



基因编辑设计和分析辅助工具的研究进展

李福权, 张瑞祥, 赵哲, 陈乐天*, 谢先荣*

华南农业大学生命科学学院, 亚热带农业生物资源保护与利用国家重点实验室, 广州 510642

* 联系人, E-mail: lotichen@scau.edu.cn; xiexianrong@scau.edu.cn

2023-12-25 收稿, 2024-05-10 修回, 2024-05-13 接受, 2024-05-14 网络版发表

国家自然科学基金(31991222, 31991223)资助

摘要 CRISPR/Cas基因编辑系统因操作简便、效率高优势, 已成为生命科学领域的革命性技术, 被广泛用于基因功能研究、医学治疗和动植物遗传改良等领域. 随着基因编辑系统的不断发展, 目前已衍生出包括基因敲除、碱基编辑、先导编辑、替换和插入、RNA编辑等类型丰富的基因编辑工具. 在运用基因编辑系统开展研究时, 应根据需求选择合适的基因编辑工具, 同时还需要充分考虑sgRNA的效率和脱靶率, 以及后续检测靶点突变和脱靶的成本与可靠性. 目前, 已开发了适用于不同基因编辑系统的靶点设计、脱靶评估和突变检测等辅助工具和方法, 使用这些工具可以极大地提高基因编辑实验的便捷性和效率. 本文简要介绍了现有基因编辑工具的类型和原理, 重点综述了用于基因编辑的靶点设计、靶点突变分析和检测的方法与辅助工具, 为利用这些方法和工具提高基因编辑实验的效率提供参考.

关键词 CRISPR/Cas, 基因编辑, 靶点设计, 脱靶, 靶点检测, 计算机辅助工具

成簇规律间隔短回文重复序列(clustered regularly interspaced short palindromic repeats, CRISPR)/CRISPR相关蛋白(CRISPR-associated, Cas)系统是广泛存在于细菌以及古细菌基因组中的一段CRISPR重复序列和Cas蛋白相关的序列, 主要参与细菌的免疫反应, 避免细菌被外来核酸的侵害^[1,2]. 在外来核酸入侵细菌后, CRISPR序列会进行转录, 生成单链引导RNA(single guide RNA, sgRNA), Cas蛋白会在sgRNA引导下对外来核酸进行切割^[3,4]. 基于Cas蛋白可以在sgRNA引导下对特定的核酸片段进行精确切割的特点, 多种由CRISPR/Cas系统改造的基因编辑工具逐渐被开发, 如碱基编辑器、先导编辑器、RNA编辑系统等^[5], 应用于基因功能研究、医学治疗和动植物遗传改良等领域^[6-9].

利用CRISPR/Cas系统对动植物进行基因编辑之

前, 应充分考虑Cas蛋白相关特性和靶点序列特征, 如不同类型的Cas蛋白所识别的PAM(proto-spacer adjacent motif)序列、编辑的核酸类型、靶点编辑效率和特异性以及期望的突变类型等因素, 从而选择合适的基因编辑工具和靶点. 此外, 随着基因编辑工具的广泛应用, 对基因编辑个体的突变检测量急剧增加. 因此, 需要根据研究目的, 考虑现有不同检测方法的成本和效率, 以便实现在保证检测准确性的前提下节约研究成本. 目前, 已开发使用的多种基因编辑靶点设计、评估和检测相关的计算机辅助工具以及检测方法, 极大地提高了基因编辑实验设计的便捷性和成功率. 本文简要总结了现有基因编辑系统的类型和特点, 重点介绍基因编辑靶点设计、靶点突变检测、脱靶分析的辅助工具和方法, 展望CRISPR/Cas基因编辑系统辅助工具和方法所面临的挑战.

引用格式: 李福权, 张瑞祥, 赵哲, 等. 基因编辑设计和分析辅助工具的研究进展. 科学通报, 2025, 70: 2449–2467

Li F Q, Zhang R X, Zhao Z, et al. Research advances in gene editing design and analysis tools (in Chinese). Chin Sci Bull, 2025, 70: 2449–2467, doi: 10.1360/TB-2023-1334

1 CRISPR/Cas基因编辑系统的类型和特点

1.1 CRISPR/Cas基因编辑技术的建立和发展

自DNA与RNA被确立为遗传物质以来,人们从未停止对DNA和RNA的分子操作,以操纵生物的目的性状.第一代基因编辑是锌指核酸酶(zinc-finger nuclease, ZFN)技术,它通过Fok I的核酸内切酶特性在基因组特定位置产生双链断裂(double-strand break, DSB)^[10],再通过真核细胞内同源重组(homology directed repair, HDR)或非同源末端连接(non-homologous end joining, NHEJ)的修复途径完成对基因组DNA的编辑^[11].但是由于编辑窗口的限制和ZFN表达载体难以构建的问题,该技术受到严重的限制.二代基因编辑技术是利用TAL(transcription activator-like)效应子与Fok I核酸酶融合而开发的转录激活因子样效应核酸酶(transcription activator-like effector nuclease, TALEN).TALEN技术同样能够在基因组上产生DSB,并实现基因敲除^[12],但这项技术的编辑特异性仍未达到要求,且载体构建较为复杂^[13],难以解决多个基因同时编辑的问题.

CRISPR系统是一种高度保守的重复序列和完全不同的间隔序列交替排列组成的序列,属于细菌和古菌应对环境压力和病毒侵害的特殊防御机制中的一部分^[14,15].在众多CRISPR系统中,用于负责识别CRISPR的典型基因为Cas基因,如*cas1*、*cas2*、*cas4*等.这些基因编码的Cas效应蛋白以不同的形式组合,在CRISPR系统中呈现不同的功能^[16].根据这些Cas效应蛋白的组成将CRISPR系统分Class I和Class II,根据其作用方式主要划分了6种类型(types I~VI).在这两类CRISPR系统中,Class I(types I, III, IV)需要多个蛋白亚基组成复合体来行使核酸酶功能,而Class II(types II, V, VI)则只需在单一Cas效应蛋白与相应sgRNA行使靶向的核酸酶功能^[17-19].近期,麻省理工学院的张锋团队发现含有 β -CASP核酸酶结构域的VII类CRISPR系统,拥有RNA靶向的RNA切割能力.这一发现导致Cas14蛋白家族被重新划分至II类CRISPR以便更好地纳入现有系统^[20].尽管目前已经挖掘并检测了多个CRISPR系统的体外DNA切割能力,但是目前在真核细胞中能够高效作用于染色体工程的系统仍然以属于Class II-Type II和Class II-Type V的SpCas9和AsCas12a为主^[21-23].以上述2种Cas效应蛋白为底盘,研究人员开发了多个基因编辑工具集,并迅速取代了ZFN和TALEN介导的基因组编辑技术.

以Cas9和Cas12家族蛋白为工程化研究平台,研究人员不仅仅满足于简单的基因敲除,而是聚焦更精准的基因组编辑操作,开发了能够从碱基尺度到DNA大片段尺度进行精确突变的一系列编辑工具.目前,碱基编辑器(base editors, BEs)可实现对4种碱基进行全部12种类型的突变, BEs包括胞嘧啶碱基编辑器(cytosine base editor, CBE)^[24]、腺嘌呤碱基编辑器(adenine base editor, ABE)^[25,26]、胞嘧啶颠换碱基编辑器(cytosine transversion base editor, CGBE)^[27]和胞嘧啶腺嘌呤双碱基编辑器(saturated targeted endogenetic mutagenesis editor, STEME)^[28].针对精确的基因片段插入,研究人员开发了先导编辑(prime editor, PE)^[29]和基因插入打靶(gene targeting, GT)^[30].在研究表观遗传时,表观编辑器(plant epigenetic editors, PEEs)逐渐成为良好的表观变异材料创制工具^[31].Cas13a(Type VI)可以应用于RNA编辑器,使得转录水平的精确调控不完全依赖于传统的RNA干扰技术(RNA interference, RNAi).此外,组织特异性基因编辑(tissue-specific gene editing, TSGE)^[32]和基于CRISPR的染色体成像^[33]等工具也在日趋完善.总而言之,基于CRISPR的基因编辑系统已成为解析复杂的基因调控网络以及基因组操作的最有效手段.

1.2 不同CRISPR/Cas效应蛋白的特点和应用

CRISPR/Cas系统广泛存在于原核生物和巨型噬菌体中,并作为免疫机制响应外来核酸的入侵^[17,34].通常CRISPR参与的免疫应答反应有3个阶段,即外源DNA获得并识别阶段、CRISPR基因座表达加工阶段和靶向外源DNA的干扰阶段.该系统结构清晰,易于工程化,很快被用于开发真核生物的基因编辑工具^[35].在众多CRISPR系统中,有4类效应器被作为底盘广泛应用,它们分别是Cas9(Class II type II)、Cas12(Class II type V)、Cas13(Class II type VI)和Cascade-Cas3(Class I type I).

最早应用于动植物细胞的CRISPR/Cas9核酸酶,Cas9蛋白需要在crRNA(CRISPR RNA)和tracrRNA(trans-activating crRNA)的招募下并在特定位置切割DNA双链(图1(a)).具体的是,sgRNA由crRNA和tracrRNA组成,它们分别用于定位基因组和招募Cas9蛋白^[35].当CRISPR系统发挥功能时,Cas9蛋白和sgRNA形成一个蛋白-RNA复合体,Cas9蛋白能够识别PAM(protospacer adjacent motif, 5'-NGG-3')序列,在sgRNA的协助下,Cas9蛋白将识别位于PAM附近、与sgRNA互补的DNA双链并切割形成DSB.随后,通过细胞内NHEJ或HDR修

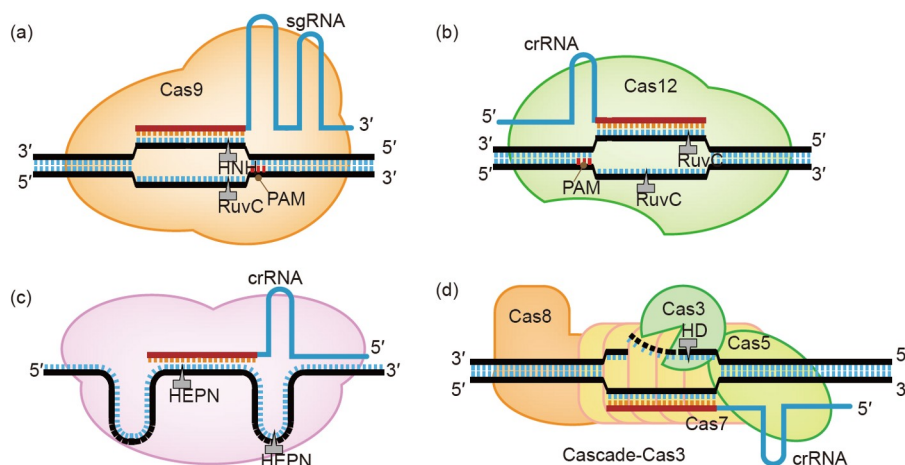


图1 基于CRISPR的主要编辑器的基本原理和特点。(a) CRISPR/Cas9作用的基本原理; (b) CRISPR/Cas12作用的基本原理; (c) CRISPR/Cas13作用的基本原理; (d) CRISPR/Cascade-Cas3作用的基本原理

Figure 1 Basic principles and features of major CRISPR-based editors. (a) Basic principle of CRISPR/Cas9 action. (b) Basic principle of CRISPR/Cas12 action. (c) Basic principle of CRISPR/Cas13 action. (d) Basic principle of CRISPR/Cascade-Cas3 action

复途径的作用,在靶位点处引入碱基插入或缺失(insertion and deletion, Indel)形成移码突变或功能结构受损而被功能敲除^[36]。Cas9作为最广泛开发的底盘蛋白,拥有HNH和RuvC两个结构域,这使得其工程化更加便利、应用更全面广泛。两个独立的核酸酶结构域分别作用于靶标链和与靶标链互补的非目标链。因而,通过改变核酸酶结构域的关键氨基酸开发的Cas9切口酶(Cas9 nickase, nCas9; D10A或H840A)和失活Cas9核酸酶(dead Cas9, dCas9; D10A和H840A)能够更灵活地作为RNA引导的基因组靶向底盘,使用nCas9和dCas9融合效应蛋白形成的各种编辑器组成了丰富的CRISPR工具箱^[37]。值得注意的是,其进化关系上的祖先IscB拥有同样的核酸酶特征结构域,其长度只有400个氨基酸左右,约为常见的SpCas9的40%^[38,39]。IscB蛋白是编码在IS200/IS605转座子家族中的核酸酶,这些由RNA引导的核酸酶被张锋团队命名为OMEGA(obligate mobile element-guided activity),其中OgeuIscB在人类基因组编辑中获得了4.4%的编辑效率^[38]。基于OgeuIscB的蛋白质进化有望开发出一系列微型的核酸酶工具,这些进展有望使得基于腺相关病毒(adeno-associated virus, AAV)的动物基因组工程和基于番茄斑萎病毒(tomato spotted wilt virus, TSWV)的植物基因编辑^[40]获得新的突破,使得基因编辑和基因组工程的应用范围进一步扩展。

另外,应用较广泛的还有Type V的Cas12家族蛋白,Cas12家族的一些成员拥有独特的特点和应用价值。例如,Cas12家族的效应蛋白大多只具备RuvC结构

域,在切割DNA双链时一些亚家族效应蛋白还需要Nuc调节结构域的参与,使其在切割DNA双链时无法同时切割,这导致DNA链因为扭转切割而产生黏性末端(图1(b))。部分Cas12亚家族成员如Cas12a不需要tracrRNA,且自身具备加工crRNA的能力,这使得Cas12a具有更易于工程化的crRNA^[41]。与Cas9不同的是Cas12家族蛋白主要识别5' T-rich的PAM序列,这使其在一些物种中全基因组靶向范围区别于Cas9^[41]。Cas12家族近期受到了更多的关注,研究人员对Cas12亚家族蛋白进行了表征和挖掘,发现了丰富的亚家族蛋白和它们一些独特的结构,如Cas12m和Cas12k具备非典型的RuvC结构域^[42-44]。Cas12家族有望作为一种新的底盘对基因组进行更全面的编辑,这弥补了Cas9在一些应用上的不足。

RNA编辑是一种区别于RNAi和CRISPRi(CRISPR interference)的技术,其基于一些能够直接精准靶向RNA的Cas蛋白如Cas13家族蛋白来实现。Cas13家族蛋白拥有两个HEPN核酸酶结构域,前间隔侧翼序列(proto-spacer flanking sequence, PFS)在细胞内作用的偏好性不明显,靶向RNA的范围广泛^[45](图1(c))。但是,Cas13蛋白具有旁系切割的活性,即识别并切割靶标RNA的同时随机切割体系中临近的单链RNA(single strand RNA, ssRNA)^[46]。这一特点便于其作为一些RNA检测的生物传感器,但限制了其在临床上的使用。杨辉团队筛选了hfCas13d,有效降低了对ssRNA的旁系降解能力,使得Cas13家族有望应用于RNA疾病的临床治疗^[47,48]。对转录水平的精确调控一直是Cas13家族蛋白

开发和应用的方向,这对一些疾病的临床治疗有极大的作用.此外在一些基因研究的过程中对其DNA直接编辑会导致细胞或受精卵的死亡,Cas13对RNA的敲低可能是这类基因研究过程中的有力技术.

CRISPR Type I系统需要*cas5*、*cas8*和*cas7*基因表达的相应蛋白形成亚基聚合结构,为crRNA引导的Cascade复合物,该复合物通过招募Cas3执行DNA切割功能(图1(d)).Cascade-Cas3在结合位点3'端附近相隔5~6个核苷酸的2个位点上的裂解,这种裂解是单向的,会使得NTS链的大量片段被消化,这一特性使得Cascade-Cas3系统被应用于长片段的基因敲除.研究人员使用属于Type I-C的Cascade-Cas3实现了长达434 kb的碱基删除^[49].此外,基于I-E和I-D的核酸酶工程开发的Cascade也成功应用于动植物细胞中^[50-52].

上述提到的四种底盘系统,被开发出一系列基因编辑工具,为科学研究提供了全面的技术支持.随着一些更新更小的Cas蛋白如IscB和TnpB被应用并开发为新的底盘,未来的基因编辑工具递送手段也将迎来全面的进步.

2 基因编辑靶点设计

2.1 影响基因编辑效率的因素

基因编辑效率指在转化/转染事件中,目标靶点被成功编辑(发生插入、缺失或替换等)的百分比.它受到

Cas蛋白密码子、Cas蛋白表达量、sgRNA启动子、靶标序列、sgRNA序列特征以及染色质可接近性等多个因素的影响^[53].研究表明,利用植物密码子优化的*Cas9*基因,在单子叶植物和双子叶植物中可进行方便高效的多重基因组编辑^[54].不同的研究发现,靶标序列GC含量的高低对编辑效率有不同影响,如水稻中GC含量较高的靶序列具有相对较高的编辑效率^[54],GC含量在65%~80%之间会增加意外突变的风险.编辑效率也可能与sgRNA的PAM近端部分的GC含量呈正相关^[55].当sgRNA靶标的3'端序列为T时,目标基因的编辑效率较低^[56].最近有研究发现DNA的拓扑结构也会影响基因编辑效率^[57].

2.2 基因敲除的靶点设计和工具

基因编辑中目标基因的靶点选择、sgRNA表达盒的设计、靶点片段的扩增与测序、脱靶分析等过程耗费大量的人力物力.近年来,利用生物信息工具辅助基因编辑的靶点设计和分析,既减少人为错误,也极大地节约时间和人力成本,使整个基因编辑过程更加简单、高效.目前,已有多种针对不同物种设计靶点的工具得到开发,如CRISPROR、CHOPCHOP、CRISPR RGEN Tools、CRISPR-GE、CRISPR-P等^[58-66](表1).

CRISPR RGEN Tools是一个可设计CRISPR/Cas、ZFN以及TALEN靶点的在线工具箱,集成了包括靶点设计、脱靶分析、微同源分析、单碱基编辑和引导编

表1 常用靶点设计工具

Table 1 Common tools for designing targets

工具	特点	网站	参考文献
CRISPROR	包含超过100个物种的基因组以及30种Cas蛋白及其变体,为靶点提供引物设计,可检测脱靶位点	https://crispor.tefor.net/	[60]
CHOPCHOP	包含超过100个物种的基因组,可选择Cas9、Cas12、Cas13以及TALEN基因编辑工具,可为基因的敲除、敲入、激活及其沉默进行靶点的设计	https://chopchop.cbu.uib.no/	[61]
CRISPR RGEN Tools	包含超过100个物种以及40种Cas蛋白及其变体,可对基因敲除、单碱基编辑和先导编辑进行靶点设计,还可对编辑结果进行分析	https://www.rgenome.net/cas-designer/	[58,62,63]
CRISPR-GE	包含超过40种植物和5种模式动物基因组,可自定义Cas蛋白特性进行靶点设计,网站还包括序列提取、脱靶评估和靶点突变解码等工具模块	http://skl.scau.edu.cn/	[59]
CRISPR-P 2.0	包含82个植物高质量基因组,可选择Cas9、Cpf1类的Cas蛋白及其变体进行靶点设计,预测每个靶点的潜在脱靶信息	http://crispr.hzau.edu.cn/CRISPR2/	[64]
CCTop	包含超过100个物种的基因组以及17种Cas蛋白及其变体,预测每个靶点的潜在脱靶位点	https://cctop.cos.uni-heidelberg.de/	[65]
CRISPRscan	包含22个物种的基因组以及7个Cas蛋白及其变体,直观显示靶点在基因组上的位置,预测每个靶点潜在的脱靶位点	https://www.crisprscan.org/	[66]

辑等多个设计和分析工具。其Cas-Designer模块可对输入的靶标序列、物种、PAM序列等信息进行处理,输出特定序列范围内可设计的靶点序列,显示核酸酶切割位置、sgRNA序列的GC含量、Out-of-frame Score等信息,同时利用Cas-Offinder算法预测每个设计的靶点在对应物种的基因组中潜在的脱靶序列^[58]。CRISPR RGEN Tools目前已支持在超过100个动植物物种的基因组中设计靶点,还可选择超过40种不同的Cas蛋白及其对应的PAM序列,便于用户根据靶序列的特性选择合适的Cas蛋白,从而设计更为高效和特异的sgRNA。

CRISPR-GE是由刘耀光院士团队开发、主要为植物基因组进行靶点设计的在线工具。其targetDesign模块中可基于输入的靶标序列进行搜索,输出含有对应PAM的所有靶点序列及其在靶标序列的位置、GC含量、靶点序列与sgRNA配对的碱基数、在对应基因组中的位置以及潜在的脱靶位点等信息,同时,CRISPR-GE工具还会对GC含量过高($\geq 80\%$)或过低($\leq 25\%$)、超过4个连续的T碱基、与sgRNA有连续配对的低质量靶点序列进行警示标记,不建议用户选择这类靶点^[59]。CRISPR-GE工具目前可选择超过40种动植物的基因组序列,还可选择线粒体以及叶绿体作为参考基因组进行靶点设计,同时,CRISPR-GE工具还可以基于无参考基因组的序列进行靶点设计。在Cas蛋白选择方面,CRISPR-GE中目前内置最为常用的SpCas9、FnCpf1、AsCpf1和LbCpf1这4种Cas蛋白,同时用户可在CRISPR-GE工具中自定义Cas蛋白的特性,输入任意的PAM序列、PAM序列在靶点上的方向以及靶点长度,CRISPR-GE会基于输入的信息对特定序列进行分析处理,搜索其中的靶点序列。常用的专门针对植物基因组靶点设计的工具还包括CRISPR-P,该工具的更新版本CRISPR-P 2.0^[64]目前收录了共82个植物基因组,包括了水稻、油菜等多个已发表品种的高质量基因组,支持Cas9、Cpf1及它们变体共14种PAM类型的靶点设计,靶点设计的结果页面提供了丰富的可视化界面交互,分析结果包括所有靶点的预测靶点编辑效率、GC含量、在基因上的位置和脱靶位点等信息,选择相关靶点后,可进一步开展sgRNA二级结构分析和靶点处切割后产生微同源修复的预测值。

2.3 单碱基编辑和先导编辑的靶点设计和工具

另外,基于CRISPR/Cas工具所开发的单碱基编辑系统和先导编辑系统也已在基因编辑领域被广泛应

用^[67]。其中,单碱基编辑除了改变特定氨基酸密码子,也可在开放阅读框中引入终止密码子,使基因表达提前终止。此外,单碱基编辑具有不确定性,在编辑窗口中,与对应的碱基相近的碱基也有可能被脱氨酶催化转变为另一种碱基。因此对单碱基编辑的靶点设计,应重点考虑几个方面的影响,如靶点编辑窗口的选择、旁观者效应、潜在的密码子变化、突变后氨基酸的变化以及脱靶效应。为了方便实验设计,现已开发了多种辅助设计的工具,如BE-Hive^[68]、DeepBaseEditor^[69]、BE-DICT^[70]等工具可预测哺乳动物单碱基编辑的效率以及旁观者效应(表2)。CRISPR RGEN Tools中的BE-Designer模块以及CRISPR-GE中的BE-target模块均可对快速分析靶标序列中的候选靶位点,输出的编辑窗口以及碱基突变后的氨基酸变化,同时可在对应物种中检测潜在的脱靶位点^[62,71]。BE-Designer模块中有超过100个参考基因组以及20种不同Cas蛋白的单碱基编辑系统可选择。BEtarget模块中包含50多个参考基因组以及SpCas9、SpCas9-NG、FnCpf1、AsCpf1、LbCpf1常见的Cas蛋白可选择,同时支持自定义PAM类型,包括PAM序列、位置、靶点的长度以及编辑的窗口,此外,工具支持对无参考基因组的序列进行靶点的设计。使用这些工具进行单碱基靶点设计的主要便利是可以根据编辑窗口快速预测密码子变化、旁观者效应和脱靶情况,以提高靶点设计的效率。

对于先导编辑,已有研究表明,pegRNA的设计对于PE系统的编辑效率有显著的影响^[29,74]。目前对PE系统中的pegRNA设计的工具主要有PnB Designer^[75]、PlantPegDesigner^[76],以及CRISPR RGEN Tools中的PE-Designer工具。在PE-Designer工具中,输入参考的序列以及编辑后所期望的序列,即可得到序列中可进行设计的靶点信息,包括靶点的位置、GC含量以及潜在的脱靶位点等^[63]。PE-Designer工具中目前可选择超过100个参考基因组以及15个用于先导编辑的Cas蛋白,用户可基于所编辑的材料以及靶标序列的特征,选择不同的参考基因组和不同PAM序列的Cas蛋白,进行PE系统的靶点设计。

Lin等人^[76]发现primer结合位点(prime binding site, PBS)与DNA的 T_m 值为30℃的pegRNA可提高PE系统在植物中的编辑效率,并且带有2个pegRNA的PE系统其编辑效率显著高于仅带有单个pegRNA的PE系统,对此,研究人员进一步开发了PlantPegDesigner在线工具,可对植物的PE编辑实验快速设计具有高活性的pegR-

表 2 单碱基编辑以及先导编辑常用靶点设计工具

Table 2 Common target design tools for base editing and prime editing

工具	特点	网站	参考文献
BE-Designer	支持CBE、ABE和CGBE编辑器类型, 可选择超过100个参考基因组以及20种不同Cas蛋白的单碱基编辑系统, 结果提供氨基酸变化、潜在脱靶位点信息	http://www.rgenome.net/be-designer/	[62]
BE-Hive	可选择33个碱基编辑系统和细胞类型的组合, 基于机器学习模型对单碱基编辑结果进行预测(包括旁观者效应), 同时预测编辑后氨基酸的变化	https://www.crisprbehive.design/	[68]
DeepBaseEditor	可选择ABE以及CEB两个单碱基编辑系统, 预测特定靶点序列的编辑效率	https://deepcrispr.info/DeepBaseEditor/	[69]
BETarget	包含50多个参考基因组(主要为植物), 可自定义PAM序列、靶点序列长度等特点, 支持无参考序列的靶点设计, 结果包含氨基酸变化、潜在脱靶位点等信息	http://skl.scau.edu.cn/betarget/	[71]
CRISPR-BETS	可选择超过100个物种的参考序列, 支持GenBank、Snapgene或Fasta格式文件的输入, 主要用于设计通过单碱基编辑引入终止密码子从而突变目的基因	https://bioinform.yzu.edu.cn/software/crispr-bets	[72]
FORECasT-BE	基于深度学习对ABE和CBE指导的编辑效率和结果频率进行预测, 模型预测的准确性较高(Pearson相关性在0.50到0.95之间)	https://partslab.sanger.ac.uk/FORECasT-BE	[73]
PnB Designer	支持PE和单碱基编辑系统, 包含6个参考基因组序列, 可根据用于预期突变结果输出合适的靶点	https://fgcz-shiny.uzh.ch/PnBDesigner/	[75]
PE-Designer	专门用于pegRNA选择的工具, 包含超过100个参考基因组以及15个用于先导编辑的Cas蛋白, 提供所有可能的目标序列、pegRNA延伸序列、和切口引导RNA序列等信息	http://www.rgenome.net/pe-designer/	[63]
PlantPegDesigner	主要用于植物pegRNA设计, 可选择双pegRNA策略, 自动设计长度、 T_m 值以及GC含量最为合适的PBS序列, 同时设计对应的引物	http://www.plantgenomeediting.net/	[76]

NA, 在输入期望编辑的序列, 以及设置PAM序列、间隔序列长度、PBS长度等各种参数后, PlantPegDesigner会在靶标序列中分析可进行编辑的靶点序列, 输出不同长度的PBS序列, 并且标记长度、 T_m 值以及GC含量最为合适的PBS序列, 同时为所推荐的PBS序列进行引物设计, 提高了设计高效pegRNA的便捷性。

3 基因编辑靶点突变分析

基因编辑在靶点产生切割后产生DSB, 在不提供外源模板时, 主要的修复方式是NHEJ^[77,78], 进而产生碱基的缺失或插入以及少量的碱基替换。当前检测突变的分析方法大致可分为非测序依赖的快速检测和基于测序的检测方法。在二倍体中, 根据编辑后在两条染色体的靶点是否产生一致的突变, 可以分为纯合突变、杂合突变和双等位突变^[54]。纯合突变是指两条染色体发生了一致的序列突变; 杂合突变是指其中一条染色体发生了突变而另一条未突变; 双等位突变是指两条染色体均发生了突变但突变序列不一致。部分二倍体(如大豆)由于转化方法易造成嵌合体现象, 对靶点序列扩增后, 会出现2种以上的突变类型。在多倍体中,

特别是异源多倍体, 由于基因组倍性复杂, 难以具体区分是哪几条染色体发生了突变, 但可以判断是否纯合或多等位突变以及对应的突变序列, 如果是基于二代测序的方法, 可以进一步根据深度差异判断突变的染色体数目。目前已有多种非测序或测序以来的突变检测方法(表3)。

3.1 非测序依赖的突变检测方法

非测序依赖的方法包括常用的酶切法、T7EI、荧光定量、HRM等。基于限制性内切酶的酶切法在设计靶点时, 所选择的靶点序列在Cas酶切割位点处必须存在相应的酶切位点, 局限性较大, 例如Cas9的切割位点通常在PAM序列上游约3~4个核苷酸处。如果靶点处发生了切割, 则导致对应的限制性内切酶的识别位点破坏而无法酶切^[79,80]。

T7EI方法是基于编辑后的靶点序列由于引入了突变, 与野生型DNA片段退火后产生异质双链DNA^[81]。T7核酸内切酶 I (T7EI)可以识别该异质双链DNA上的不完全配对的位点, 然后进行双链切割, 因此通过琼脂糖凝胶电泳即可显示酶切后的条带, 从而判定基因编

表 3 不同检测方法的比较

Table 3 Comparison of different detection methods

比较项	非测序依赖的方法	测序依赖的方法	
		基于Sanger测序	基于二代测序
检测类型	InDel等常见的简单突变	常见的杂合、双等位和纯合突变	所有突变类型
推荐方法	T7EI、酶切法、Cc-qPCR、HRM等	DSD、DSDecodeM、TIDE等	Hi-TOM、CrisprStitch、Cas-Analyzer等
预估成本	根据所使用的试剂, 1~8元/靶点	12~18元/靶点	低至1元/靶点
特点	快速	快速、准确	通量高、成本低
适用范围	初筛发生了突变的样本, 无需获知具体突变序列	少量二倍体样本的突变鉴定	大量样本的突变鉴定, 嵌合突变和多倍体样本

辑效果。该方法与基于限制性内切酶的酶切法相比, 不依赖于特定的酶切位点, 检测的靶点范围受限制少。但是, T7EI除了识别因错配产生的异质双链DNA外, 还能识别一些DNA高级结构, 如十字结构、Holiday连接体等, 所以就有可能在检测的时候出现假阳性。

此外, 还有利用定量PCR方法检测突变的方法, 如Li等人^[82]开发的通过在体外Cas9酶切割辅助的单管双联定量实时PCR(Cc-qPCR)方法。Cc-qPCR利用体外表达的sgRNA(和载体构建所用的sgRNA是同一个)和Cas9蛋白同时靶向切割待检测的基因组DNA, 然后进行qPCR检测。如果待检测位点为野生型, 那么Cas9可以将两条染色体上的DNA均切开, 在后续qPCR时得不到产物; 如果是杂合突变, 会有一半可以正常扩增, 产量相对于野生型增加一半; 如果是纯合突变, 那么两条染色体上的DNA均可以正常扩增。以上3种结果反映在qPCR曲线上就是3种不同的状态, 进而可以进行快速检测突变情况。

上述通过酶切、定量等方法只能评估靶点处是否发生突变, 但无法获得具体的突变序列, 且灵敏度和可靠性较差。基于测序的方法, 包括一代Sanger测序和二代测序的方法, 可以准确获得具体的突变类型和序列, 并且由于测序成本的下降, 基于测序的鉴定方法已成为了主要的检测手段。

3.2 基于Sanger测序的突变分析方法和工具

在二倍体中进行编辑后, 如果靶点发生了杂合突变或双等位突变时, 直接对包含靶点的序列进行测序时, 所得的测序峰图会在突变起点处产生复杂的双峰或套峰, 而无法直接读取序列。在没有开发相关方法和工具对Sanger峰图进行解码分析前, 常用的做法是通过

将PCR片段克隆至T载体转化至细菌后, 挑选若干单克隆进行测序, 从而判断突变的序列和类型, 过程繁琐、耗时。由于Cas蛋白在靶点处产生的DSB是由NHEJ介导修复, 这过程中以引入核苷酸的缺失为主, 以及1~2 nt的核苷酸插入或替换。基于分析Cas9介导的突变特征, Ma等人^[83]提出了一种简单的测序重叠波峰图的解码方法DSD, 通过读取包含因突变(杂合突变和双等位突变)产生的重叠波峰, 利用通用的序列分析软件如DNASTAR/EditSeq等, 在野生型序列中查找出对应的碱基序列, 从而解码出两条等位序列。基于DSD方法, 研究团队进一步开发了网页版的在线工具DSDecode^[84], 用户只需要上传野生型的参考序列和测序峰图文件, 程序即可对文件直接解码, 输出最终的突变序列和类型。针对部分测序峰图因峰图质量无法自动解码的情况, 该程序还提供了一个半自动解码工具SaDS-Decode, 用户将人工读取的锚定序列和简并序列输入后, 程序会自动判别2个等位序列。升级版的DSDecodeM(<http://skl.scau.edu.cn/dsdecode/>)^[59], 可以同时解码多个测序文件。

此外, 也有类似的通过分析Sanger测序峰图检测突变的方法(表4)。通过峰图比较的方法, Brinkman等人^[85]开发了缺失序列追踪分析软件(tracking of indels by DE composition, TIDE), TIDE采用分解算法, 通过比较编辑样本和野生型对照样本的扩增子之间的Sanger测序峰图, 识别编辑样本中存在InDel的等位序列, 量化样本中主要插入和缺失类型的频率, 但对于替换或者插入类型的突变, 无法确定具体的序列。该方法适用于细胞系编辑后产生的突变分析和效率评估, 同时, 该团队进一步开发了用于分析插入/替换编辑的突变分析工具TIDER^[86], 评估目标细胞群中含有插入/替换的比例。与

表 4 Sanger测序突变分析工具的比较

Table 4 Comparison of Sanger sequencing mutation analysis tools

分析工具	优点	限制	参考文献
DSDecode/DSDecodeM	可快速对Sanger测序的峰图进行解码,并根据参考序列以及靶点序列检测靶点处的突变,DSDecodeM最多可同时处理20个样本的Sanger测序文件	无法分析嵌合突变或含有3种或以上突变类型的测序文件	[59,84]
TIDE	可有效检测样本中的样本中主要插入和缺失类型的频率	主要针对缺失类型的突变分析,无法分析大于1 nt以上插入的突变类型.需手动调整算法的参数对不同的样本进行检测	[85]
ICE	可对多种类型实验样本进行变异检测,不须对检测参数进行调整,可同时检测大量的样本	主要针对缺失类型的突变分析,插入突变仅能判定插入的碱基数(结果中使用N代替插入碱基)	[87]
EditR	分析速度快,根据荧光的百分比检测单碱基编辑的结果,准确性高	仅用于单碱基编辑的分析,受Sanger测序质量的影响较大	[88]
BEAT	可同时处理多批量的样本测序数据,快速生成多个高质量单碱基编辑结果图	仅用于单碱基编辑的突变结果分析,对于突变频率较低的检测不可靠	[89]

TIDE工具相似, ICE(Inference of CRISPR Edits)通过对比编辑样本与野生型样本的 Sanger 测序峰图检测样本中的突变类型, 但相对于TIDE, ICE可同时处理单基因编辑、多基因编辑、单碱基编辑以及同源重组的实验测序结果, 提高了基因编辑检测的便利性^[87]. 对于单碱基编辑的 Sanger 测序结果, EditR工具^[88]和BEAT(base editing analysis tools)工具^[89]可检测编辑窗口中4种碱基的荧光百分比, 从而判断单碱基突变率.

3.3 基于二代测序的突变分析工具

利用Sanger测序分析靶点突变适用于单个样本中靶点突变类型少于或等于2种的情况, 而对于嵌合突变. 例如大豆(虽然为二倍体, 但经常产生嵌合突变), 以及多倍体基因组的物种则存在明显的局限性. 同时, 单个样本Sanger测序的成本较高(12~18元/反应), 当检测的样本量较大时, 成本急剧增加. 随着以华大智造、illumina为主流的二代测序仪的通量增加、测序价格的下降(20~40元/Gb)以及测序周期的缩短, 利用二代测序的方法分析靶点突变可有效降低检测成本, 解决复杂突变类型的解码难题.

早期开发的利用二代测序进行突变序列分析的工具包括AGEseq、Cas-Analyzer、CRISPR-GA和CRISPResso等(表5). Xue和Tsai^[90]开发的AGEseq用于分析靶点或潜在脱靶位点的扩增子二代测序, 底层通过调用了BLAT比对软件从而分析突变序列, 软件支持本地界面版和基于Galaxy的分析. 此外, AGEseq也支持Sanger文件的比对, 但仅限于纯合突变类型, 不支持存在双

峰的测序文件. Cas-Analyzer、BE-Analyzer以及PE-Analyzer均为CRISPR RGEN Tools中的突变分析工具, 是基于JavaScript语言开发的网页工具^[62,63,91], 完全依赖于用户界面端处理二代测序数据从而分析基因编辑产生的突变序列和频率, 因此, 一定程度上规避了因数据上传导致的耗时问题, 但由于部分用户端计算资源有限而分析缓慢, 因此无法分析较大数据量的测序数据. CRISPResso是一款基于Linux/Mac系统的分析软件^[92], 虽然提供了网页版本, 但部分功能受限. CRISPResso可以自动地将靶点的测序文件比对到参考基因组, 然后分析突变的序列和频率, 并且提供了丰富的可视化结果. 在提供了外显子序列后, 软件可进一步判别突变对剪接点的影响. 升级版的CRISPResso2更新了分析流程, 添加了对单碱基编辑的突变分析, 以及当参考序列为杂合基因组的比对分析^[93].

以上工具在分析二代测序数据时, 虽然允许用户同时分析多个靶点, 即同个样本包含不同靶点的多个扩增子混合测序, 但如果是多个样本的混合扩增子建库时, 需要额外进行测序数据的拆分(部分测序公司可以提供拆分服务), 然后将拆分后的文件逐个进行分析, 因此在一定程度上限制了多样本检测情况下的使用, 且没有提供具体的多样本建库的方法.

Liu等人^[94]通过在扩增靶点序列时在每个样本的扩增子中引入标识碱基(barcode), 然后进行混合建库测序, 实现了多样本的高通量鉴定, 同时提供了分析工具Hi-TOM. 具体的检测步骤是在96孔板中进行PCR反应, 在第一轮特异扩增时, 在特异扩增引物的两端分别添

表 5 二代测序突变分析工具的比较

Table 5 Comparison of next-generation sequencing analysis tools

分析工具	优点	限制	参考文献
AGEseq	具有网页工具以及Window本地化软件工具, 便于使用, 同时可用于SNP分型、宏基因组分析等多方面的应用	本地版需要用户额外配置安装环境, 单次分析只能处理1个样本的测序文件, 多样本时必须由测序公司提前拆分	[90]
Cas-Analyzer	虽是网页版工具, 但依赖用户端进行样本的突变检测, 避免了数据上传造成的耗时	受限用户的计算资源配置, 无法处理数据量较大的测序文件, 且单次仅可分析单个样本的测序数据	[91]
BE-Analyzer	快速比较对照样品和编辑样品的在特定靶点处的单碱基突变	仅适用于单碱基编辑的突变分析, 单次分析仅处理单个样本的测序数据	[62]
PE-Analyzer	可选择特定突变类型, 对先导编辑样本进行突变检测	仅适用于PE编辑的突变分析, 网页版数据上传缓慢	[63]
CRISPResso2	可快速、准确地检测大批量的样本数据, 包括单碱基编辑数据, 结果分析丰富, 提供了本地版和网页版工具	网页版工具仅支持分析最多4个样本的数据, 本地版的安装依赖一定的生物信息学基础	[93]
Hi-TOM	可同时快速检测最多96个样本的突变, 可输出每个变异所覆盖的reads数目, 可靠性和灵敏度高	网页工具, 测序数据较大时, 文件上传时间较久	[94]
Hi-TOM 2.0	除了高通量检测基因编辑突变, 还可以用于低频检测, 增加了靶点特异区域突变检测以及甲基化的检测	网页工具, 测序数据较大时, 文件上传时间较久	[95]
BATCH-GE	使用简便, 可同时处理大量的样本数据	为Perl语言脚本工具, 限制部分对Perl语言不熟练的用户使用	[96]
CRIS.py	可用于ZFN、TALEN、Cas9等相关基因编辑工具的编辑结果, 短时间内可检测大量的样本测序数据	受靶点旁侧序列的影响, 为Python语言脚本, 限制部分对编程语言不熟练的用户使用	[97]
CrisprStitch	本地版软件, 不受操作平台限制, 对结果进行了较为丰富的可视化, 并统计各种类型的突变	软件不自带测序文件质量控制功能, 输入文件前需用户进行质量控制和过滤低质量reads	[98]

加了一段搭桥序列, 然后在接下来的第二轮扩增时, 分别在正向和反向引物中引入标记96孔板中行信息的barcode(共8行)和标记列信息的barcode(共12列), 方法提供了相应的barcode和引物序列。因此, 经过两轮扩增后, 每个样本靶点序列的扩增子在两端标记了唯一的barcode对, 用以区分不同样本。每个测序文库最高混合样本数为96个。将扩增子混合后进行建库测序(必需使用双端测序), 在Hi-TOM网页中上传参考序列和双端测序文件后, 软件会自动输出每个样本对应的突变序列和每个等位点覆盖的reads数。通过混合建库的方法, 检测成本可降至约1元/靶点, 尤其适合大量样本的突变鉴定, 以及细胞系、愈伤和多倍体编辑的突变检测。在Hi-TOM 2.0工具中, 添加的引物设计模块为样品突变库的构建提供了便利, 同时添加了位置特异性的突变检测以及甲基化的变异检测, 极大地提高了样本编辑检测的丰富度^[95]。类似的工具还包括BATCH-GE^[96]、CRIS.py^[97]以及CrisprStitch^[98], 用于基于扩增子高通量测序进行多种突变类型的鉴定和编辑效率的

自动化分析, 其中CrisprStitch工具提供了简便的图形操作界面, 可在本地计算机上快速地进行批量化的高通量数据分析。

4 脱靶分析方法

尽管CRISPR/Cas系统为科学研究、疾病治疗、作物品种改良等方面提供了一个便捷高效的基因编辑工具, 然而, 研究表明, 使用CRISPR/Cas系统进行基因编辑时, 会出现对与靶位点具有一定相似度的序列进行切割, 产生脱靶(off-target)现象^[99]。

目前, 在常用的CRISPR/Cas9基因编辑系统, 包括SpCas9和SaCas9, 均发现存在一定概率的脱靶现象^[100,101]。在CRISPR/Cas12a即Cpf1系统中, 常用的核酸酶为AsCpf1、LbCpf1和FnCpf1系统, Kleinstiver等人^[102]利用多种脱靶检测的方法对比了Cas9系统和Cpf1系统的脱靶率, 结果显示Cpf1系统的脱靶率低于Cas9系统的脱靶率, 说明不同的Cas系统在编辑特异性方面存在差异。

4.1 脱靶的原因

Cas核酸酶的脱靶率与靶点序列和sgRNA的选择密切相关,对于SpCas9,当sgRNA与基因组序列有3个碱基错配时,SpCas9仍具有一定的编辑活性^[103].对于LbCpf1和AsCpf1,当靶点序列与sgRNA有一个碱基不匹配,且错配的碱基位于靶点序列的第3、4、5和17位时,LbCpf1和AsCpf1的编辑效率显著下降,而当错配的碱基位于靶点序列的第21位,LbCpf1和AsCpf1的编辑效率与未错配序列的编辑效率相当,当错配的碱基位于靶点序列的最后一位时,其编辑效率甚至高于未错配序列的编辑效率,在第21和23位两个碱基同时错配的情况下,LbCpf1和AsCpf1相对于未错配的序列仍保留有80%的编辑活性^[104].另外,sgRNA的长度也会对Cas核酸酶的脱靶率造成影响,对于SpCas9,常设计的靶点序列为20 nt,而使用17 nt或者18 nt靶点序列的sgRNA,能够显著减少SpCas9进行基因编辑的脱靶率^[105].

同时,Cas核酸酶的脱靶率与Cas蛋白有关,Kleistner等人^[106]通过开发高保真的SpCas9,即SpCas9-HF1,可显著降低SpCas9的脱靶率;Kulcsár等人^[107]对SpCas9进行突变,开发了一套SpCas9的变体,即CRISPR-Recise,提高了SpCas9的保真程度,并降低其脱靶率.因此,通过蛋白进化等方法有助于提升Cas蛋白的特异性,降低脱靶率.

在单碱基编辑器中,存在着两种脱靶情况:其一,与Cas核酸酶靶向DNA的脱靶原理相同,由于基因组中靶序列与非靶序列的相似性较高,依赖于Cas核酸酶的碱基编辑器产生非目标编辑;其二,胞苷脱氨酶的广谱脱氨活性,具有同时作用于DNA和RNA的能力,导致非靶向性结合从而引发脱靶事件^[108].研究表明,CBE系统可在人类细胞中编辑时会产生严重的DNA和RNA脱靶效应,引起DNA和RNA链上的单核苷酸突变(single nucleotide variations, SNVs)^[109,110].与CBE不同,自然界没有一种能够作用于DNA的腺嘌呤脱氨酶,Gaudelli等人利用多重抗生素介导的连续进化获得了一种腺苷脱氨酶,它拥有极高的DNA靶向性,有效地降低了第二种类型的脱靶^[111].Zhou等人^[112]对ABE以及CBE系统进行一系列的脱靶检测发现,与未处理的细胞相比,ABE处理后细胞会产生5~10倍的SNVs,CBE处理后的细胞则会产生5~40倍的SNVs.因此,如何降低单碱基编辑工具的脱靶率是未来进行科学研究以及疾病治疗所要解决的问题.而对CBE以及ABE中的脱氨酶进行改造,

可有效降低脱氨酶引起的脱靶效应,如目前已改造的变体SECURE-CEB^[109]、YE1-BE4^[113]、SECURE-AEB^[110]、ABE^{F148A}^[112]等,可在不影响DNA靶向编辑的效率下,显著降低脱靶率.

4.2 脱靶的危害

Zuo等人^[114]利用GOTI技术对CRISPR/Cas9系统进行脱靶率的检测,结果表明设计良好的CRISPR/Cas9系统并不具有明显的脱靶效应;而利用GOTI技术对CRISPR/Cas9系统衍生的BE3和ABE单碱基编辑系统的脱靶效率进行检测的时,发现ABE系统未出现明显脱靶效应,但BE3单碱基编辑系统则产生非常严重的脱靶效应.脱靶突变大部分是由G突变成A或者由C突变成T,表明这些突变缺失是由BE3所导致,并且大部分突变位于基因编码区,甚至位于原癌基因和抑癌基因中,即这些脱靶产生的突变可能会导致癌症的发生.在未来临床应用中,应特别注意基因编辑脱靶所带来的风险.

在植物中,Jin等人^[115]利用全基因组测序统计单碱基编辑系统转化的56株T₀代水稻材料和21株对照材料发现,在有无sgRNA的情况下,BE3系统均可在水稻基因组中产生大量的SNVs,大部分突变位于基因的转录区,且产生脱靶的位点不是现有软件可预测的位点.植物中虽然可以通过后代分离的方法筛选掉脱靶位点,但为了避免脱靶导致育种周期延长,特别是对于依赖于无性繁殖的部分植物,在实验设计时,需要借助辅助设计工具尽可能降低脱靶风险.

4.3 脱靶检测技术

对于CRISPR/Cas系统的脱靶现象,目前研究人员开发了一系列的脱靶检测技术(表6).根据各种脱靶检测技术的原理,脱靶检测技术主要可分为细胞内脱靶检测技术以及细胞外脱靶检测技术两大类型.

4.3.1 细胞内脱靶检测技术

细胞内脱靶检测方法主要将基因编辑工具转入细胞内,并对细胞内的编辑情况进行检测,可有效地反映基因编辑工具在细胞内基因组的编辑情况.其中,GUIDE-seq技术是目前在细胞和动物研究中,使用频率最高的细胞内脱靶检测技术之一,该方法依赖细胞NHEJ的DNA修复机制对脱靶位点进行检测,当Cas蛋白对核酸序列进行切割造成DNA双链断裂时,利用DNA的NHEJ修复机制,将一段双链寡核苷酸序列(dou-

表6 脱靶检测技术特点

Table 6 Characteristics of off-target detection methods

方法	特点	限制	参考文献
WGS	可准确检测全基因组水平上的脱靶情况, 检测出各种变异类型	测序成本较高, 且只能检测脱靶率较高的位点, 难以检测大量样品中脱靶率较低的位点	[125]
GUIDE-seq	成本低, 敏感性高, 能准确检测潜在的脱靶位点, 假阳性率低	依赖于dsODN的插入, 未整合dsODN的位点无法检测	[116]
DISCOVER-seq	灵敏度高, 能准确检测脱靶位点的体内检测技术	只能捕获MRE11结合的正在修复的位点, 未结合MRE11蛋白的位点无法检测, 且需进行ChIP-seq实验, 成本较高	[117]
PEAC-seq	可准确检测Cas蛋白在细胞内的脱靶情况, 且能鉴定染色体易位	标签的插入效率受pegRNA以及脱靶位点的影响	[118]
VIVO	灵敏度高, 准确性高	只针对CIRCLE-seq所检测的脱靶位点	[119]
GOTI	可有效检测潜在脱靶位点, 且能检测单碱基编辑系统的脱靶效应	需进行流式细胞分选技术分选出被编辑的细胞以及未被编辑的细胞, 存在一定假阳性	[114]
GOAT	能有效检测CRISPR系统以及单碱基编辑系统的脱靶效应, 且不需要通过流式细胞分选技术对细胞进行分选	需进行全基因组水平上的测序, 成本较高	[120]
SURRO-seq	可同时评估超过10000个脱靶位点, 且能同时评估细胞中数百个Cas蛋白的脱靶情况	只评估预测的脱靶位点, 寡核苷酸链的合成错误会导致脱靶位点数据的丢失	[121]
SITE-seq	测序深度要求低, 背景噪音小	脱靶位点检测数量受Cas9/sgRNA复合物的浓度以及处理时间影响	[122]
Digenome-seq	灵敏度高, 能检测0.1%甚至更低频率的脱靶位点	需进行全基因组测序, 且需要较高的测序深度, 成本较高	[123]
CIRCLE-seq	测序深度要求低, 背景噪音小, 能检测频率小于0.1%的脱靶位点	假阳性率高, 需进行下游实验验证	[124]

ble-stranded oligodeoxynucleotide, dsODN)整合到双链断裂处, 提取细胞基因组DNA后随机打断基因组序列, 在打断后的DNA片段两头连接接头序列, 利用已知的dsODN序列以及接头序列, 对片段进行文库的构建以及二代测序, 即可检测核酸酶的切割位点, 从而检验该核酸酶的脱靶率^[116]。GUIDE-seq检测技术简便, 可信度较高, 但须在靶点切割后, 连接了dsODN的靶点序列方可被检测, 因此GUIDE-seq检测技术所检测的脱靶位点一般较少。目前, 已有许多研究利用GUIDE-seq技术检测不同类型的Cas蛋白编辑特异性, 如Shi等人^[57]利用GUIDE-seq方法, 检测*Alicyclobacillus tengchongensis*菌株中鉴定的AtCas9蛋白的脱靶率, 表明AtCas9为低脱靶性的核酸酶。DISCOVER-seq是一种细胞内蛋白介导的脱靶检测技术。该技术通过追踪DSB的识别因子MRE11来判断Cas酶的切割位点, 通过ChIP-seq实验, 获取与MRE11结合的相关DNA片段, 即可检测Cas核酸酶切割位点, 从而获得脱靶位点^[117](图2)。DISCOVER-seq可有效地检测Cas核酸酶在全基因组中的脱靶位点, 但由于需进行ChIP-seq实验, 使用该方法进行检测难度较大, 成本较高。PEAC-seq技术是基于先导

编辑系统所改造的脱靶检测技术, Yu等人^[118]将pegRNA上的反转录模板替换为一段标签序列, 并将Cas9n替换为可引起双链断裂的Cas9, 使得被编辑的基因组序列带有标签序列, 利用Tn5转座酶处理后进行PCR富集以及二代测序, 从而检测Cas9在全基因组中脱靶情况。然而, 不论是dsODN插入DSB, 还是MRE11识别的DSB, 都可能多种因素导致, 因此, GUIDE-seq、DISCOVER-seq以及PEAC-seq检出的脱靶位点也可能存在一定的假阳性, 需要通过DSB所在序列与sgRNA相似度进一步确认。

以上两种方法主要是基于检测全基因组的DSB位点, 从而推断脱靶位点。此外, 还有部分方法是先通过预测全基因组的潜在脱靶位点, 然后通过改进的扩增子测序、全基因组测序或探针法进行脱靶位点的检测, 包括VIVO、GOTI、GOAT、SURRO-seq等。例如, VIVO先通过CIRCLE-seq方法在体外检测Cas核酸酶在基因组上潜在的脱靶位点, 随后利用核酸酶处理靶向组织, 并对体外检测出的位点进行扩增测序, 在体内验证评估体外检测的位点是否存在突变^[119]。GOTI是在全基因组测序前, 先通过红色荧光蛋白对被编辑的受精



图 2 DISCOVER-seq脱靶检测原理
Figure 2 The principle of DISCOVER-seq

卵其中一个卵裂球进行标记，然后利用流式细胞分选技术分选出被编辑的细胞以及未被编辑的细胞，再对两组细胞分别进行全基因组测序，通过比较两组的变异点差异，即可检测Cas核酸酶编辑时的脱靶位点^[114]。GOAT是基于GOTI技术改良的一种脱靶检测技术，与GOTI不同的是，使用GOAT技术对卵裂球进行编辑时，需将两个卵裂球分开，只对其中一个卵裂球进行编辑，然后通过全基因组测序比对编辑以及未编辑胚胎的差异，从而检测脱靶位点，研究人员使用GOAT技术对LbCas12a以及AsCas12a两个基因编辑工具进行检测，结果表明Cas12a蛋白脱靶率较低^[120]。SURRO-seq是一种大规模且精确的脱靶分析方法，通过各种预测工具预测Cas核酸酶潜在脱靶位点后，基于预测的脱靶位点合成寡核苷酸链并克隆到病毒载体中，将病毒载体文库分别转入到含有Cas9的细胞以及不含有Cas9的细胞中，富集转入病毒载体的细胞后，再通过二代测序分析Cas核酸酶编辑情况，检测脱靶率^[121]。研究人员通过将SURRO-seq与T7E1、GUIDE-seq和CIRCLE-seq三个脱靶检测方法进行对比，结果显示SURRO-seq可检测出T7E1、GUIDE-seq和CIRCLE-seq方法检测的大部分脱靶位点，表明SURRO-seq是细胞内脱靶检测有效的评估技术。

4.3.2 细胞外脱靶检测技术

与细胞内脱靶检测技术不同, 细胞外检测技术需先将细胞中的DNA提取, 再使用基因编辑工具对所得核酸进行编辑, 从而检测Cas蛋白的编辑位点, 与细胞内的脱靶检测位点相比, 细胞外检测技术更为便捷, 快速, 但由于提取DNA时需去除各种蛋白质, 因此细胞外脱靶检测技术不能准确地反映出基因编辑工具在细胞的真实编辑情况。常用的细胞外检测脱靶技术主要包括SITE-seq^[122]、Digenome-seq^[123]、CIRCLE-seq^[124]等。其中, Digenome-seq在检测技术是在对基因组DNA进行核酸酶处理之前, 对基因组DNA进行打断和在片段两头接上接头序列, 利用Cas9对接上接头的片段进行酶切后再通过二代测序技术检测切割位点^[125]。Digenome-seq脱靶检测技术最大的优点是其检测脱靶位点

具有高敏感性, 能够检测到0.1%甚至更低频率的脱靶位点, 是一种稳定、敏感和重复性高的脱靶检测技术.

CIRCLE-seq是一种快速、简便、高灵敏度的体外脱靶检测技术, 可在全基因组范围内检测Cas9的脱靶位点。首先将基因组DNA随机打断, 并在片段的两端接上接头序列, 然后进行片段的酶切和连接, 使得片段环化, 接着用核酸酶消化降解未环化的基因组片段, 使用Cas9蛋白对环化后的片段进行酶切。此时, 被切割的片段重新线性化, 未切割的片段仍保持环状的结构, 对线性化的片段加上接头序列, 最后通过二代测序, 检测Cas9在全基因组范围内的切割位点以及脱靶情况^[124]。为了验证CIRCLE-seq检测技术的灵敏性, 研究人员对*HBB*基因进行靶向切割以及脱靶的检测, 同时与Digenome-seq脱靶检测技术进行对比。结果表明, CIRCLE-seq可鉴定出Digenome-seq已检测出29个脱靶位点中的26个, 同时还检测出156个新的脱靶位点, 表明CIRCLE-seq具有更高的脱靶检测灵敏度。

4.4 预测脱靶位点的软件工具

目前形式多样的实验系统已被运用于研究CRISPR/Cas系统的脱靶效应, 也有许多不同的算法和软件平台已被设计应用于脱靶效应的预测. 大多数靶点设计平台都包含有脱靶预测的功能或模块, 如RGEN、CRISPR-GE等. Stemmer等人^[65]设计开发了CCTop的CRISPR/Cas9在线预测工具, 其提供了一个直观的用户界面. 根据一段给定的DNA序列, CCTop可以根据候选的sgRNA靶位点的脱靶质量, 识别并显示出所有的候选sgRNA靶位点. Cas-OFFinder是一个通用的脱靶位点预测工具, 可以根据用户给定的DNA序列预测基因组中的脱靶位点. Cas-OFFinder是RGEN的脱靶预测模块, 与其他脱靶位点预测工具相比, 其可在不限制PAM序列以及错配碱基数量的情况下快速预测潜在的脱靶位点^[126]. CRISPR-GE是一个集合多种基因编辑设计工具的网页平台, 其offTarget模块可用于快速预测目的物种基因组中的脱靶位点^[59]. Stortz等人^[127]使用深度学习的方法, 对CRISPR/Cas系统的大型脱靶切割

数据库crisprSQL中的数据集进行训练,构建模型,从而准确预测潜在的脱靶位点。尽管已有大量计算机辅助工具可快速预测在目的基因组上潜在的脱靶位点,但这些预测方法通常是基于sgRNA序列进行计算,无法充分考虑细胞核中染色质可及性等方面的环境因素,因此通过计算机辅助工具预测脱靶位点还需进一步的实验验证。

5 总结与展望

CRISPR/Cas基因编辑技术已成为生命科学领域的革命性技术,被广泛用于基因功能研究、基因治疗和遗传改良等。近十年来,基因编辑系统发展呈现多样化,开发了多种类型的基因编辑设计和分析的辅助工具,极大地帮助了科研人员利用基因编辑技术开展科学研究,特别是基因编辑靶点设计和编辑位点突变检测相关工具。然而,由于基因编辑技术在细胞内编辑的不确定性以及编辑需求的差异,相关工具仍需进一步发展和完善。

由于细胞活动的复杂性,Cas酶和sgRNA复合体在细胞内行使切割功能的效率与细胞状态、染色体开放程度以及DNA表观修饰都存在相关性^[53]。例如,Bai等基于DNA靶序列处的染色质开放性对基因编辑效率具有显著影响,通过共表达一个促进染色质开放的人源RNA m6A去甲基化酶(hFTO),可以显著提高大豆和水稻中各种基因编辑工具的效率^[128]。然而,目前评估切割效率主要以细胞系中开展的大量靶点编辑产生的结果数据为模型,而缺乏基于体内不同细胞类型的大规模测试数据开发的预测模型。随着单细胞测序、多种细胞的基因编辑数据的日益增加,结合深度学习等方法,有望提高基因编辑在体内切割效率和特异性评估的可靠性。此外,在脱靶评估方面,目前的研究普遍认为在潜在脱靶序列和目标靶向序列的PAM和靠近PAM区序列存在的差异越多,脱靶的可能性越低,反之在距离PAM更远的碱基存在少量的差异(1~2 nt),脱靶的可能性较高^[129]。然而,目前用于编辑效率和脱靶评估的打分矩阵主要是基于CRISPR/Cas9和CRISPR/Cas12a系统的基因编辑数据,而对于其他类型的CRISPR/Cas系统(如Cas12f、Cas13等)、编辑衍生体系(如PE编辑),它们的Cas蛋白和sgRNA复合体识别靶点序列的特性可能存在差异,因此,开发这类编辑系的辅助设计工具需要更多的实验数据支撑。利用新型Cas酶进行靶点设

计时,需要特别注意现有靶点设计工具的评估模型是否适用。其次,针对于高度重复的复杂基因组,例如异源多倍体植物,其重复序列多,等位基因间存在复杂的拷贝形式,重复序列间仅有少量变异,如果目的是特异性地编辑其中一种拷贝,对其进行靶点设计时需要谨慎考虑脱靶的可能性。因此,未来在需要使用基因编辑技术大规模开展疾病治疗或植物遗传改良时,需要针对特定细胞或植物开展前期的基因编辑效率和特异性分析和建模,以提高编辑的可靠性和效率。

虽然现有的基因编辑突变检测方法能够满足多数靶点的检测需求,但随着多种编辑策略和编辑器的出现,目前基于Sanger测序或是二代混合建库测序的方法仍存在一定的提升空间。例如,在多倍体(如八倍体甘蔗品种)中开展基因编辑时,现有检测方法仅能做到是多少个等位基因发生了编辑,但难以区分突变是发生在哪些染色体上的等位基因。其次,当在特定区域内存在连续多个靶点的同时编辑时,例如启动子平铺删除,由于不同靶点间均存在片段删除的可能,将在一定程度上增加检测的难度,需要根据实际的编辑目的,利用多种检测方法筛选需要的特定突变个体。最后,需要注意的是,在植物中基于农杆菌介导的遗传转化递送基因编辑载体时,虽然多数突变是发生在愈伤侵染阶段,通过检测T₀代植株可以判断植株的靶点突变序列,然而,如果没有分离去除Cas基因和sgRNA,它们在后代的植株中还会发生一定几率的编辑作用,因此,获得编辑植株后,建议使用去除了Cas基因和sgRNA成分的后代作为稳定植株,用以后续的表型和其他研究实验。此外,利用农杆菌介导植物基因编辑敲除过程中,在靶位点产生DSB后,时常会出现插入载体的整个T-DNA区的情况^[130],如果在T₀代靶点两侧设计引物进行扩增,会出现无法扩增出条带的情况或者是因仅能扩增出其中一个等位突变而判断为纯合突变,因此需要通过后代检测进一步进行判断。

随着基因编辑技术在基因治疗、生物育种等多个领域内的广泛应用,开发从基因选择、编辑位点选择、性状预测、靶点分析和评估以及高通量突变检测为一体化的综合设计平台,有助于进一步提高利用基因编辑开展相关研究或应用的效率、可靠性,节约成本。从基础研究、编辑技术开发、应用研究、信息技术等领域内的深度交叉融合,对于建成智慧化的基因编辑工具平台具有重要意义。

参考文献

- 1 Horvath P, Barrangou R. CRISPR/Cas, the immune system of Bacteria and Archaea. *Science*, 2010, 327: 167–170
- 2 Fineran P C, Charpentier E. Memory of viral infections by CRISPR-Cas adaptive immune systems: Acquisition of new information. *Virology*, 2012, 434: 202–209
- 3 Sander J D, Joung J K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*, 2014, 32: 347–355
- 4 Wright A V, Nuñez J K, Doudna J A. Biology and applications of CRISPR systems: Harnessing nature's toolbox for genome engineering. *Cell*, 2016, 164: 29–44
- 5 Zhang Y L, Wang X H, Li G S, et al. Research advances in novel DNA base editors (in Chinese). *J South China Agricult Univ*, 2022, 43: 1–16 [张雅玲, 王铮和, 李构思, 等. 新型DNA碱基编辑器的研究进展. 华南农业大学学报, 2022, 43: 1–16]
- 6 Lai Z S Y, Huang Z T, Sun J T, et al. The recent progress of CRISPR/Cas genome editing technology and its application in crop improvement (in Chinese). *Chin Sci Bull*, 2022, 67: 1923–1937 [赖郑诗雨, 黄赞唐, 孙洁婷, 等. CRISPR/Cas基因组编辑技术及其在农作物品种改良中的应用. 科学通报, 2022, 67: 1923–1937]
- 7 Yan L, Zhang J S, Zhu J K, et al. Genome editing and its applications in crop improvement (in Chinese). *J Agricult Sci Technol*, 2022, 24: 78–89 [闫磊, 张金山, 朱健康, 等. 基因编辑技术及其在农作物中的应用进展. 中国农业科技导报, 2022, 24: 78–89]
- 8 Ren J, Cao Y X, Huang Y, et al. Development and application of genome editing technology in rice (in Chinese). *China Rice*, 2021, 27: 92–100 [任俊, 曹跃炫, 黄勇, 等. 基因编辑技术及其水稻中的发展和应用. 中国稻米, 2021, 27: 92–100]
- 9 Wang C, Wang K J. Advances in CRISPR-Cas-mediated genome editing system in plants (in Chinese). *Chin J Biotech*, 2017, 33: 1712–1722 [王春, 王克剑. CRISPR-Cas系统在植物基因组编辑中的研究进展. 生物工程学报, 2017, 33: 1712–1722]
- 10 Bibikova M, Golic M, Golic K G, et al. Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics*, 2002, 161: 1169–1175
- 11 Urnov F D, Rebar E J, Holmes M C, et al. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*, 2010, 11: 636–646
- 12 Li T, Huang S, Zhao X, et al. Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res*, 2011, 39: 6315–6325
- 13 Cong L, Zhou R, Kuo Y C, et al. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun*, 2012, 3: 968
- 14 Ishino Y, Shinagawa H, Makino K, et al. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol*, 1987, 169: 5429–5433
- 15 Bhaya D, Davison M, Barrangou R. CRISPR-Cas systems in Bacteria and Archaea: Versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet*, 2011, 45: 273–297
- 16 Jansen R, Embden J D A, Gaastra W, et al. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*, 2002, 43: 1565–1575
- 17 Koonin E V, Makarova K S, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol*, 2017, 37: 67–78
- 18 Shmakov S, Smargon A, Scott D, et al. Diversity and evolution of class 2 CRISPR–Cas systems. *Nat Rev Microbiol*, 2017, 15: 169–182
- 19 Makarova K S, Wolf Y I, Alkhnbashi O S, et al. An updated evolutionary classification of CRISPR–Cas systems. *Nat Rev Microbiol*, 2015, 13: 722–736
- 20 Altae-Tran H, Kannan S, Suberski A J, et al. Uncovering the functional diversity of rare CRISPR-Cas systems with deep terascale clustering. *Science*, 2023, 382: eadi1910
- 21 Cong L, Ran F A, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 2013, 339: 819–823
- 22 Zetsche B, Gootenberg J S, Abudayyeh O O, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, 2015, 163: 759–771
- 23 Tsuchida C A, Zhang S, Doost M S, et al. Chimeric CRISPR-CasX enzymes and guide RNAs for improved genome editing activity. *Mol Cell*, 2022, 82: 1199–1209.e6
- 24 Zong Y, Wang Y, Li C, et al. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat Biotechnol*, 2017, 35: 438–440
- 25 Li C, Zong Y, Wang Y, et al. Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol*, 2018, 19: 59
- 26 Tan J, Zeng D, Zhao Y, et al. PhieABEs: A PAM-less/free high-efficiency adenine base editor toolbox with wide target scope in plants. *Plant Biotechnol J*, 2022, 20: 934–943
- 27 Chen L, Park J E, Paa P, et al. Programmable C:G to G:C genome editing with CRISPR-Cas9-directed base excision repair proteins. *Nat*

- [Commun](#), 2021, 12: 1384
- 28 Li C, Zhang R, Meng X, et al. Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. [Nat Biotechnol](#), 2020, 38: 875–882
 - 29 Lin Q, Zong Y, Xue C, et al. Prime genome editing in rice and wheat. [Nat Biotechnol](#), 2020, 38: 582–585
 - 30 Lu Y, Tian Y, Shen R, et al. Targeted, efficient sequence insertion and replacement in rice. [Nat Biotechnol](#), 2020, 38: 1402–1407
 - 31 Pan C, Sretenovic S, Qi Y. CRISPR/dCas-mediated transcriptional and epigenetic regulation in plants. [Curr Opin Plant Biol](#), 2021, 60: 101980
 - 32 Lei J, Dai P, Li J, et al. Tissue-specific CRISPR/Cas9 system of cotton pollen with *GhPLIMP2b* and *GhMYB24* Promoters. [J Plant Biol](#), 2020, 64: 13–21
 - 33 Kim J H, Rege M, Valeri J, et al. LADL: Light-activated dynamic looping for endogenous gene expression control. [Nat Methods](#), 2019, 16: 633–639
 - 34 Al-Shayeb B, Sachdeva R, Chen L X, et al. Clades of huge phages from across Earth's ecosystems. [Nature](#), 2020, 578: 425–431
 - 35 Doudna J A, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. [Science](#), 2014, 346: 1258096
 - 36 Shan Q, Wang Y, Li J, et al. Targeted genome modification of crop plants using a CRISPR-Cas system. [Nat Biotechnol](#), 2013, 31: 686–688
 - 37 Zhang R X, Li B B, Yang Z G, et al. Dissecting plant gene functions using CRISPR toolsets for crop improvement. [J Agric Food Chem](#), 2022, 70: 7343–7359
 - 38 Altae-Tran H, Kannan S, Demircioglu F E, et al. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. [Science](#), 2021, 374: 57–65
 - 39 Chai N, Liu Y X, Zhang R X, et al. Development of genome editing tools: From CRISPR/Cas9 to TnpB (in Chinese). [Genomics Appl Biol](#), 2023, 42: 1267–127 [柴楠, 刘雨馨, 张瑞祥, 等. 基因组编辑工具的发展: 从CRISPR/Cas9到TnpB. [基因组学与应用生物学](#), 2023, 42: 1267–127]
 - 40 Liu Q, Zhao C, Sun K, et al. Engineered biocontainable RNA virus vectors for non-transgenic genome editing across crop species and genotypes. [Mol Plant](#), 2023, 16: 616–631
 - 41 Zhang Y, Ren Q, Tang X, et al. Expanding the scope of plant genome engineering with Cas12a orthologs and highly multiplexable editing systems. [Nat Commun](#), 2021, 12: 1944
 - 42 Yan W X, Hunnewell P, Alfonse L E, et al. Functionally diverse type V CRISPR-Cas systems. [Science](#), 2019, 363: 88–91
 - 43 Makarova K S, Wolf Y I, Iranzo J, et al. Evolutionary classification of CRISPR–Cas systems: A burst of class 2 and derived variants. [Nat Rev Microbiol](#), 2020, 18: 67–83
 - 44 Chen W, Ma J, Wu Z, et al. Cas12n nucleases, early evolutionary intermediates of type V CRISPR, comprise a distinct family of miniature genome editors. [Mol Cell](#), 2023, 83: 2768–2780.e6
 - 45 Xu C, Zhou Y, Xiao Q, et al. Programmable RNA editing with compact CRISPR–Cas13 systems from uncultivated microbes. [Nat Methods](#), 2021, 18: 499–506
 - 46 Abudayyeh O O, Gootenberg J S, Essletzbichler P, et al. RNA targeting with CRISPR–Cas13. [Nature](#), 2017, 550: 280–284
 - 47 Xiao Q, Xu Z, Xue Y, et al. Rescue of autosomal dominant hearing loss by *in vivo* delivery of mini dCas13X-derived RNA base editor. [Sci Transl Med](#), 2022, 14: eabn0449
 - 48 Tong H, Huang J, Xiao Q, et al. High-fidelity Cas13 variants for targeted RNA degradation with minimal collateral effects. [Nat Biotechnol](#), 2023, 41: 108–119
 - 49 Csörgő B, León L M, Chau-Ly I J, et al. A compact Cascade–Cas3 system for targeted genome engineering. [Nat Methods](#), 2020, 17: 1183–1190
 - 50 Pickar-Oliver A, Black J B, Lewis M M, et al. Targeted transcriptional modulation with type I CRISPR–Cas systems in human cells. [Nat Biotechnol](#), 2019, 37: 1493–1501
 - 51 Young J K, Gasior S L, Jones S, et al. The repurposing of type I-E CRISPR-Cascade for gene activation in plants. [Commun Biol](#), 2019, 2: 383
 - 52 Osakabe K, Wada N, Murakami E, et al. Genome editing in mammalian cells using the CRISPR type I-D nuclease. [Nucleic Acids Res](#), 2021, 49: 6347–6363
 - 53 Zhu C, Bortesi L, Baysal C, et al. Characteristics of genome editing mutations in cereal crops. [Trends Plant Sci](#), 2017, 22: 38–52
 - 54 Ma X, Zhang Q, Zhu Q, et al. A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. [Mol Plant](#), 2015, 8: 1274–1284
 - 55 Ren X, Yang Z, Xu J, et al. Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in *Drosophila*. [Cell Rep](#), 2014, 9: 1151–1162
 - 56 Xu H, Xiao T, Chen C H, et al. Sequence determinants of improved CRISPR sgRNA design. [Genome Res](#), 2015, 25: 1147–1157
 - 57 Shi Y J, Duan M, Ding J M, et al. DNA topology regulates PAM-Cas9 interaction and DNA unwinding to enable near-PAMless cleavage by thermophilic Cas9. [Mol Cell](#), 2022, 82: 4160–4175.e6
 - 58 Park J, Bae S, Kim J S. Cas-Designer: A web-based tool for choice of CRISPR-Cas9 target sites. [Bioinformatics](#), 2015, 31: 4014–4016
 - 59 Xie X, Ma X, Zhu Q, et al. CRISPR-GE: A convenient software toolkit for CRISPR-based genome editing. [Mol Plant](#), 2017, 10: 1246–1249

- 60 Haeussler M, Schönig K, Eckert H, et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol*, 2016, 17: 148
- 61 Montague T G, Cruz J M, Gagnon J A, et al. CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res*, 2014, 42: W401–W407
- 62 Hwang G H, Bae S. Web-based base editing toolkits: BE-Designer and BE-Analyzer. *Methods Mol Biol*, 2021, 2189: 81–88
- 63 Hwang G H, Jeong Y K, Habib O, et al. PE-Designer and PE-Analyzer: Web-based design and analysis tools for CRISPR prime editing. *Nucleic Acids Res*, 2021, 49: W499–W504
- 64 Liu H, Ding Y, Zhou Y, et al. CRISPR-P 2.0: An improved CRISPR-Cas9 tool for genome editing in plants. *Mol Plant*, 2017, 10: 530–532
- 65 Stemmer M, Thumberger T, del Sol Keyer M, et al. CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS ONE*, 2015, 10: e0124633
- 66 Moreno-Mateos M A, Vejnar C E, Beaudoin J D, et al. CRISPRscan: Designing highly efficient sgRNAs for CRISPR-Cas9 targeting *in vivo*. *Nat Methods*, 2015, 12: 982–988
- 67 Xu X, Liu M J. Recent advances and applications of base editing systems (in Chinese). *Chin J Biotech*, 2021, 37: 2307–2321 [徐鑫, 刘明军. 碱基编辑系统研究最新进展及应用. *生物工程学报*, 2021, 37: 2307–2321]
- 68 Arbab M, Shen M W, Mok B, et al. Determinants of base editing outcomes from target library analysis and machine learning. *Cell*, 2020, 182: 463–480.e30
- 69 Song M, Kim H K, Lee S, et al. Sequence-specific prediction of the efficiencies of adenine and cytosine base editors. *Nat Biotechnol*, 2020, 38: 1037–1043
- 70 Marquart K F, Allam A, Janjuha S, et al. Predicting base editing outcomes with an attention-based deep learning algorithm trained on high-throughput target library screens. *Nat Commun*, 2021, 12: 5114
- 71 Xie X, Li F, Tan X, et al. BEdtarget: A versatile web-based tool to design guide RNAs for base editing in plants. *Comput Struct Biotechnol J*, 2022, 20: 4009–4014
- 72 Wu Y, He Y, Sretenovic S, et al. CRISPR-BETS: A base-editing design tool for generating stop codons. *Plant Biotechnol J*, 2022, 20: 499–510
- 73 Pallasen A, Peets E M, Koeppl J, et al. Predicting base editing outcomes using position-specific sequence determinants. *Nucleic Acids Res*, 2022, 50: 3551–3564
- 74 Lin Q P, Zhu X L, Ma L S, et al. Recent advances in prime editing system (in Chinese). *J South China Agric Univ*, 2023, 45: 159–171 [林秋鹏, 朱秀丽, 马琳莎, 等. 引导编辑系统研究进展. *华南农业大学学报*, 2023, 45: 159–171]
- 75 Siegner S M, Karasu M E, Schröder M S, et al. PnB Designer: A web application to design prime and base editor guide RNAs for animals and plants. *BMC BioInf*, 2021, 22: 101
- 76 Lin Q, Jin S, Zong Y, et al. High-efficiency prime editing with optimized, paired pegRNAs in plants. *Nat Biotechnol*, 2021, 39: 923–927
- 77 Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012, 337: 816–821
- 78 Hsu P D, Lander E S, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 2014, 157: 1262–1278
- 79 Lloyd A, Plaisier C L, Carroll D, et al. Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. *Proc Natl Acad Sci USA*, 2005, 102: 2232–2237
- 80 Voytas D F. Plant genome engineering with sequence-specific nucleases. *Annu Rev Plant Biol*, 2013, 64: 327–350
- 81 Guschin D Y, Waite A J, Katibah G E, et al. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol*, 2010, 649: 247–256
- 82 Li X, Li R, Yuan Z, et al. One versatile Cas9-integrated single-tube duplex quantitative real-time PCR system for rapid analysis of CRISPR/Cas-induced mutants. *Anal Chem*, 2022, 94: 10832–10840
- 83 Ma X, Chen L, Zhu Q, et al. Rapid decoding of sequence-specific nuclease-induced heterozygous and biallelic mutations by direct sequencing of PCR products. *Mol Plant*, 2015, 8: 1285–1287
- 84 Liu W, Xie X, Ma X, et al. DSDecode: A web-based tool for decoding of sequencing chromatograms for genotyping of targeted mutations. *Mol Plant*, 2015, 8: 1431–1433
- 85 Brinkman E K, Chen T, Amendola M, et al. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res*, 2014, 42: e168
- 86 Brinkman E K, Kousholt A N, Harmsen T, et al. Easy quantification of template-directed CRISPR/Cas9 editing. *Nucleic Acids Res*, 2018, 46: e58
- 87 Conant D, Hsiao T, Rossi N, et al. Inference of CRISPR edits from sanger trace data. *CRISPR J*, 2022, 5: 123–130
- 88 Kluesner M G, Nedveck D A, Lahr W S, et al. EditR: A method to quantify base editing from sanger sequencing. *CRISPR J*, 2018, 1: 239–250
- 89 Xu L, Liu Y, Han R. BEAT: A python program to quantify base editing from sanger sequencing. *CRISPR J*, 2019, 2: 223–229
- 90 Xue L J, Tsai C J. AGEseq: Analysis of genome editing by sequencing. *Mol Plant*, 2015, 8: 1428–1430

- 91 Park J, Lim K, Kim J S, et al. Cas-Analyzer: An online tool for assessing genome editing results using NGS data. *Bioinformatics*, 2017, 33: 286–288
- 92 Pinello L, Canver M C, Hoban M D, et al. Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol*, 2016, 34: 695–697
- 93 Clement K, Rees H, Canver M C, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol*, 2019, 37: 224–226
- 94 Liu Q, Wang C, Jiao X, et al. Hi-TOM: A platform for high-throughput tracking of mutations induced by CRISPR/Cas systems. *Sci China Life Sci*, 2019, 62: 1–7
- 95 Sun T, Liu Q, Chen X, et al. Hi-TOM 2.0: An improved platform for high-throughput mutation detection. *Sci China Life Sci*, 2024, 67: 1532–1534
- 96 Boel A, Steyaert W, De Rocker N, et al. BATCH-GE: Batch analysis of next-generation sequencing data for genome editing assessment. *Sci Rep*, 2016, 6: 30330
- 97 Connelly J P, Pruett-Miller S M. CRIS.py: A versatile and high-throughput analysis program for CRISPR-based genome editing. *Sci Rep*, 2019, 9: 4194
- 98 Han Y, Liu G, Wu Y, et al. CrisprStitch: Fast evaluation of the efficiency of CRISPR editing systems. *Plant Commun*, 2024, 5: 100783
- 99 Guo C, Ma X, Gao F, et al. Off-target effects in CRISPR/Cas9 gene editing. *Front Bioeng Biotechnol*, 2023, 11: 1143157
- 100 Fu Y, Foden J A, Khayter C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*, 2013, 31: 822–826
- 101 Ran F A, Cong L, Yan W X, et al. *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature*, 2015, 520: 186–191
- 102 Kleinstiver B P, Tsai S Q, Prew M S, et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat Biotechnol*, 2016, 34: 869–874
- 103 Hsu P D, Scott D A, Weinstein J A, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*, 2013, 31: 827–832
- 104 Kim D, Kim J, Hur J K, et al. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol*, 2016, 34: 863–868
- 105 Fu Y, Sander J D, Reyon D, et al. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol*, 2014, 32: 279–284
- 106 Kleinstiver B P, Pattanayak V, Prew M S, et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 2016, 529: 490–495
- 107 Kulcsár P I, Tálás A, Ligeti Z, et al. A cleavage rule for selection of increased-fidelity SpCas9 variants with high efficiency and no detectable off-targets. *Nat Commun*, 2023, 14: 5746
- 108 Tao J, Bauer D E, Chiarle R. Assessing and advancing the safety of CRISPR-Cas tools: From DNA to RNA editing. *Nat Commun*, 2023, 14: 212
- 109 Grünwald J, Zhou R, Garcia S, et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature*, 2019, 569: 443–437
- 110 Grünwald J, Zhou R, Lyer S, et al. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat Biotechnol*, 2019, 37: 1041–1048
- 111 Gaudelli N M, Komor A C, Rees H A, et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, 2017, 551: 464–471
- 112 Zhou C, Sun Y, Yan R, et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature*, 2019, 571: 278–278
- 113 Doman J L, Raguram A, Newby G A, et al. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat Biotechnol*, 2020, 38: 620–628
- 114 Zuo E, Sun Y, Wei W, et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science*, 2019, 364: 289–292
- 115 Jin S, Zong Y, Gao Q, et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science*, 2019, 364: 292–295
- 116 Tsai S Q, Zheng Z, Nguyen N T, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol*, 2015, 33: 187–197
- 117 Wienert B, Wyman S K, Richardson C D, et al. Unbiased detection of CRISPR off-targets *in vivo* using DISCOVER-Seq. *Science*, 2019, 364: 286–289
- 118 Yu Z, Lu Z, Li J, et al. PEAC-seq adopts prime editor to detect CRISPR off-target and DNA translocation. *Nat Commun*, 2022, 13: 7545
- 119 Akcakaya P, Bobbin M L, Guo J A, et al. *In vivo* CRISPR editing with no detectable genome-wide off-target mutations. *Nature*, 2018, 561: 416–419
- 120 Wei Y, Zhou Y, Liu Y, et al. Indiscriminate ssDNA cleavage activity of CRISPR-Cas12a induces no detectable off-target effects in mouse embryos. *Protein Cell*, 2021, 12: 741–745

- 121 Pan X, Qu K, Yuan H, et al. Massively targeted evaluation of therapeutic CRISPR off-targets in cells. [Nat Commun](#), 2022, 13: 4049
- 122 Cameron P, Fuller C K, Donohoue P D, et al. Mapping the genomic landscape of CRISPR–Cas9 cleavage. [Nat Methods](#), 2017, 14: 600–606
- 123 Kim D, Bae S, Park J, et al. Digenome-seq: Genome-wide profiling of CRISPR–Cas9 off-target effects in human cells. [Nat Methods](#), 2015, 12: 237–243
- 124 Tsai S Q, Nguyen N T, Malagon-Lopez J, et al. CIRCLE-seq: A highly sensitive *in vitro* screen for genome-wide CRISPR–Cas9 nuclease off-targets. [Nat Methods](#), 2017, 14: 607–614
- 125 Manghwar H, Li B, Ding X, et al. CRISPR/Cas systems in genome editing: Methodologies and tools for sgRNA design, off-target evaluation, and strategies to mitigate off-target effects. [Adv Sci](#), 2020, 7: 1902312
- 126 Bae S, Park J, Kim J S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. [Bioinformatics](#), 2014, 30: 1473–1475
- 127 Stortz F, Mak J K, Minary P. piCRISPR: Physically informed deep learning models for CRISPR/Cas9 off-target cleavage prediction. [Artif Intell Life Sci](#), 2023, 3: 100075
- 128 Bai M, Lin W, Peng C, et al. Expressing a human RNA demethylase as an assister improves gene-editing efficiency in plants. [Mol Plant](#), 2024, 17: 363–366
- 129 Doench J G, Fusi N, Sullender M, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. [Nat Biotechnol](#), 2016, 34: 184–191
- 130 Zhang Q, Xing H L, Wang Z P, et al. Potential high-frequency off-target mutagenesis induced by CRISPR/Cas9 in *Arabidopsis* and its prevention. [Plant Mol Biol](#), 2018, 96: 445–456

Summary for “基因编辑设计和分析辅助工具的研究进展”

Research advances in gene editing design and analysis tools

Fuquan Li, Ruixiang Zhang, Zhe Zhao, Letian Chen^{*} & Xianrong Xie^{*}

College of Life Sciences, South China Agricultural University, State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangzhou 510642, China

^{*} Corresponding authors, E-mail: lotichen@scau.edu.cn; xiexianrong@scau.edu.cn

The CRISPR/Cas-based gene editing system, owing to its simplicity and high efficiency, has emerged as a revolutionary technology in the field of life sciences. This technology has been widely employed in gene-functional studies, medical research, and genetic improvement in plants and animals. Recently, various gene editing tools have been derived with the continuous improvement of the gene editing systems, including gene knockout, base editing, prime editing, replacement and insertion, and RNA editing. Use of this CRISPR/Cas systems to conduct gene editing experiments in eukaryotes involves a series of steps, including the design of single guide RNAs (sgRNAs), preparation of expression cassette(s), delivery of the Cas-sgRNA component or vector, detection of editing results, and potential off-target sites. Among these steps, appropriate gene editing tools should be selected based on research needs. In addition, the efficiency and off-target properties of sgRNA must be fully considered, as well as the cost and reliability of subsequent detection of target mutations and off-targets. Various bioinformatics tools and methods have been developed for assisting the target design, off-target evaluation and editing outcome detection that greatly facilitate the performance of gene editing.

In this review, we first provide a concise overview of the types, principles and features of existing CRISPR/Cas-based gene editing tools and compare the differences of major editors with diverse CRISPR/Cas nucleases. We next give a comprehensive introduction of the factors influencing editing efficiency and available bioinformatics tools to design target sites, a key step of choosing appropriate target sites(s) for the Cas nucleases. With the wide applications of gene editing technologies in different research fields ranging from individual editing to high-throughput mutant screening, a reliable and cost-effective method is of significance for detecting the editing results. Therefore, we further focus on the comparison of currently methods to evaluate the editing outcomes, including sequencing-independent detection methods, and tools by directly analyzing Sanger or next generation sequencing files. The advantages and limits of these analysis tools are systematically compared in this review. In addition, we summarize the possible reasons that might cause off-target of sgRNAs and introduce bioinformatics tools to evaluate the off-target potential of sgRNAs.

We also discuss the existing problems and perspective of the methods and auxiliary tools for CRISPR/Cas-based gene editing in eukaryotes. With the discovery and use of more Cas nucleases or other editors, their distinct editing properties and efficiency have to been considered, and thus the development of corresponding design tools need to take these editors' characteristics into full consideration. In the future, more smart platforms that integrate automatic methods and tools assisting researchers to perform all steps of gene editing design and experiments would be greatly accelerate the application of gene editing technologies in basic researches and genetic improvement.

Overall, this review can provide feasible references for facilitating the usage of gene editing experiments by utilizing these methods and tools.

CRISPR/Cas, gene editing, target design, off-target, target detection, bioinformatics tools

doi: [10.1360/TB-2023-1334](https://doi.org/10.1360/TB-2023-1334)