**基因编辑**

2019级神经所基正邦整理

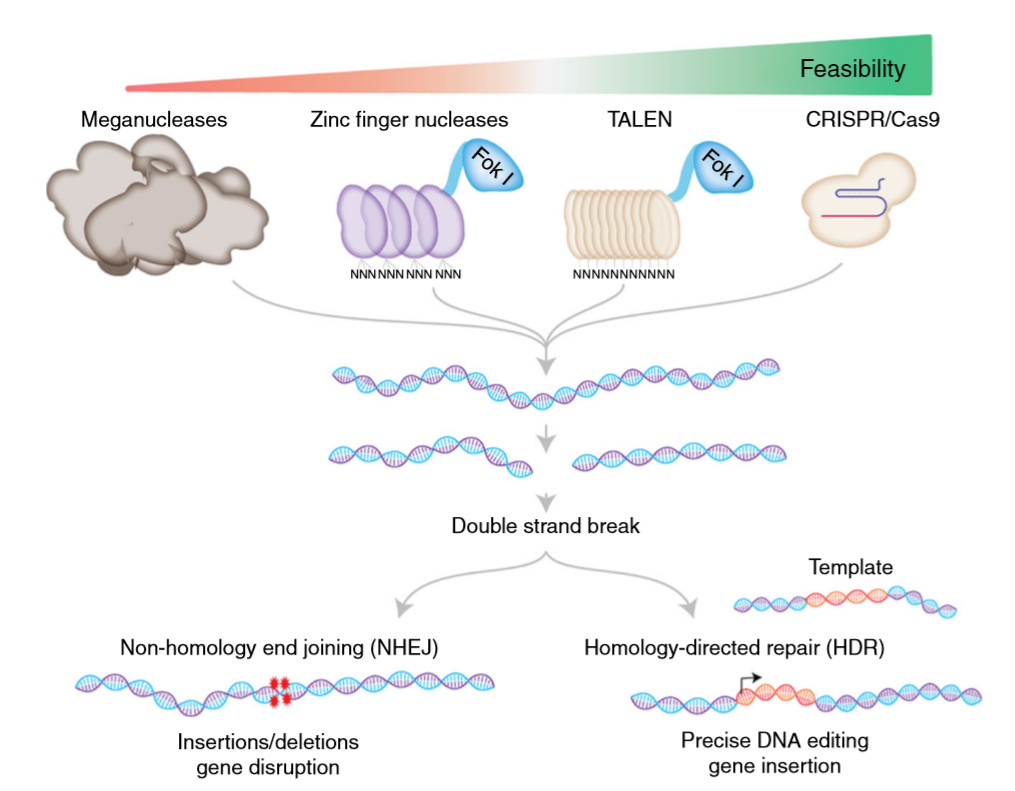
仅供复习用，错误之处，还请指正！

**重点可能是：基因编辑工具有哪些应用，基因编辑技术有哪些、原理是什么、应用。**

**推荐阅读：网络文章**《**基因编辑技术CRISPR的发展简史：从发现到爆炸》和**

**Mazhar Adli1的The CRISPR tool kit for genome editing and beyond**

1. **基因编辑技术有哪些**？**原理是什么、应用（补充三大技术的对比）**



Cited from Mazhar A, The CRISPR tool kit for genome editing and beyond, Nature communication, 2018

以下总结只是简略版，详见

<https://www.cyagen.com/cn/zh-cn/community/frontier/information-20150721.html>

**基因组编辑三大利器：TALEN、ZFN和CRISPR/Cas9**

转录激活样效应因子核酸酶（transcription activator-like effector nuclease, TALEN）技术与锌指核酸酶（Zinc-finger nuclease, ZFN）技术组成了一大类强有力的基因组编辑工具，这一大类技术的发展重新划定了生物学研究的边界。这些嵌合核酸酶由两部分组成——一个可编码的序列特异性DNA结合模块与一个非特异性的DNA切割结构域。通过诱导DNA双链断裂（DNA double-strand break）来刺激容易出错的非同源末端连接或在特定基因所在的位置进行的同源定向修复，TALEN和ZFN能够完成一系列遗传学编辑修饰操作。  
  
成簇规律间隔短回文重复（clustered regulatoryinterspaced short palindromic repeat, CRISPR）技术是最新出现的一种基因组编辑工具，它能够完成RNA导向的DNA识别及编辑。与其它基因组编辑工具相比，CRISPR技术更易于操作，具有更强的可扩展性。

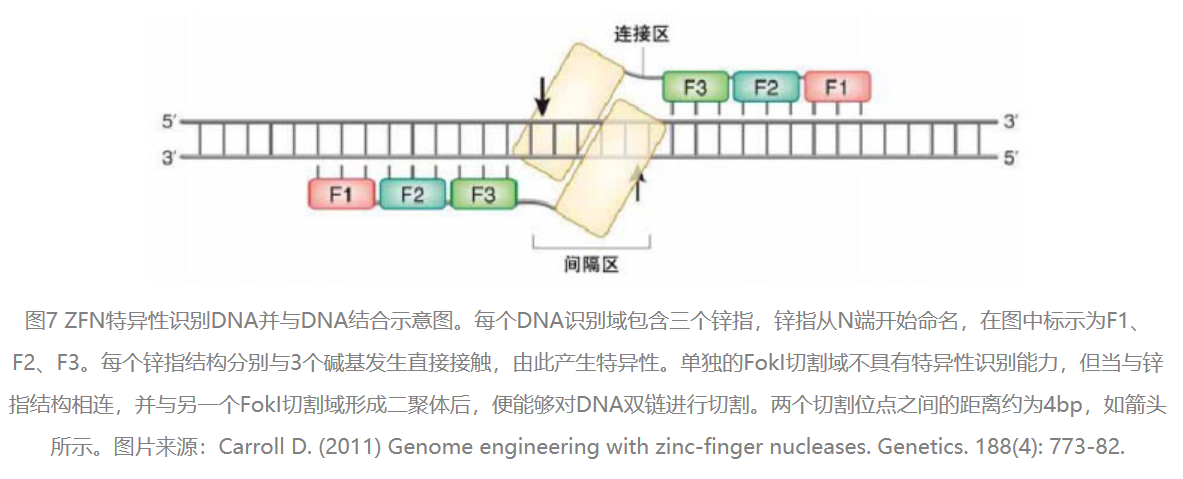
1. **ZFN**

锌指核酸酶（Zinc-finger nuclease, ZFN）又名锌指蛋白核酸酶（ZFPN），它是一类人工合成的限制性内切酶，由锌指DNA结合域（zinc finger DNA-binding domain）与限制性内切酶的DNA切割域（DNA cleavage domain）融合而成。研究者可以通过加工改造ZFN的锌指DNA结合域，靶向定位于不同的DNA序列，从而使得ZFN可以结合复杂基因组中的目的序列，并由DNA切割域进行特异性切割。此外，通过将锌指核酸酶技术和胞内DNA修复机制结合起来，研究者还可以自如地在生物体内对基因组进行编辑。

**（1）原理**

**结构：**ZFN由负责特异性识别序列的锌指DNA结合域和进行非特异性限制性内切酶切割的DNA切割域两部分组成。其中锌指DNA结合域部分一般包含3个独立的锌指（Zinc finger, ZF）重复结构，每个锌指结构能够识别3个碱基，因而一个锌指DNA结合域可以识别9bp长度的特异性序列（而ZFN二聚体，则包含6个锌指，可以识别18bp长度的特异性序列）。目前最常用的ZF结构为Cys2His2锌指，其结构由大约30个氨基酸包裹一个锌原子构成。

ZFN的切割域与DNA结合域通过连接区（linker）结合。在ZFN中应用最广泛的DNA切割域来自IIS型限制性内切酶FokI。由于切割域与DNA链的结合能力较弱，因此DNA切割域必须以二聚体的形式发挥作用。构建锌指核酸酶时，应针对DNA各链上的邻近区域设计两条ZFN，使其DNA切割域能够位于双链的同一位置，以达到最佳的切割效果。两条ZFN之间具有被称为“间隔区”的spacer结构，该结构的长度以5～6bp为宜，7bp也能正常工作，合理的“间隔区”设计才能保证ZFN二聚体拥有最佳的工作空间。



**原理：**针对目的基因序列设计并合成ZFN后，使之对DNA进行特异性切割，从而形成DNA双链断裂区（Double-StrandedBreaks,DSB）；通过破坏非同源末端链接（non-homologous end joining, NHEJ）使目的基因失活，或借助同源重组（homologous recombination, HR）等方式完成DNA的修复连接，可以使断裂的DNA双链重新黏合。将以上两步结合起来便可以完成一般的基因组编辑操作

**（2）应用**

ZFN技术具有重大的应用价值。在科研和农业领域，该技术既可用于基因的敲除失活，也可用于导入目标基因，使基因激活或阻断，或者人为改造基因序列，使之符合人们的要求。在医疗领域，经ZFN技术改造后导入治疗性基因的质粒或干细胞可被导入人体，实现基因治疗。此外，ZFN技术也可以直接用于有害基因的修补替换或是直接删除，以达到相关治疗目的。ZFN技术具有极佳的特异性和效率，因此能将基因/基因组错误修改的风险降到最低。从理论上来说，研究人员甚至可以在任何物种中，对处于任意生长时期的细胞进行ZFN操作，可以自如地修改其基因，而还不破坏细胞状态。

ZFN技术的先导之一，Sangamo Biosciences公司正在和宾夕法尼亚大学合作，研究ZFN技术通过介导核酸酶引起CCR5基因座的破坏。这一研究成果在治疗HIV中具有广阔的应用前景，而且ZFN技术也已尝试应用于杜氏肌营养不良症（Duchenne muscular dystrophy）、21三体综合征等遗传疾病的基因治疗。

**（3）缺陷**

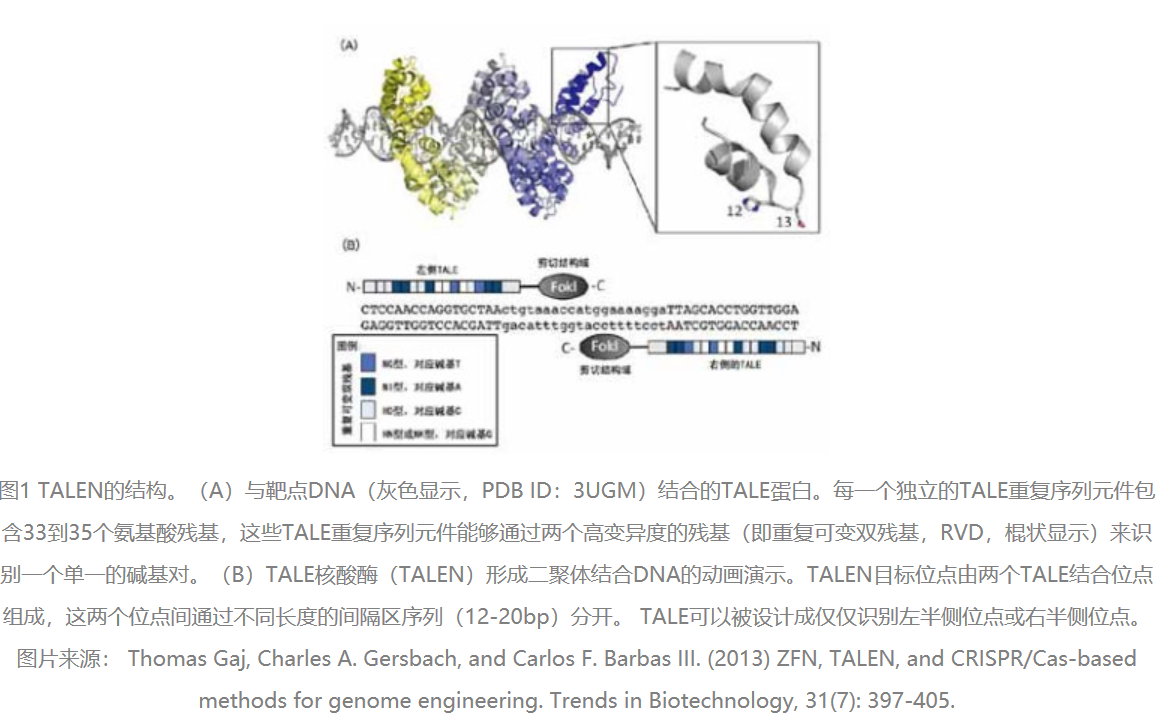
ZFN 技术虽然简易实用，但也具有一定缺陷。ZFN对DNA的剪切需要两个FokI切割区域的二聚化，并且需要至少一个识别单元结合DNA。DNA识别域虽然具 有较强的特异性识别能力，但由于ZFN剪切的过程并不完全依赖同源二聚体的形成，所以一旦形成异源二聚体，就很可能造成脱靶效应，并最终可能导致DNA的 错配和序列改变，产生较强的细胞毒性。当这些不良影响积累过多，超过细胞修复机制承受的范围时，便会引起细胞的凋亡。另一方面，该手段仍然受到现有生物学 领域研究手段的限制，因此在细胞内部操作的精确程度和后果都较难预料。如果ZFN引起相关基因突变，则可能会导致一系列意想不到的后果，在与人体相关的应 用领域，甚至可能引发癌症。另外，ZFN作为基因治疗的手段之一，如果在生物体内使用，可能会引发免疫反应。现有的研究手段尚不能预测引入的ZNF蛋白是 否会引起免疫系统的进攻。并且到目前为止，ZFN技术只能用于体外操作（in vitro），在对人体提取的细胞进行处理之后，再导入回输到病人体内。而直接向患者体内导入相关ZFN元件进行基因编辑处理则具有较大的潜在风险，且效 率不高。以上诸多限制导致人体相关的ZFN操作较为繁琐，难以推广应用。

1. **TALENs(transcription activator-like (TAL) effector nucleases)，[转录激活因子](https://baike.baidu.com/item/%E8%BD%AC%E5%BD%95%E6%BF%80%E6%B4%BB%E5%9B%A0%E5%AD%90/3471949" \t "_blank)样**[**效应物**](https://baike.baidu.com/item/%E6%95%88%E5%BA%94%E7%89%A9/8061762)[**核酸酶**](https://baike.baidu.com/item/%E6%A0%B8%E9%85%B8%E9%85%B6/311972)

TAL效应因子（TAL effector, TALE）最初是在一种名为黄单胞菌（Xanthomonas sp.）的植物病原体中作为一种细菌感染植物的侵袭策略而被发现的。这些TALE通过细菌 III类分泌系统（bacterial type III secretion system）被注入植物细胞中，通过靶定效应因子特异性的基因启动子来调节转录，来促进细菌的集落形成。由于TALE具有序列特异性结合能力，研究者通过将FokI核酸酶与一段人造TALE连接起来，形成了一类具有特异性基因组编辑功能的强大工具，即TALEN。

1. **原理**

**结构**：典型的 TALEN由一个包含核定位信号（Nuclear localization signal, NLS）的N端结构域、一个包含可识别特定 DNA序列的典型串联TALE重复序列的中央结构域，以及一个具有FokI核酸内切酶功能的C端结构域组成。



**原理：**通过 DNA识别模块将TALEN元件靶向特异性的DNA位点并结合，然后在FokI核酸酶的作用 下完成特定位点的剪切，并借助于细胞内固有 的同源定向修复（HDR）或非同源末端连接 途径（NHEJ）修复过程完成特定序列的插入 （或倒置）、删失及基因融合。

**构建：**实现TALEN的关键就在于完成DNA的特异性识别功能，一般说来分为两个步骤:

1. 构建TAL靶点识别模块  
   TAL的DNA特异性识别单位是间隔32个恒定氨基酸残基的二联氨基酸。二联氨基酸与AGCT这4个核苷酸碱基有一一对应的关系：腺嘌呤（A）由NI识别、胸腺嘧啶（T）由NG识别、鸟嘌呤（G）由NN识别，而胞嘧啶（C）则由HD识别。实验操作中，我们通过 靶位点的DNA序列可以反推能特异性识别这一序列的二联氨基酸序列，从而构建TAL靶点识别模块。
2. TAL靶点识别模块的克隆与表达  
   根据之前对TALEN结构的介绍，我们需要将上一步骤中根据目标DNA序列构建好的一对TAL靶点识别模块与N端的核定位序列、C端的FokI酶连接起来，才能得到一个完整的TALEN元件。一般来说，我们可以采用专门用于构建TALEN的真核表达载体体系，将一对特异性的TAL靶点识别模块克隆进该载体中，再通过转染等方式导入细胞内。这种体系一般由供体质粒（donor plasmid，提供单基、二联及三联等类型的TAL模块）和骨架质粒（backbone plasmid，用于构建TALEN并表达构建好的TALEN）两类质粒构成。
3. **应用**

自2010年正式发明 TALEN技术以来，全球范围内多个研究小组利用体外培养细胞、酵母、拟南芥、水稻、果蝇及斑马鱼等多个动植物体系应用了TALEN做基因组编辑。

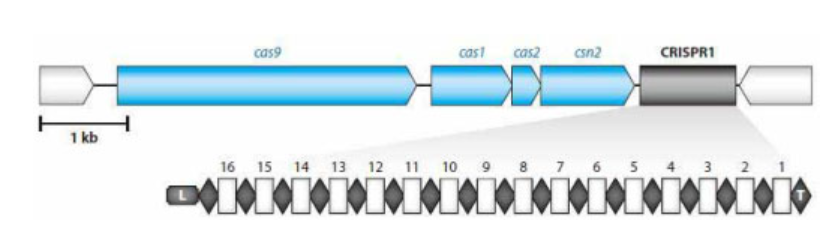
以斑马鱼为模式动物，使用TALEN技术在活体内完成定向突变、特定 DNA的删除、人工DNA插入等较为复杂的操作。植物、大小鼠的基因组改造。

**3.CRISPR/Cas系统**

不论是TALEN技术还是ZFN技术，其定向打靶都依赖于DNA序列特异性结合蛋白模块的合成，这一步骤非常繁琐费时。而CRISPR/Cas技术作为一种 最新涌现的基因组编辑工具，能够完成RNA导向的DNA识别及编辑。CRISPR/Cas技术使用一段序列特异性向导RNA分子（sequence- specific guide RNA）引导核酸内切酶到靶点处，从而完成基因组的编辑。CRISPR/Cas系统的开发为构建更高效的基因定点修饰技术提供了全新的平台。

**(1)CRISPR/Cas系统元件与特征**

CRISPR/Cas 系统最早是在细菌的天然免疫系统内发现的，其主要功能是对抗入侵的病毒及外源DNA。1987年大阪大学（OsakaUniversity）的研究人员在 E.coliK12的碱性磷酸酶基因附近发现了成簇的规律间隔的短回文重复序列（Clustered regularly interspaced short palindromic repeat, CRISPR），其结构如图所示，目前普遍认为有40%的细菌基因组具有这样的结构。

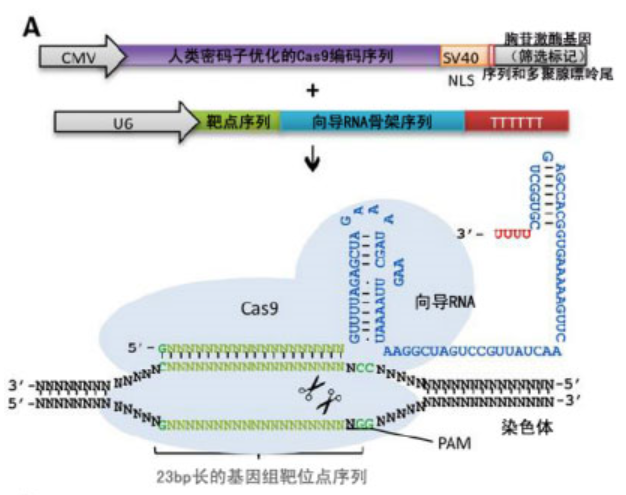


CRISPR/Cas系统由CRISPR序列元件与Cas基因家族组成。其中CRISPR由一系列高度保守 的重复序列（repeat）与同样高度保守的间隔序列（spacer）相间排列组成。而在CRISPR附近区域还存在着一部分高度保守的CRISPR相关 基因（CRISPR-associated gene, Cas gene），这些基因编码的蛋白具有核酸酶活性的功能域，可以对DNA序列进行特异性的切割。

**(2)原理**

CRISPR区域第一个重复序列上游有一段CRISPR的前导序列（Leader sequence），该序列作为启动子来启动后续CRISPR序列的转录，转录生成的RNA被命名为CRISPRRNA（简称crRNA）。

CRISPR/Cas 系统中crRNA与tracrRNA（反式激活的crRNA）形成嵌合RNA分子，即单向导RNA（Single guide RNA,sgRNA）。sgRNA可以介导Cas9蛋白在特定序列处进行切割，形成DNA双链断裂（Double-Stranded Break, DSB），完成基因定向编辑等的各类操作。



人类细胞的RNA介导基因打靶：设计 C末端包含SV40核定位信号 （nuclear localization signal）的Cas9蛋白和一个或一个以上的向导RNA（guide RNA, gRNA）的共表达（上半部分共表达两个质粒的结构示意图），这一过程由人类U6聚合酶III（U6 polymerase III）的启动子介导。Cas完成DNA双链的解聚并在gRNA（guide RNA）的识别下切割特定的DNA双链位置，该过程前提是其3’端有序列正确的原间隔物模块（protospacer-adjacent motif, PAM）。

1. **应用**

张锋的研究组使用CRISPR技术完成了多重基因组编辑， Church的研究组首次使用CRISPR技术完成了RNA介导的人类基因组编辑。周琪研究员利用CRISPR-Cas技术在大鼠中实现了多基因同步敲除； Jaenisch利用CRISPR-Cas技术构建了条件敲除的小鼠转基因模型；瞿礼嘉教授课题组利用CRISPR-Cas系统成 功地实现了对水稻特定基因的定点突变； Gersbach研究组尝试使用CRISPR技术进行基因治疗。

**4.三大技术的对比**

**（1）共同点**

**a.本质**

本质上均是联合特异性DNA的靶向识别及核酸内切酶产生特定位点的双链断裂，利用非同源末端链接途径（NHEJ）修复和同源重组（HR）修复，使DNA序列改变。

**b.应用的模式**

它们在各类应用中的基本模式是相似的，例如在转基因大鼠的构建上，三种技术均是以显微注射的方式进入大鼠胚胎的。

**（2）不同点**

TALEN技术是目前商业化最成功的技术，虽然将单个的 TALEN模块进行组装需要大量的分子克隆和测序操作，十分繁琐，但是很多商业公司可以提供组装好的三联密码子TALEN模块，甚至四联密码子TALEN 模块，这样就大大缩短了构建TALEN元件的实验周期。不过也正是因为如此，绝大多数实验室都难以自行完成TALEN技术的完整操作，对其推广造成了障 碍。

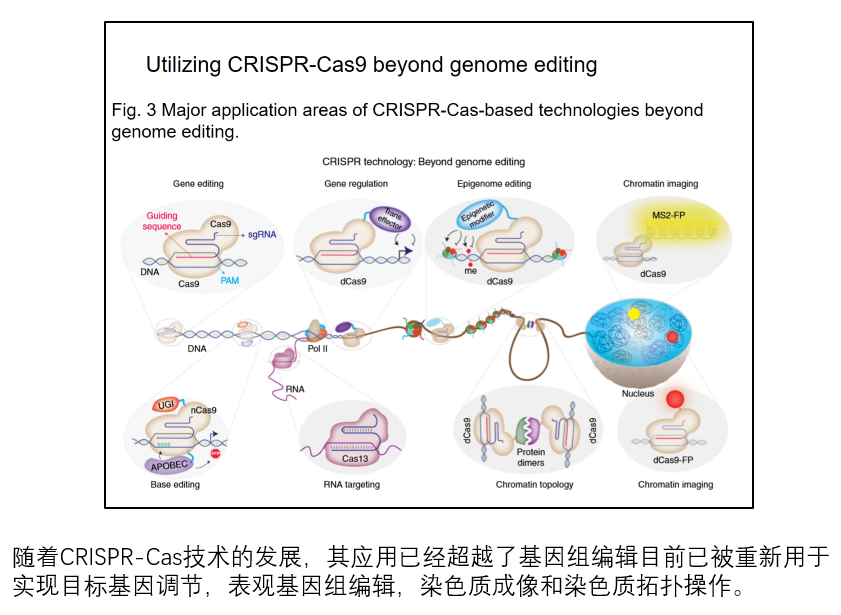
ZFN技术则是最早被广泛使用的基因组定点修饰技术，各大平台均比较完善，有很多可以直接使用的资源，然而由于其自身的三联属性，其设计比TALEN更为繁琐，而且高度依赖于目标序列及其上下游序列，还具有脱靶率高及细胞毒性大等诸多限制性因素。

CRISPR/Cas 技术摆脱了合成并组装具有特异性DNA识别能力蛋白模块的繁琐操作，其gRNA的设计和合成工作量远远小于TALEN和ZFN技术的DNA识别模块的构建 过程，且毒性远远低于ZFN技术。然而CRISPR/Cas技术也有上下文依赖性，目前只能应用于上游有PAM序列的靶位。

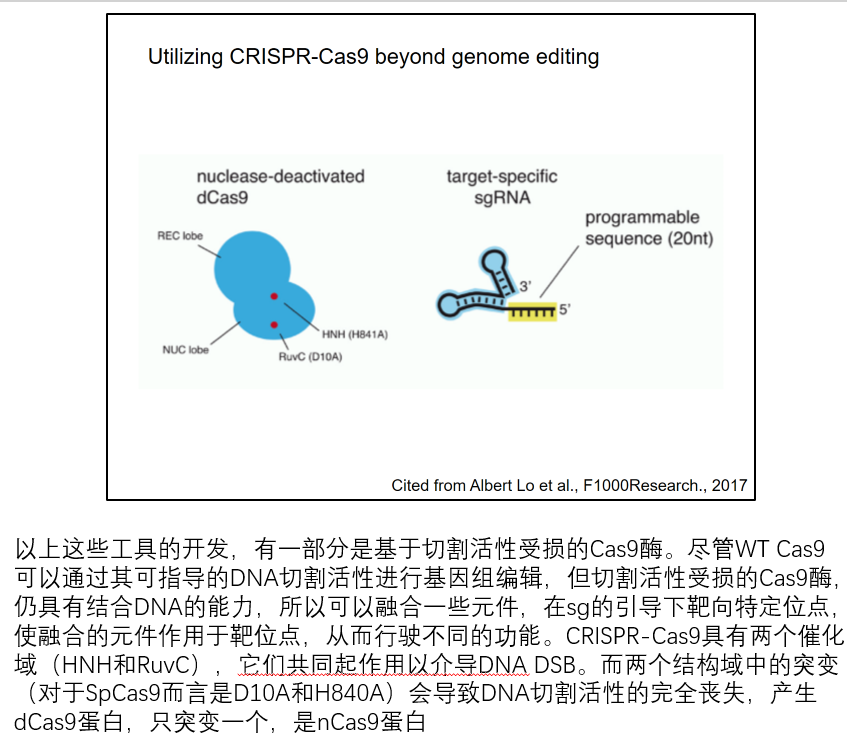


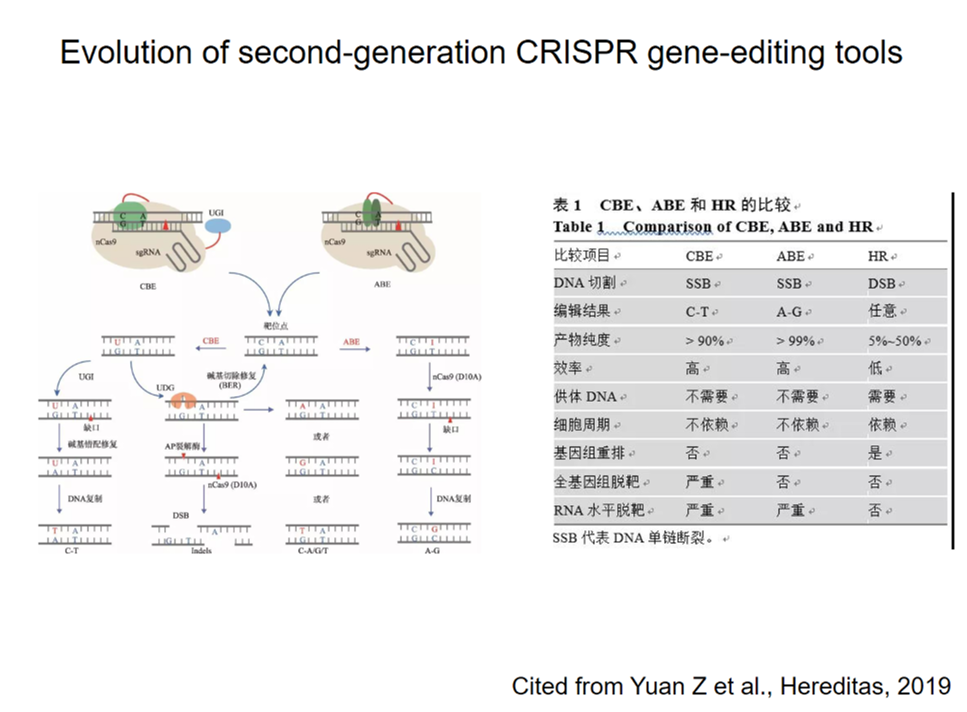
1. **CRISPR相关工具及其应用**

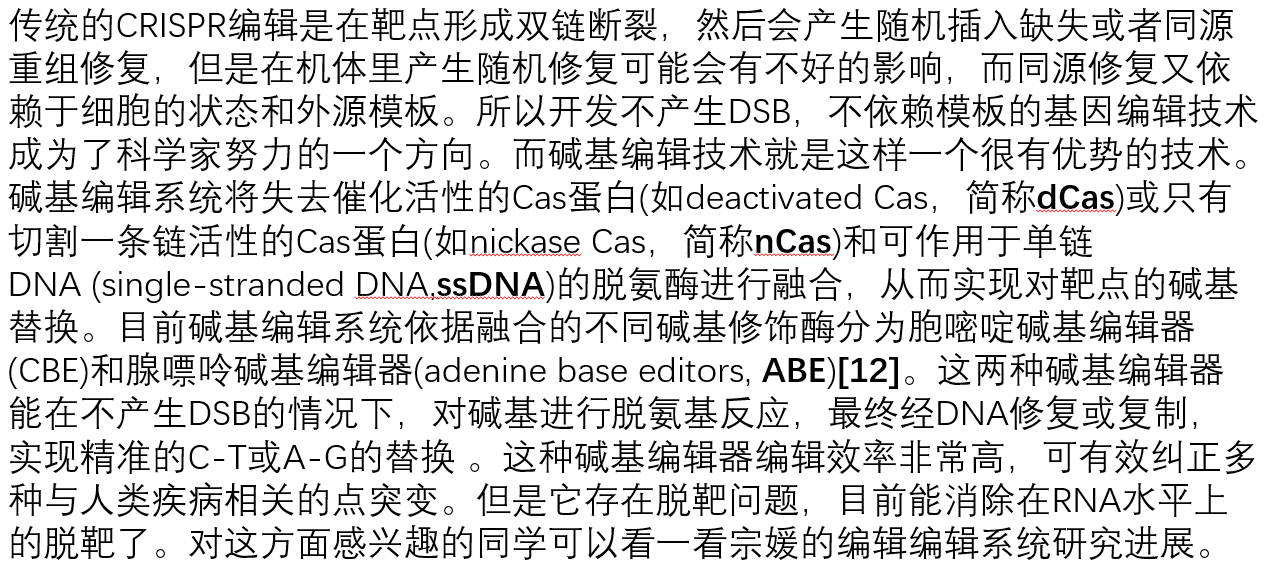
以下只是简单概述，如需了解CRISPR相关工具具体原理，可参考PPT上写的引用文献。

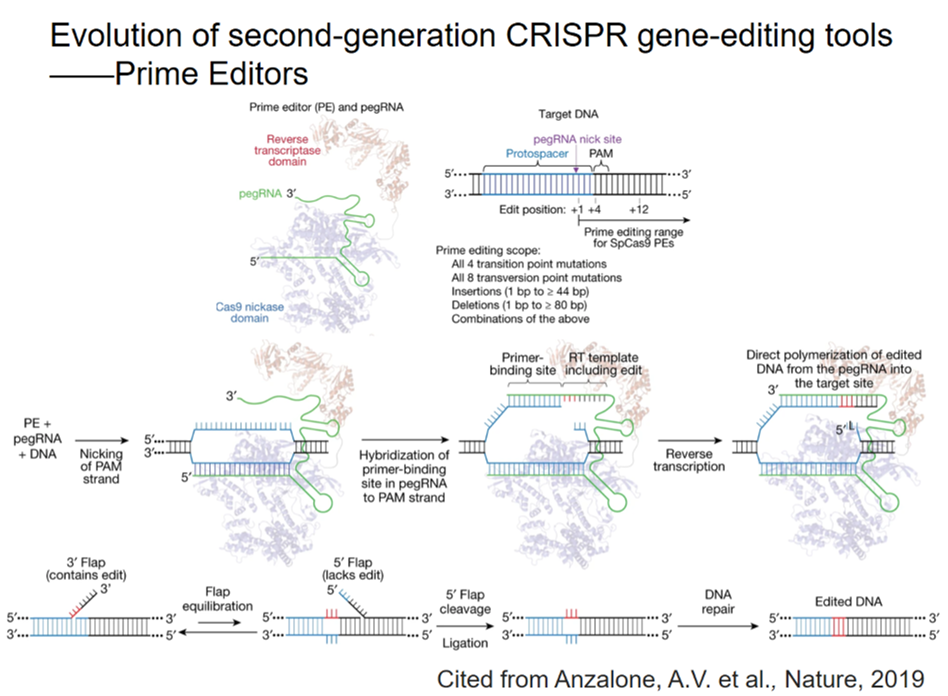


Cited from Mazhar A, The CRISPR tool kit for genome editing and beyond, Nature communication, 2018



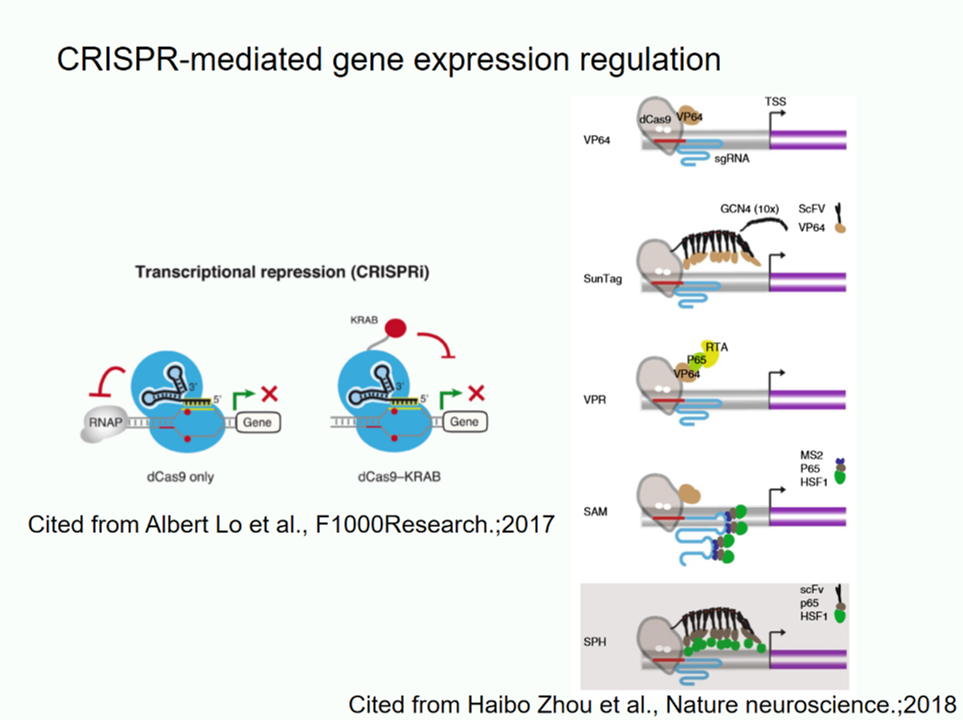






新工具PE是以CRISPR-Cas9系统为基础，在两方面加以改造：首先是改造单链引导RNA（sgRNA），其3’末端增加了一段RNA序列，新获得的RNA被称作pegRNA；第二则是将Cas9切口酶（H840A突变型，只切断含PAM的靶点DNA链）与逆转录酶融合获得新的融合蛋白。pegRNA的3’端序列有双重角色，一段序列作为引物结合位点（PBS），与断裂的靶DNA链3’末端互补以起始逆转录过程，另一端序列则是逆转录的模板（RT模板），其上携带有目标点突变或插入缺失突变以实现精准的基因编辑。

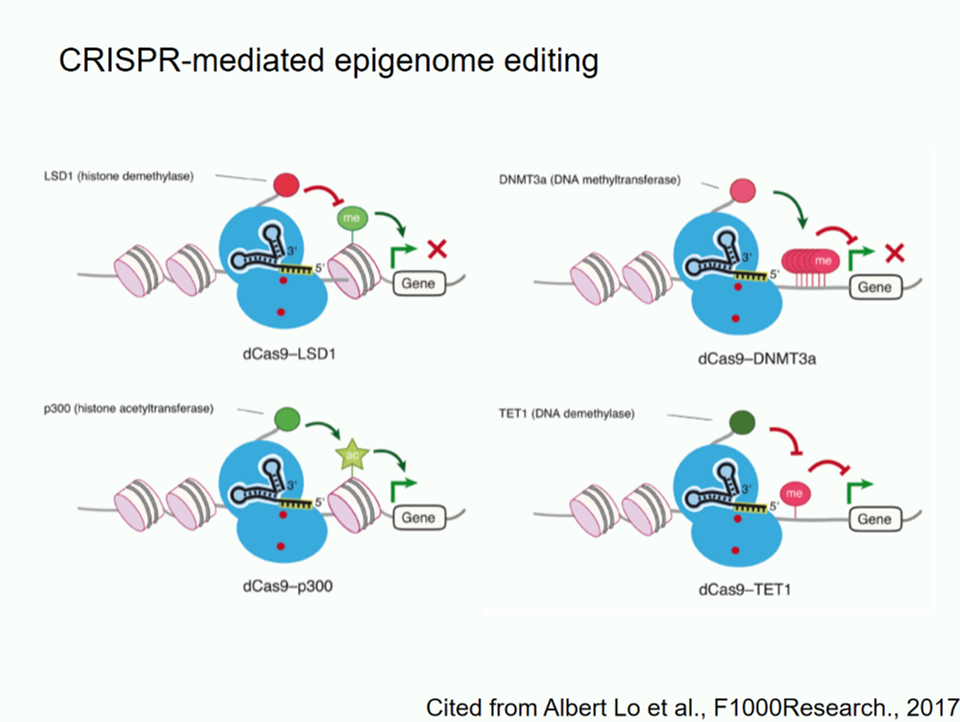
在pegRNA的引导下，Cas9 H840切口酶切断含PAM的靶点DNA链，断裂的靶DNA链与pegRNA的3’末端PBS序列互补并结合，之后逆转录酶发挥功能，沿RT模板序列开始逆转录反应。反应结束后DNA链的切口处会形成处在动态平衡中的5’-和3’-flap结构，其中3’flap结构的DNA链携带有目标突变，而5’flap结构的DNA链则无任何突变。细胞内5’flap结构易被结构特异性内切酶识别并切除，之后经DNA连接和修复后靶位点处便实现了精准的基因编辑。



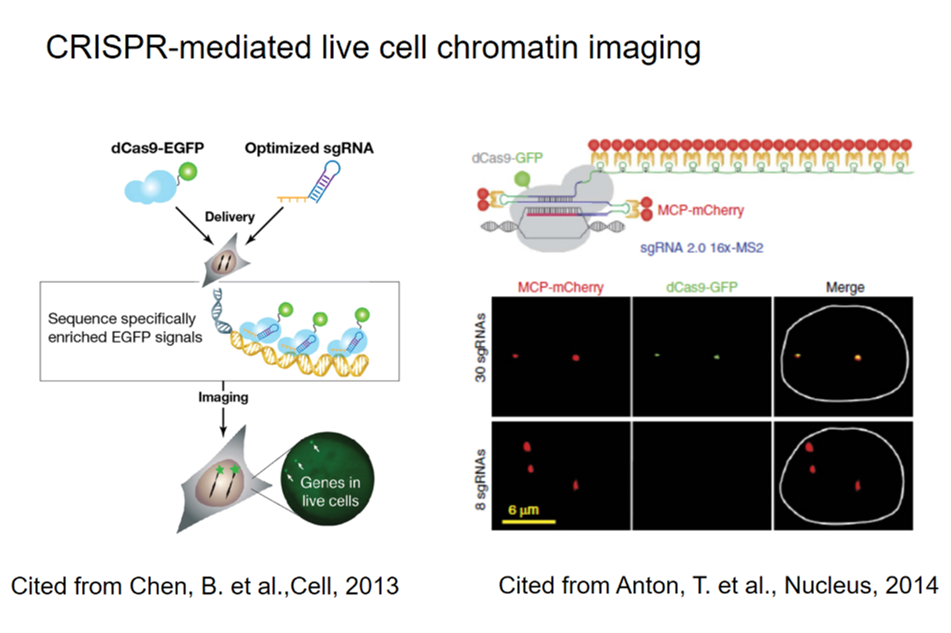
我们之前提到的dCas9也可以用来介导基因表达调控，一方面是抑制基因表达，一方面是激活内源基因表达。

dCas9与DNA靶序列牢固结合，会干扰其他DNA结合蛋白（例如内源转录因子和RNA聚合酶II，阻止了转录过程，从而敲低了（KD）基因表达。值得注意的是，将诸如Kruppelassociated Box（KRAB）之类的强大阻遏物复合物与dCas9融合，会产生比单独的dCas9更强大，更特异性的基因阻遏物。

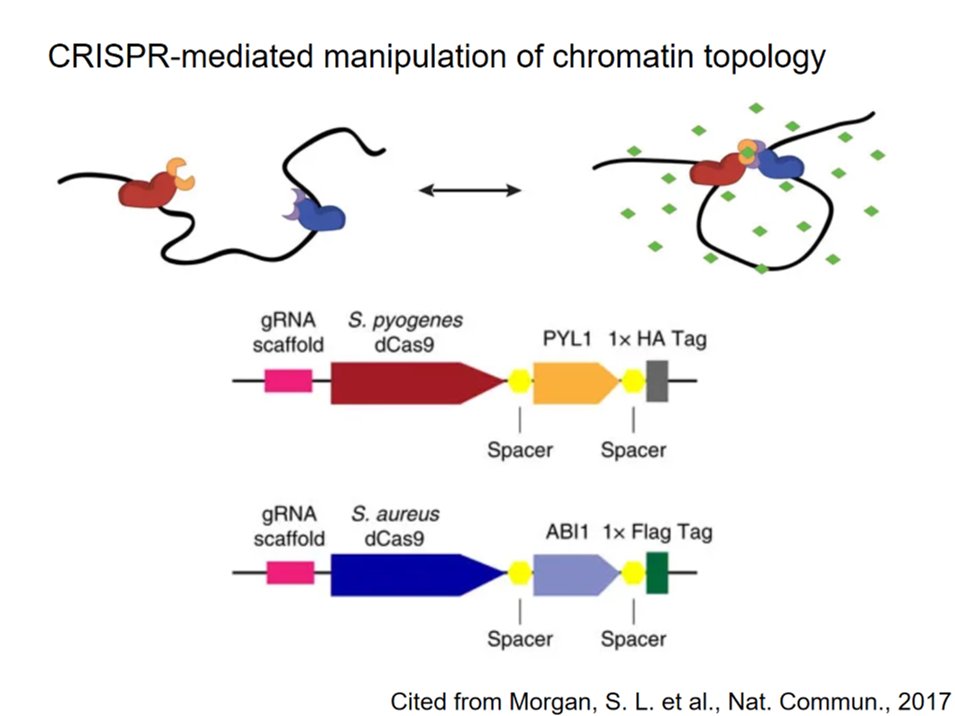
与dCas9-KRAB介导的基因阻抑相反，使用dCas9靶向平台募集强大的转录激活因子会导致基因表达的强烈诱导。之前有上面四种版本的激活系统，后来杨老师实验室在SunTag系统和SAM系统的基础上建立了SPH系统，其激活能力显著强于以往的系统，并且这个系统在神经科学方面得到了应用，通过激活几个特定的基因，这个系统实现了人为将胶质细胞转分化为神经元。



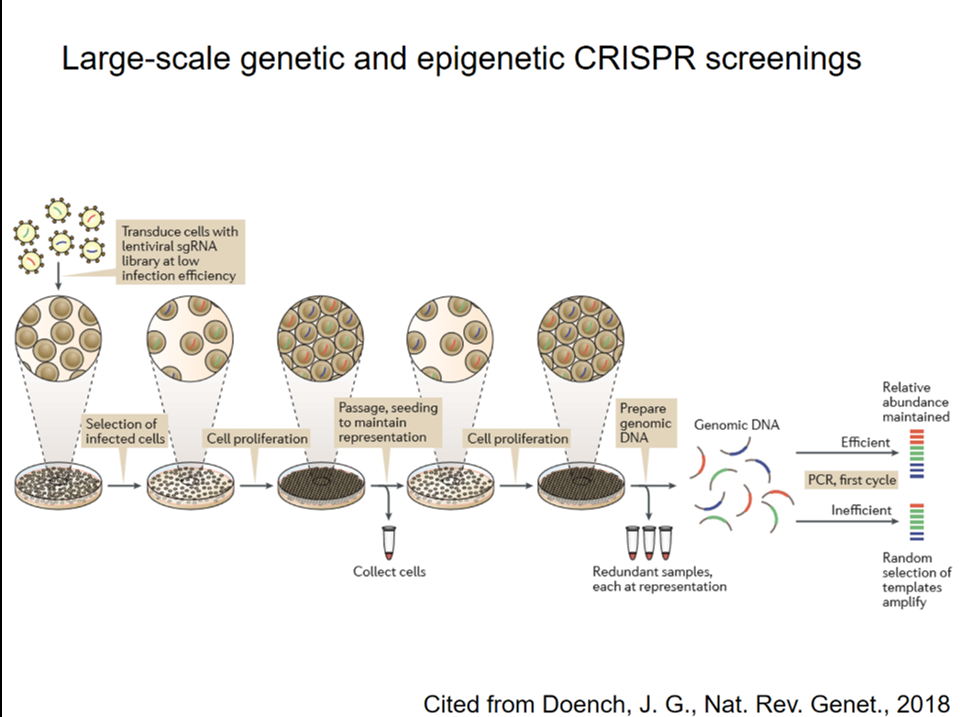
基因组中有多层表观遗传调控机制，有一些疾病也是和表观遗传有关联的。表观遗传层面包括DNA甲基化，组蛋白翻译后修饰和非编码RNA。通过在dCas9上面融合一些效应蛋白，可以实现对表观修饰的写入或者擦除，调控靶基因的表达和细胞的状态。但是这项技术目前研究的还不是特别清楚，也存在比较普遍的脱靶。



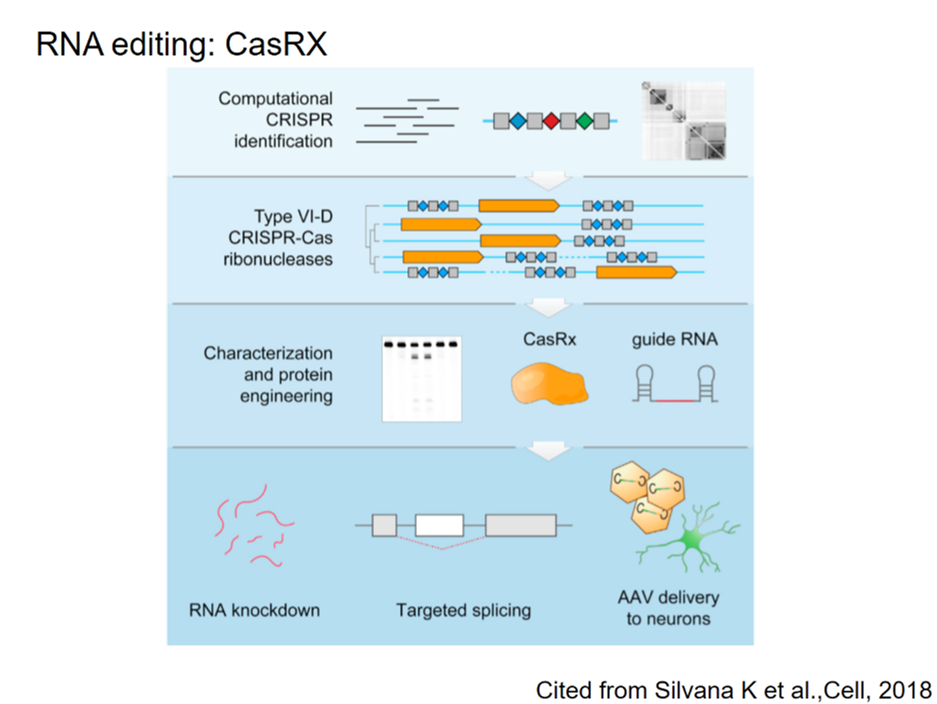
CRISPR技术也可以用于活细胞内的染色质成像，传统的Fish研究的都是死细胞，然后用荧光探针和基因组杂交，这样不能看到活细胞中真实的状态和一些动态变化。2013年，科学家将荧光蛋白和dCas9融合，利用sg将这个复合物带到靶位点附近，即可观察到靶位点在细胞中的位置，但是这个方法有一个问题，dCas9-EGFP复合物在细胞核里是自由浮动的，所以有很强的背景信号，信噪比很低，对于非重复区域要转30个左右特异的sg，使荧光在这个位点聚集，才能提高信噪比。为了解决这个问题，作者实验室开发了一种sgRNA支架，包含多个MS2结合模块，通过在靶位点招募与MCP融合的荧光蛋白，来实现靶位点的信号放大，可以看到用8个的sg的时候，这个系统可以标记出dCas9-EGFP标记不到的信息。



CRISPR技术在研究染色质功能这一方面，还有一个应用就是可以操纵内源染色质的结构，从而了解染色质的功能，对基因表达有什么贡献，也许将来可以用于克服一些遗传缺陷，从而应用于基因治疗。比如2017年这篇文献，就通过将一个蛋白的二聚体拆成两部分分别和两个dCas9融合，利用蛋白的二聚化作用强制染色质成环，激活了沉默的基因。



CRISPR还可以用于大规模的遗传或者表观遗传筛选，使用数百数千个靶向不同基因的sgRNA(一个基因6到10个sg),引导Cas9或者dCas9靶向不同的基因，最终筛选出影响特定表型的基因。这个方法使利用低感染能力的病毒将Cas9/sgRNA递送到细胞中，保证一个细胞只接受一个或者零个sg。病毒将sg整合到细胞基因组中，如果这个sgKO了一个细胞增殖必须的基因的话，那么后期对细胞基因组进行扩增，定量分析各个sg的丰度的时候，会发现含有这个sg的细胞数目减少了，由此可以筛出一个有重要功能的基因。



前面提到的CRISPR都是靶向DNA的，但是在基因治疗层面，靶向RNA比DNA更加安全，因为这个过程是可逆转的。科学家们通过生物信息学分析找到了细菌DNA库里潜在的CRISPR，并且鉴定出了一种能靶向RNA的Cas蛋白，Cas13d，改造成了RNA编辑工具CasRx.这个系统可以用来做RNA knock down，也可以通过AAVdeliver到神经元中，纠正一些RNA错误的剪切形式。Cas13由于体积小，没有PAM，且对基因没有不可逆的影响，所以有望应用于成体治疗，但是它设别靶点后会对周围的RNA进行随机切割，需要进行工程化改造才能应用。