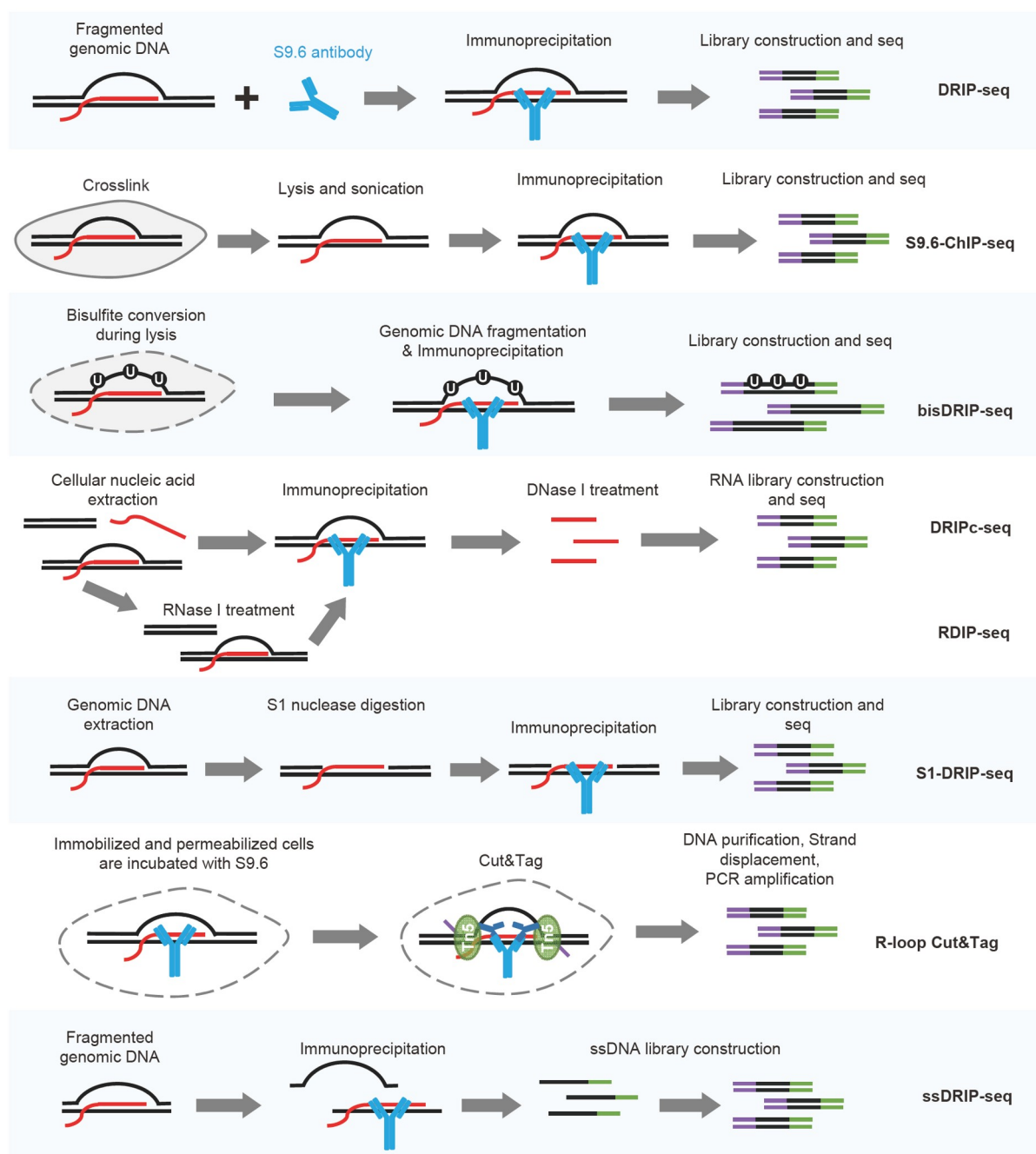


图 4 基于S9.6抗体的R-loop高通量测序技术

Figure 4 S9.6-based high-throughput strategies for detecting R-loops

究指出S9.6抗体对AU序列富集的dsRNA有微弱的结合活性^[39,40]。ssDRIP-seq利用S9.6捕获的RNA:DNA杂交链中的单链DNA进行建库,只要抗体足量,便可排除dsRNA竞争结合S9.6的影响,因此相较于提取RNA的建库方法(DRIPc-seq和RDIP-seq),ssDRIP-seq获得

的R-loop信号背景较低。(ii) 该方法对RNA:DNA中的DNA进行3'末端连接并建库,可从后续的分析中提取RNA:DNA的链特异性信息。(iii) 操作简便,不需要表达RNase H的实验材料,不受染色质凝聚状态的影响(而己知CUT&Tag方法中所用到的Tn5酶活性会受到

染色质开放程度的影响)。目前ssDRIP-seq已经成功应用在拟南芥^[11,13,14]、水稻^[41]、玉米^[42]、酵母(尚未发表)、斑马鱼^[43]、哺乳动物细胞^[44-46]等模式生物中。

1.3 全基因组R-loop测序技术展望

如前所述,目前已发表的R-loop测序技术方式众多,但进一步提高分辨率是所有测序方法需要面临的挑战。无论是超声还是酶切打断基因组,都无法避免S9.6抗体或突变型RNase H蛋白富集到的RNA:DNA片段两端边界含有非RNA:DNA片段(ssDNA或者双链DNA末端)。bisDRIP-seq附加了单链中胞嘧啶被转换的信息,但也无法在单碱基水平做到以非胞嘧啶为边界的R-loop区域确定。最近Chedin课题组^[47]利用三代测序技术结合非变性条件下重亚硫酸盐处理基因组,从单分子水平检测基因组中ssDNA的位置,并发现和已发表的DRIP-seq数据重合性较高,然而此方法同样也无法精确定位非胞嘧啶为边界的R-loop区域。如何更精准地从基因组中解离RNA:DNA杂交链,或者如何从S9.6抗体(或突变型RNase H)富集的核酸片段中进一步精确RNA:DNA杂交链的边界(直至单碱基水平边界界定),将会是进一步提高R-loop高通量测序分辨率的重要发展方向。此外,近年来单细胞测序技术的日益成熟使人们能更清晰地认识不同细胞类型中的单个细胞转录水平、表观修饰水平的动态变化^[48-50],单细胞R-loop测序作为R-loop检测技术的另一个发展方向将会帮助人们从时间和空间两个维度更加精准地理解R-loop的生物学功能。

2 植物R-loop的生物学功能和调控机制

尽管R-loop结构在20世纪70年代就被发现,但其功能及调节方式直到2013年才在植物中首次报道^[9]。近年来,随着R-loop测序技术的开发和R-loop调节因子遗传材料的逐一收集鉴定,以及成熟研究体系的逐渐建立,植物中R-loop的研究得到了较大的发展^[9-14,16-18,42,51-55]。

2.1 植物中R-loop的分布特征

利用ssDRIP-seq技术,本课题组^[11]率先揭示了拟南芥基因组上R-loop的分布情况,并挖掘出R-loop在

拟南芥全基因组范围内的分布特征,例如,(i) R-loop广泛存在于拟南芥基因组上,并在AT和GC偏离区富集;(ii) R-loop与CG序列甲基化呈负相关,然而在与转录抑制型组蛋白修饰H3K9me2和H3K27me1有一定的强相关性;(iii) R-loop在基因启动子区和基因上显著富集,与非编码RNA区和基因组重复区域高度相关。最近,通过绘制拟南芥在不同发育时期和环境刺激下R-loop动态变化的综合图谱,本课题组发现在营养生长阶段,以及多种环境因素包括光谱、光周期和温度等的刺激下,R-loop的积累模式都比较稳定,但在世代转换过程中变化较为剧烈。并且R-loop的动态变化与RNA丰度相关性不强,这表明R-loop形成和RNA积累的调控机制是独立的。此外,本课题组^[14]还注意到R-loop在转录因子结合区域有富集,暗示R-loop可能是潜在的顺式转录元件。

在其他高等植物中,Liu等人^[42]借助ssDRIP-seq对玉米进行全基因组测序,发现玉米中正义链R-loop主要富集于启动子和转录终止位点,但是在基因主体(gene body)上相对少一些,这不同于拟南芥中正义链R-loop主要位于基因主体的现象。从染色体尺度上分析,玉米的R-loop富集于近着丝粒周围的异染色质区域,其中很大一部分R-loop来源于转座元件。在着丝粒中,R-loop优先形成于着丝粒特异性组蛋白CENH3的结合区域内,且着丝粒逆转录转座子与R-loop的形成密切相关。并且,利用原子力显微镜单分子成像技术,Liu等人^[42]观察到了位于着丝粒区域逆转录转座子上的R-loop。在水稻中,Fang等人^[52]发现,R-loop的峰值与基因表达水平之间的关系取决于R-loop在基因上的位置,并且R-loop的分布位置与DNA甲基化和组蛋白甲基化修饰区域有一定的重叠。

2.2 植物细胞核基因组中R-loop的功能研究

已有的研究表明,R-loop在植物中的生物学功能是丰富多样的,在细胞核中主要发现R-loop参与调控基因的转录、DNA复制、RNA加工、基因组稳定性以及染色质三维构象(图5A)。在拟南芥中,R-loop遏制反义lncRNAs *COOLAIR*转录来影响开花抑制基因*FLC*的表达,从而参与调控开花时期^[9,56]。最近研究发现*COOLAIR*启动子区域R-loop中的单链DNA容易形成二级结构,阻止DNA复制叉前进^[57]。LncRNA *APOLO*通过短序列互补配对识别拟南芥基因组上多个远端位

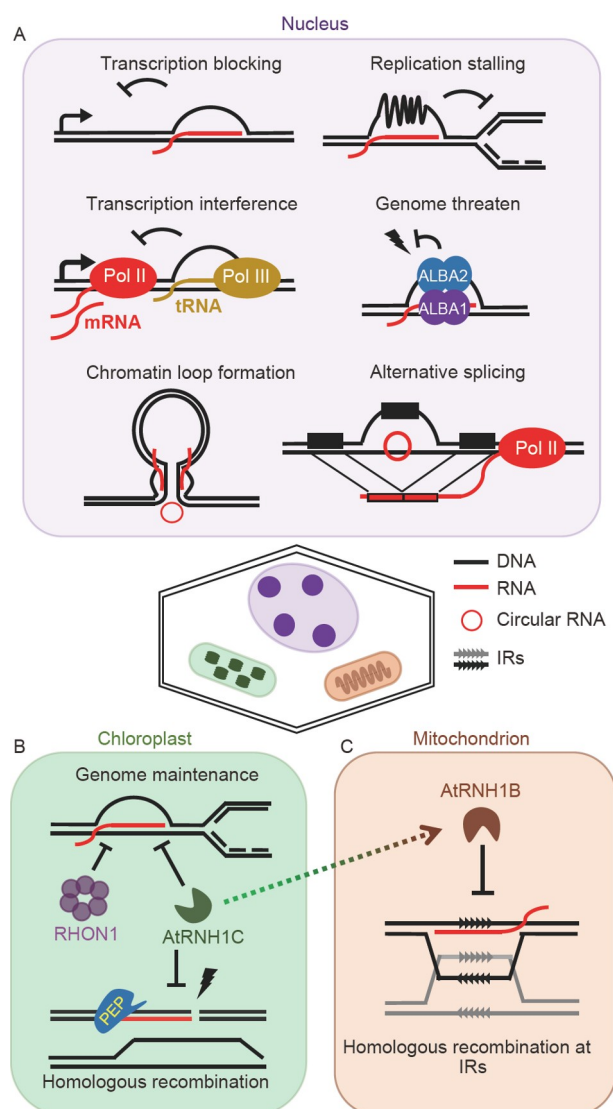


图5 R-loop在植物细胞中的功能. R-loop在植物(A)细胞核、(B)叶绿体和(C)线粒体中的生物学功能模式图

Figure 5 Overview of R-loop functions in plant cells. Working models for plant R-loops in (A) nucleus, (B) chloroplast, (C) mitochondrion

点形成R-loop, 从而诱导H3K27me3识别蛋白LHP1对局部染色质3D结构进行调节, 精细调控多个基因的表达, 影响生长素响应和侧根的发育^[54]. 最近本课题组^[16]发现, 拟南芥基因内tRNA形成的R-loop通过调控RNA聚合酶II和III之间的转录干扰来调节tRNA宿主基因的表达, 从而参与植物氧化胁迫响应. 来源于SEP3基因上第6外显子区的环形RNA(circular RNAs, circRNAs)在该基因区形成R-loop, 增加宿主基因的可变剪切效率, 造成缺少第6外显子的转录本SEP3.3的

大量积累, 从而导致拟南芥花器官同源异形^[51]. 在水稻中, 生长素相关基因位点上的R-loop可以影响基因转录, 参与生长素的极性运输和根的向地性生长^[10]. Liu等人^[55]通过对毛果杨次生木质部的高通量测序分析, 发现环状RNA在R-loop结构和选择性拼接事件区域显著富集, 并且过表达和次生长相关的环形RNA circ-IRX7会促进R-loop增加, 可能影响宿主基因转录产物的选择性剪接. 在玉米中发现, 由着丝粒反转录转座子(centromeric retrotransposons, CRMs)产生的环形RNA以R-loop结构结合在着丝粒上, 并且通过一个环形RNA形成两个R-loop促进CRM1位点处染色质环的形成^[53].

基于这一系列研究, 近期内多个植物细胞核R-loop的调节因子被鉴定到. 单链DNA结合蛋白AtNDX通过结合R-loop结构中的单链DNA维持COOLAIR启动子区R-loop的稳定^[9], 而RNA结合蛋白FCA和FY通过对RNA 3'末端加工消除COOLAIR处形成的R-loop^[56]. ALBA是进化上保守的核酸结合蛋白, 在拟南芥中AtALBA1结合RNA:DNA杂合链, AtALBA2识别R-loop中的单链DNA, AtALBA1与AtALBA2形成异源二聚体作为R-loop阅读器, 保护基因区R-loop DNA不受损伤^[13]. 水稻中DNA拓扑异构酶TOP1(DNA topoisomerase 1)通过去除RNA聚合酶在转录过程中产生的正、负DNA超螺旋防止R-loop的过量积累^[10]. RNA核酸酶RNase H作为调控R-loop水平的重要因子, 其同源蛋白在植物界中也普遍存在, 其中AtRNH1A是定位于拟南芥细胞核的RNase H1同源蛋白^[12], 其具体的生物学功能尚待揭示. 此外, 植物中还有大量含有类RNase H结构域的蛋白家族, 这些蛋白也可能是R-loop的潜在调控因子^[58].

3 内共生半自主细胞器中R-loop的功能研究进展

线粒体和叶绿体被认为是一系列内共生事件的产物^[59,60]. 目前推测, 线粒体由一个被原始真核宿主吞噬的 α -变形杆菌祖先产生, 叶绿体是后来(大约10亿~15亿年前)从一个被已经有线粒体的真核生物吞噬的蓝藻祖先中产生^[61]. 虽然线粒体和叶绿体保留了大部分原核生物的生化特性, 但在基因组层面可能只保留了各自祖先所拥有的少量基因组^[62]. 随着时间的推移,

叶绿体和线粒体基因组的大小已经发生了巨大的变化. 此外, 单个线粒体和叶绿体中的基因组拷贝数目显著提升, 在几个至几百个之间^[63,64]. 细胞器基因组上R-loop水平对维持细胞器的正常功能和生物体的健康发育也具有重要作用.

3.1 线粒体中R-loop的形成与功能

作为细胞的能量加工厂, 线粒体为细胞活动提供所需的能量, 其基因组编码参与能量转换和基因表达的关键因子. 线粒体基因组上R-loop最早于1995年由Xu和Clayton^[65]报道. 该课题组通过体外转录实验, 在酿酒酵母和人源线粒体DNA(mitochondrial DNA, mtDNA)中观察到持续、稳定形成的RNA:DNA杂合链结构^[65,66]. 后续通过二维琼脂糖凝胶电泳(two-dimensional agarose gel electrophoresis, 2D-AGE)和RNA印迹等实验证实哺乳动物线粒体R-loop的存在^[67,68]. 后续的高通量测序也表明R-loop广泛分布于酵母和人源线粒体基因组上^[34,69].

在哺乳动物线粒体中, 绝大多数的R-loop积累于调控线粒体基因组转录和复制的控制区(control region, CR), 其中的RNA链作为起始引物介导线粒体先导链的复制^[70]. 同时, 在先导链合成的过程中, 预先形成的转录本可能会与空置出的另一条单链DNA配对形成RNA:DNA杂合链结构, 促进后随链的合成^[71]. Holt^[72]还推测, 控制区的R-loop可能作为一种特殊结构使线粒体DNA锚定在线粒体内膜上(inner mitochondrial membrane, IMM). Brown等人^[68]利用原子力显微镜(atomic force microscope, AFM)发现, R-loop的存在使小鼠线粒体DNA保持更加开放的状态. 此外, R-loop还能促进mtDNA分子之间物理上的分离^[73].

不同于哺乳动物的线粒体基因组是只有15~17 kb的紧凑环形, 主要依赖于控制区的D-loop和R-loop调控复制起始^[72], 被子植物线粒体基因组要大得多, 从大约200 kb到超过10 Mb^[74], 并且结构也更加复杂, 包括环形、线性和多分枝单、双链分子等, 并主要以线性结构存在^[64]. 这些特点使植物mtDNA不太可能采用类似于动物mtDNA的复制方式. 事实上在植物和真菌中, mtDNA被认为主要通过类似噬菌体T4 DNA的重组依赖性复制^[75], 但是关于重组依赖性的复制机理仍然缺乏了解. 本课题组^[15]最近的研究发现, 拟南芥线粒体R-loop能显著促进mtDNA上中等重复区(inter-

mediate-sized repeats, IRs)的异常同源重组(ectopic homologous recombination, eHR)(图5C). 这暗示R-loop也可能参与大重复区(>1 kb, large repeats)的高频相互重组介导的基因复制.

3.2 线粒体R-loop的调控

与核基因组上R-loop水平需要严格调控一样, 细胞器中的R-loop也需要维持在稳定的状态. 人源线粒体中R-loop的异常积累会导致基因组的不稳定, 产生致病性^[69]. 而成人神经肌肉疾病(adult-onset neuromuscular disease)与mtDNA上R-loop水平过低有关^[73]. 尽管目前已经鉴定了许多调控真核生物细胞核基因组上R-loop水平的因子, 但对于线粒体中R-loop调节因子的挖掘还较少. 在哺乳动物和酵母中, 核糖核酸酶RNase H1除能定位于细胞核外, 还被转运到线粒体中, 是mtDNA上R-loop的重要调节因子. 在哺乳动物线粒体中, RNase H1被认为主要调节CR区R-loop的状态, 影响mtDNA的复制^[72]. 新生的R-loop结构不能直接启动mtDNA合成, 必须先经过RNase H1处理生成能被DNA聚合酶 γ 用于起始DNA合成的3'末端^[76], 而在复制起始后R-loop中作为引物的RNA分子需要被RNase H1消除, 保证复制能顺利完成^[70,77]. RNase H1的缺失严重影响小鼠mtDNA的复制、转录, 导致线粒体功能紊乱, 造成胚胎死亡^[70,78,79]. 在酵母中, 目前认为RNase H1在线粒体中的作用与在细胞核中的类似, 主要涉及共转录R-loop的清除, 而不是直接参与mtDNA复制过程^[34]. 线粒体降解小体(mitochondrial degradosome, mtEXO)是在人源线粒体中鉴定到的另一种R-loop去除因子, mtDNA上R-loop具有广泛的调节作用^[69]. 由于mtEXO是一种从细菌到人的功能保守的RNA监测和降解复合物^[80], 其在各物种中可能也具有R-loop调节功能. RNase H1的同源蛋白广泛存在于植物中, 并且不同于动物和真菌一般只有一个RNase H1同源蛋白, 植物中可能有多个同源蛋白负责不同基因区室的R-loop调控^[12]. 十字花科拟南芥线粒体中有3个RNase H1同源蛋白, AtRNH1A, AtRNH1B和AtRNH1C^[12]. 正常条件下AtRNH1B负责线粒体R-loop的调控, AtRNH1C负责叶绿体中R-loop的调控. 有趣的是, 在线粒体中缺乏AtRNH1B时, AtRNH1C也能被转运到线粒体中弥补AtRNH1B的功能(图5B和C)^[15].

3.3 叶绿体中R-loop的形成与功能

叶绿体是陆生植物、藻类和少数原生动物中常见的具有独立遗传物质的多功能细胞器。与线粒体或核基因组相比, 植物叶绿体基因组在结构、基因数量和基因组成方面具有更高的保守性^[61,81]。叶绿体基因组通常以环形结构表示, 包括一个小的单拷贝区域 (small-single copy region, SSC) 和一个大的单拷贝区域 (large single-copy region, LSC), 这两个区域被一对反向重复区域分开, 形成典型的四部结构, 但在细胞内其主要以线形或分枝低聚体形式存在^[82]。高等植物叶绿体基因组大小一般为120~160 kb, 包括大约120个基因, 主要编码光合作用系统的核心元件和质体基因表达相关的组分, 包括RNA聚合酶亚基、核糖体蛋白亚基、rRNA和tRNA^[83]。其中rDNA区是转录活跃区域, 而且其内部含有两个基因组复制起始位点。由于rRNA在叶绿体中高峰表达, 因此rDNA区天然具有高频的转录-复制迎面碰撞冲突 (transcription-replication head-on conflicts, HO-TRCs)。本课题组^[12]通过DRIP-qPCR在该区域内检测到大量的R-loop积累。本课题组^[12]发现, 叶绿体定位的RNase H1蛋白AtRNH1C可与DNA拓扑异构酶II同原蛋白AtGyrase协作维持rDNA区R-loop稳态并解除该区域的HO-TRCs, 从而减少DNA损伤, 维持叶绿体基因组的稳定性。除AtRNH1C外, 本课题组^[18]还发现解旋酶RHON1作为另一个已知的叶绿体R-loop消除因子, 与AtRNH1C平行调控rDNA上R-loop的水平, 维持叶绿体基因组的稳定(图5B)。

此外, 本课题组^[17]发现, 在叶绿体基因组的断裂位点处有叶绿体基因编码的RNA聚合酶(plastid-encoded RNA polymerase, PEP)转录瞬时产生的RNA:DNA杂合链, AtRNH1C识别此结构并降解其中的RNA, 随后与单链DNA结合蛋白WHY1/3和重组酶RecA1协作抑制断裂位点处微同源序列介导的断裂修复 (microhomology-mediated break-induced repair, MMBIR), 以促进更精准的同源重组 (homologous recombination, HR) 修复来维持叶绿体基因组完整性(图5B)。

4 总结与展望

虽然早在20世纪70年代, R-loop结构已经在体外

实验中被观测到, 但直到最近十年, R-loop生物学功能研究才随着S9.6抗体的普及和多种高通量测序技术的发展逐渐被更多研究者重视。虽然很长时间内R-loop被认为是基因转录的副产物, 但目前越来越多的研究发现R-loop可以作为真核生物染色质稳定组分, 在转录调控、基因组复制、DNA损伤修复、表观修饰等过程中发挥重要功能。

由转录-复制碰撞产生的R-loop作为基因组不稳定性的的重要因素, 需要被一系列机制来消除^[84]。叶绿体作为光合作用的主要场所, 可认为是实现“碳中和”宏伟目标的最基本单位, 其基因组的稳定性与光合作用的效率息息相关。在植物叶绿体中, 本课题组分别鉴定到核酸酶AtRNH1C和解旋酶RHON1可直接消除R-loop结构, 减小复制压力, 但还尚未在植物细胞核基因组中发现类似的机制。与质体基因组复制不同, 真核生物核基因组复制还伴随着染色质状态的改变, 如核小体解聚和重组装等过程, 因此核基因组复制过程中的R-loop解除机制更为复杂。近期, 有研究指出, 在动物细胞核中染色质重塑蛋白也可消除由R-loop介导的转录-复制冲突^[85,86], 这是一个理解R-loop代谢过程的新角度。除了对基因组稳定性起负面作用, 已有研究发现R-loop也参与基因组稳定性的维持。在叶绿体和细胞核基因组中的研究指出, 在DNA损伤位点处产生的一类瞬时RNA:DNA杂合链可作为早期同源重组信号帮助机体完成有效的同源重组修复^[17,87-89]。由此可见, R-loop可以作为“双刃剑”从正反两个方面实现某一生物学意义, 其丰富多彩的生物学功能也将吸引更多的关注。

除了上文提到的R-loop在植物细胞核中参与基因转录、可变剪切、复制等过程, 动物细胞中的研究揭示出R-loop还参与基因组上的表观修饰调控, 如抑制DNA甲基化修饰^[26,90,91], 促进H3K9me3和H3K27me3修饰^[92-94], 在果蝇中PRE(polycomb response elements)位点处富集的R-loop结构也能被PcG(polycomb group)蛋白识别进而调控发育基因上的组蛋白修饰^[95]。拟南芥中的ssDRIP-seq数据显示, R-loop与异染色质修饰信号H3K9me2和H3K27me1高度关联^[11], 但其背后的机制和生物学意义尚需进一步研究。

目前, R-loop的全基因组检测方法主要基于RNA:DNA识别蛋白和S9.6抗体的免疫沉淀(图3和4)。两类方法在蛋白/抗体特异性、免疫沉淀方式、建库方法

等方面有不同, 导致两类方法的测序结果存在一定差异^[96], 但目前均需要向单碱基分辨率和低起始量(甚至到单细胞水平)这两个方向优化, 以满足不同检测样品的需要. 此外, 随着计算生物学的发展, 基于热力学^[97,98]、模式搜索^[99]、形式文法^[100]、隐马尔可夫模型^[26]的R-loop预测工具相继被开发出来, 这些方法能在一定程度上预测R-loop结构, 帮助人们从“干”实验的角度理解R-loop结构形成原则(现阶段已经上线的R-loop高通量测序数据库链接RLBase: [https://gccri.](https://gccri.bishop-lab.uthscsa.edu/rlbase/)

[bishop-lab.uthscsa.edu/rlbase/](https://gccri.bishop-lab.uthscsa.edu/rlbase/); R-loopDB: <http://rloop.bii.a-star.edu.sg/>; R-loopAtlas: <http://bioinform.kib.ac.cn/R-loopAtlas/>; R-loopBase: <https://rloopbase.nju.edu.cn/>). 未来, 随着R-loop测序数据的不断产生以及计算工具日益成熟, 基于测序数据的深度学习也将有可能会是R-loop预测工具发展的新方向.

综上所述, 随着R-loop检测技术的提升、研究手段的成熟, 不同生理过程中R-loop的重要生物学功能将被逐一揭示.

参考文献

- García-Muse T, Aguilera A. R loops: from physiological to pathological roles. *Cell*, 2019, 179: 604–618
- Niehrs C, Luke B. Regulatory R-loops as facilitators of gene expression and genome stability. *Nat Rev Mol Cell Biol*, 2020, 21: 167–178
- Thomas M, White R L, Davis R W. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc Natl Acad Sci USA*, 1976, 73: 2294–2298
- Westover K D, Bushnell D A, Kornberg R D. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science*, 2004, 303: 1014–1016
- Kitani T, Yoda K, Ogawa T, et al. Evidence that discontinuous DNA replication in *Escherichia coli* is primed by approximately 10 to 12 residues of RNA starting with a purine. *J Mol Biol*, 1985, 184: 45–52
- Aguilera A, García-Muse T. R loops: from transcription byproducts to threats to genome stability. *Mol Cell*, 2012, 46: 115–124
- Liu J H, Xi K, Zhang X, et al. Structural flexibility of DNA-RNA hybrid duplex: stretching and twist-stretch coupling. *Biophys J*, 2019, 117: 74–86
- Tolokh I S, Pabit S A, Katz A M, et al. Why double-stranded RNA resists condensation. *Nucleic Acids Res*, 2014, 42: 10823–10831
- Sun Q, Csorba T, Skourti-Stathaki K, et al. R-loop stabilization represses antisense transcription at the *Arabidopsis FLC* locus. *Science*, 2013, 340: 619–621
- Shafiq S, Chen C, Yang J, et al. DNA topoisomerase I prevents R-loop accumulation to modulate auxin-regulated root development in rice. *Mol Plant*, 2017, 10: 821–833
- Xu W, Xu H, Li K, et al. The R-loop is a common chromatin feature of the *Arabidopsis* genome. *Nat Plants*, 2017, 3: 704–714
- Yang Z, Hou Q, Cheng L, et al. RNase H1 cooperates with DNA gyrases to restrict R-loops and maintain genome integrity in *Arabidopsis* chloroplasts. *Plant Cell*, 2017, 29: 2478–2497
- Yuan W, Zhou J, Tong J, et al. ALBA protein complex reads genic R-loops to maintain genome stability in *Arabidopsis*. *Sci Adv*, 2019, 5: eaav9040
- Xu W, Li K, Li S, et al. The R-loop atlas of *Arabidopsis* development and responses to environmental stimuli. *Plant Cell*, 2020, 32: 888–903
- Cheng L, Wang W, Yao Y, et al. Mitochondrial RNase H1 activity regulates R-loop homeostasis to maintain genome integrity and enable early embryogenesis in *Arabidopsis*. *PLoS Biol*, 2021, 19: e3001357
- Liu K, Sun Q. Intragenic tRNA-promoted R-loops orchestrate transcription interference for plant oxidative stress responses. *Plant Cell*, 2021, 33: 3574–3591
- Wang W, Li K, Yang Z, et al. RNase H1C collaborates with ssDNA binding proteins WHY1/3 and recombinase RecA1 to fulfill the DNA damage repair in *Arabidopsis* chloroplasts. *Nucleic Acids Res*, 2021, 49: 6771–6787
- Yang Z, Li M, Sun Q. RHON1 co-transcriptionally resolves R-loops for *Arabidopsis* chloroplast genome maintenance. *Cell Rep*, 2020, 30: 243–256.e5
- Bhatia V, Barroso S I, García-Rubio M L, et al. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature*, 2014, 511: 362–365

- 20 Ramirez P, Crouch R J, Cheung V G, et al. R-loop analysis by dot-blot. *J Vis Exp*, 2021, doi: 10.3791/62069
- 21 Yu K, Chedin F, Hsieh C L, et al. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol*, 2003, 4: 442–451
- 22 Gómez-González B, Aguilera A. Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proc Natl Acad Sci USA*, 2007, 104: 8409–8414
- 23 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: 636–647
- 24 Cerritelli S M, Crouch R J. Ribonuclease H: the enzymes in eukaryotes. *FEBS J*, 2009, 276: 1494–1505
- 25 Nowotny M, Gaidamakov S A, Ghirlando R, et al. Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol Cell*, 2007, 28: 264–276
- 26 Ginno P A, Lott P L, Christensen H C, et al. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell*, 2012, 45: 814–825
- 27 Chen L, Chen J Y, 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: 745–757.e5
- 28 Yan Q, Shields E J, Bonasio R, et al. Mapping native R-loops genome-wide using a targeted nuclease approach. *Cell Rep*, 2019, 29: 1369–1380.e5
- 29 Wulfridge P, Sarma K. A nuclease- and bisulfite-based strategy captures strand-specific R-loops genome-wide. *eLife*, 2021, 10: e65146
- 30 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: eabe3516
- 31 Boguslawski S J, Smith D E, Michalak M A, et al. Characterization of monoclonal antibody to DNA-RNA and its application to immunodetection of hybrids. *J Immunol Methods*, 1986, 89: 123–130
- 32 Hu Z, Zhang A, Storz G, et al. An antibody-based microarray assay for small RNA detection. *Nucleic Acids Res*, 2006, 34: e52
- 33 Dutrow N, Nix D A, Holt D, et al. Dynamic transcriptome of *Schizosaccharomyces pombe* shown by RNA-DNA hybrid mapping. *Nat Genet*, 2008, 40: 977–986
- 34 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: e1004716
- 35 Dumelie J G, Jaffrey S R. Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. *eLife*, 2017, 6: e28306
- 36 Sanz L A, Hartono S R, Lim Y W, et al. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol Cell*, 2016, 63: 167–178
- 37 Wahba L, Costantino L, Tan F J, et al. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev*, 2016, 30: 1327–1338
- 38 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: 46
- 39 Phillips D D, Garboczi D N, Singh K, et al. The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. *J Mol Recognit*, 2013, 26: 376–381
- 40 König F, Schubert T, Längst G. The monoclonal S9.6 antibody exhibits highly variable binding affinities towards different R-loop sequences. *PLoS ONE*, 2017, 12: e0178875
- 41 Xiao Q, Huang X, Zhang Y, et al. The landscape of promoter-centred RNA-DNA interactions in rice. *Nat Plants*, 2022, 8: 157–170
- 42 Liu Y, Liu Q, Su H, et al. Genome-wide mapping reveals R-loops associated with centromeric repeats in maize. *Genome Res*, 2021, 31: 1409–1418
- 43 Chen L, Xu W, Liu K, et al. 5' Half of specific tRNAs feeds back to promote corresponding tRNA gene transcription in vertebrate embryos. *Sci Adv*, 2021, 7: eabh0494
- 44 Yang X, Liu Q L, Xu W, et al. m⁶A promotes R-loop formation to facilitate transcription termination. *Cell Res*, 2019, 29: 1035–1038
- 45 Yan P, Liu Z, Song M, et al. Genome-wide R-loop landscapes during cell differentiation and reprogramming. *Cell Rep*, 2020, 32: 107870
- 46 Li Y, Song Y, Xu W, et al. R-loops coordinate with SOX2 in regulating reprogramming to pluripotency. *Sci Adv*, 2020, 6: eaba0777
- 47 Malig M, Hartono S R, Giafaglione J M, et al. Ultra-deep coverage single-molecule R-loop footprinting reveals principles of R-loop formation. *J*

- [Mol Biol](#), 2020, 432: 2271–2288
- 48 Pollen A A, Nowakowski T J, Shuga J, et al. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. [Nat Biotechnol](#), 2014, 32: 1053–1058
- 49 Tedesco M, Giannese F, Lazarević D, et al. Chromatin Velocity reveals epigenetic dynamics by single-cell profiling of heterochromatin and euchromatin. [Nat Biotechnol](#), 2022, 40: 235–244
- 50 Wimmers F, Donato M, Kuo A, et al. The single-cell epigenomic and transcriptional landscape of immunity to influenza vaccination. [Cell](#), 2021, 184: 3915–3935.e21
- 51 Conn V M, Hugouvieux V, Nayak A, et al. A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation. [Nat Plants](#), 2017, 3: 17053
- 52 Fang Y, Chen L, Lin K, et al. Characterization of functional relationships of R-loops with gene transcription and epigenetic modifications in rice. [Genome Res](#), 2019, 29: 1287–1297
- 53 Liu Y, Su H, Zhang J, et al. Back-spliced RNA from retrotransposon binds to centromere and regulates centromeric chromatin loops in maize. [PLoS Biol](#), 2020, 18: e3000582
- 54 Ariel F, Lucero L, Christ A, et al. R-loop mediated trans action of the APOLO long noncoding RNA. [Mol Cell](#), 2020, 77: 1055–1065.e4
- 55 Liu X, Gao Y, Liao J, et al. Genome-wide profiling of circular RNAs, alternative splicing, and R-loops in stem-differentiating xylem of *Populus trichocarpa*. [J Integr Plant Biol](#), 2021, 63: 1294–1308
- 56 Xu C, Wu Z, Duan H C, et al. R-loop resolution promotes co-transcriptional chromatin silencing. [Nat Commun](#), 2021, 12: 1790
- 57 Baxter C L, Šviković S, Sale J E, et al. The intersection of DNA replication with antisense 3' RNA processing in *Arabidopsis FLC* chromatin silencing. [Proc Natl Acad Sci USA](#), 2021, 118: e2107483118
- 58 Li S, Liu K, Sun Q. Comprehensive classification of the RNaseH-like domain-containing proteins in plants. [bioRxiv](#), 2019: 572842
- 59 Drouin G, Daoud H, Xia J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. [Mol Phylogenet Evol](#), 2008, 49: 827–831
- 60 Green B R. Chloroplast genomes of photosynthetic eukaryotes. [Plant J](#), 2011, 66: 34–44
- 61 Cavalier-Smith T. Only six kingdoms of life. [Proc R Soc Lond B](#), 2004, 271: 1251–1262
- 62 Timmis J N, Ayliffe M A, Huang C Y, et al. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. [Nat Rev Genet](#), 2004, 5: 123–135
- 63 Sakamoto W, Miyagishima S Y, Jarvis P. Chloroplast biogenesis: control of plastid development, protein import, division and inheritance. [Arabidopsis Book](#), 2008, 6: e0110
- 64 Gualberto J M, Newton K J. Plant mitochondrial genomes: dynamics and mechanisms of mutation. [Annu Rev Plant Biol](#), 2017, 68: 225–252
- 65 Xu B, Clayton D A. A persistent RNA-DNA hybrid is formed during transcription at a phylogenetically conserved mitochondrial DNA sequence. [Mol Cell Biol](#), 1995, 15: 580–589
- 66 Xu B, Clayton D A. RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers.. [EMBO J](#), 1996, 15: 3135–3143
- 67 Yasukawa T, Reyes A, Cluett T J, et al. Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. [EMBO J](#), 2006, 25: 5358–5371
- 68 Brown T A, Tkachuk A N, Clayton D A. Native R-loops persist throughout the mouse mitochondrial DNA genome. [J Biol Chem](#), 2008, 283: 36743–36751
- 69 Silva S, Camino L P, Aguilera A. Human mitochondrial degradosome prevents harmful mitochondrial R loops and mitochondrial genome instability. [Proc Natl Acad Sci USA](#), 2018, 115: 11024–11029
- 70 Holmes J B, Akman G, Wood S R, et al. Primer retention owing to the absence of RNase H1 is catastrophic for mitochondrial DNA replication. [Proc Natl Acad Sci USA](#), 2015, 112: 9334–9339
- 71 Reyes A, Kazak L, Wood S R, et al. Mitochondrial DNA replication proceeds via a ‘bootlace’ mechanism involving the incorporation of processed transcripts. [Nucleic Acids Res](#), 2013, 41: 5837–5850
- 72 Holt I J. The mitochondrial R-loop. [Nucleic Acids Res](#), 2019, 47: 5480–5489
- 73 Akman G, Desai R, Bailey L J, et al. Pathological ribonuclease H1 causes R-loop depletion and aberrant DNA segregation in mitochondria. [Proc Natl Acad Sci USA](#), 2016, 113: E4276–85

- 74 Wu Z, Cuthbert J M, Taylor D R, et al. The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. [Proc Natl Acad Sci USA](#), 2015, 112: 10185–10191
- 75 Cupp J D, Nielsen B L. Minireview: DNA replication in plant mitochondria. [Mitochondrion](#), 2014, 19: 231–237
- 76 Posse V, Al-Behadili A, Uhler J P, et al. RNase H1 directs origin-specific initiation of DNA replication in human mitochondria. [PLoS Genet](#), 2019, 15: e1007781
- 77 Al-Behadili A, Uhler J P, Berglund A K, et al. A two-nuclease pathway involving RNase H1 is required for primer removal at human mitochondrial OriL. [Nucleic Acids Res](#), 2018, 46: 9471–9483
- 78 Cerritelli S M, Frolova E G, Feng C, et al. Failure to produce mitochondrial DNA results in embryonic lethality in *Rnaseh1* null mice. [Mol Cell](#), 2003, 11: 807–815
- 79 Lima W F, Murray H M, Damle S S, et al. Viable *RNaseH1* knockout mice show RNaseH1 is essential for R loop processing, mitochondrial and liver function. [Nucleic Acids Res](#), 2016, 44: 5299–5312
- 80 Szczesny R J, Borowski L S, Malecki M, et al. RNA degradation in yeast and human mitochondria. [Biochim Biophys Acta](#), 2012, 1819: 1027–1034
- 81 Reyes-Prieto A, Weber A P M, Bhattacharya D. The origin and establishment of the plastid in algae and plants. [Annu Rev Genet](#), 2007, 41: 147–168
- 82 Bendich A J. DNA abandonment and the mechanisms of uniparental inheritance of mitochondria and chloroplasts. [Chromosome Res](#), 2013, 21: 287–296
- 83 Pfalz J, Pfannschmidt T. Essential nucleoid proteins in early chloroplast development. [Trends Plant Sci](#), 2013, 18: 186–194
- 84 Gómez-González B, Aguilera A. Transcription-mediated replication hindrance: a major driver of genome instability. [Genes Dev](#), 2019, 33: 1008–1026
- 85 Bayona-Feliu A, Barroso S, Muñoz S, et al. The SWI/SNF chromatin remodeling complex helps resolve R-loop-mediated transcription-replication conflicts. [Nat Genet](#), 2021, 53: 1050–1063
- 86 Prendergast L, McClurg U L, Hristova R, et al. Resolution of R-loops by INO80 promotes DNA replication and maintains cancer cell proliferation and viability. [Nat Commun](#), 2020, 11: 4534
- 87 Ohle C, Tesorero R, Schermann G, et al. Transient RNA-DNA hybrids are required for efficient double-strand break repair. [Cell](#), 2016, 167: 1001–1013.e7
- 88 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: 1314–1329.e10
- 89 Yasuhara T, Kato R, Hagiwara Y, et al. Human Rad52 promotes XPG-mediated R-loop processing to initiate transcription-associated homologous recombination repair. [Cell](#), 2018, 175: 558–570.e11
- 90 Grunseich C, Wang I X, Watts J A, et al. Senataxin mutation reveals how R-loops promote transcription by blocking DNA methylation at gene promoters. [Mol Cell](#), 2018, 69: 426–437.e7
- 91 Arab K, Karaulanov E, Musheev M, et al. GADD45A binds R-loops and recruits TET1 to CpG island promoters. [Nat Genet](#), 2019, 51: 217–223
- 92 Groh M, Lufino M M P, 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: e1004318
- 93 Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot N J. R-loops induce repressive chromatin marks over mammalian gene terminators. [Nature](#), 2014, 516: 436–439
- 94 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: 930–945.e4
- 95 Alecki C, Chiwara V, Sanz L A, et al. RNA-DNA strand exchange by the *Drosophila* polycomb complex PRC2. [Nat Commun](#), 2020, 11: 1781
- 96 Miller H E, Montemayor D, Abdul J, et al. Quality-controlled R-loop meta-analysis reveals the characteristics of R-Loop consensus regions. [bioRxiv](#), 2021: 2021.11.01.466823
- 97 Huppert J L. Thermodynamic prediction of RNA-DNA duplex-forming regions in the human genome. [Mol Biosyst](#), 2008, 4: 686
- 98 Stolz R, Sulthana S, Hartono S R, et al. Interplay between DNA sequence and negative superhelicity drives R-loop structures. [Proc Natl Acad Sci USA](#), 2019, 116: 6260–6269
- 99 Jenjaroenpun P, Wongsurawat T, Yenamandra S P, et al. QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis

of R-loop forming sequences. *Nucleic Acids Res*, 2015, 43: W527–W534

- 100 Jonoska N, Obatake N, Poznanović S, et al. Modeling RNA:DNA hybrids with formal grammars. In: Segal R, Shtylla B, Sindi S, eds. *Using Mathematics to Understand Biological Complexity*. Association for Women in Mathematics Series. Cham: Springer, 2021. 35–54

Recent progress on R-loop biology: from detection methods to biological functions

ZHOU JinCong^{1,2}, WANG WenJie^{1,2} & SUN QianWen^{1,2}

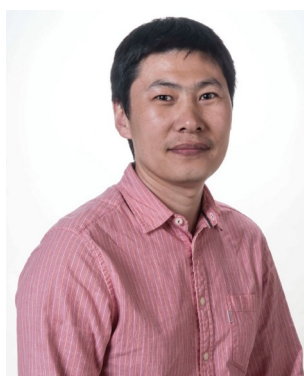
1 Center for Plant Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China;

2 Tsinghua-Peking Center for Life Sciences, Beijing 100084, China

R-loops are three-stranded chromatin structures composed of a single-stranded DNA and an RNA:DNA hybrid with important biological functions in numerous organisms. Numbers of R-loop regulators have been found to affect genomic events, such as transcription, replication, DNA damage, and repair, through modulating R-loop homeostasis. Over the last few years, with the development of R-loop detection methods, an increasing number of R-loop biological functions are being discovered. In this review, we provide an overview of R-loop detection methods and focus on the research progress of R-loop functions in plants.

RNA:DNA hybrids, R-loops, genome regulation

doi: [10.1360/SSV-2021-0422](https://doi.org/10.1360/SSV-2021-0422)



孙前文, 博士, 清华大学生命科学学院副教授, 清华大学-北京大学生命科学联合中心研究员. 2019年获得国家自然科学基金优秀青年科学基金项目资助. 孙前文实验室长期专注R-loop生物学与植物基因组调控研究, 在领域内有一系列重要的发现, 如在植物中发现R-loop在调控基因转录和生长发育的新功能, 绘制植物R-loops全基因组分布特征图谱, 发现R-loop调节叶绿体基因组稳定性的全新途径等. 以通讯或共同通讯作者在*Nature Plants*, *The Plant Cell*, *PLOS Biology*, *Science Advances*以及*Molecular Plant*等杂志发表多篇论文.