

基因编辑技术

陈金东

广州市朝利良生物科技有限公司 技术总监
国家外专局A类外国专家
美国国际癌症研究协会AACR资深成员
贵州医疗卫生援黔专家团核心专家
遵义医学院候鸟专家

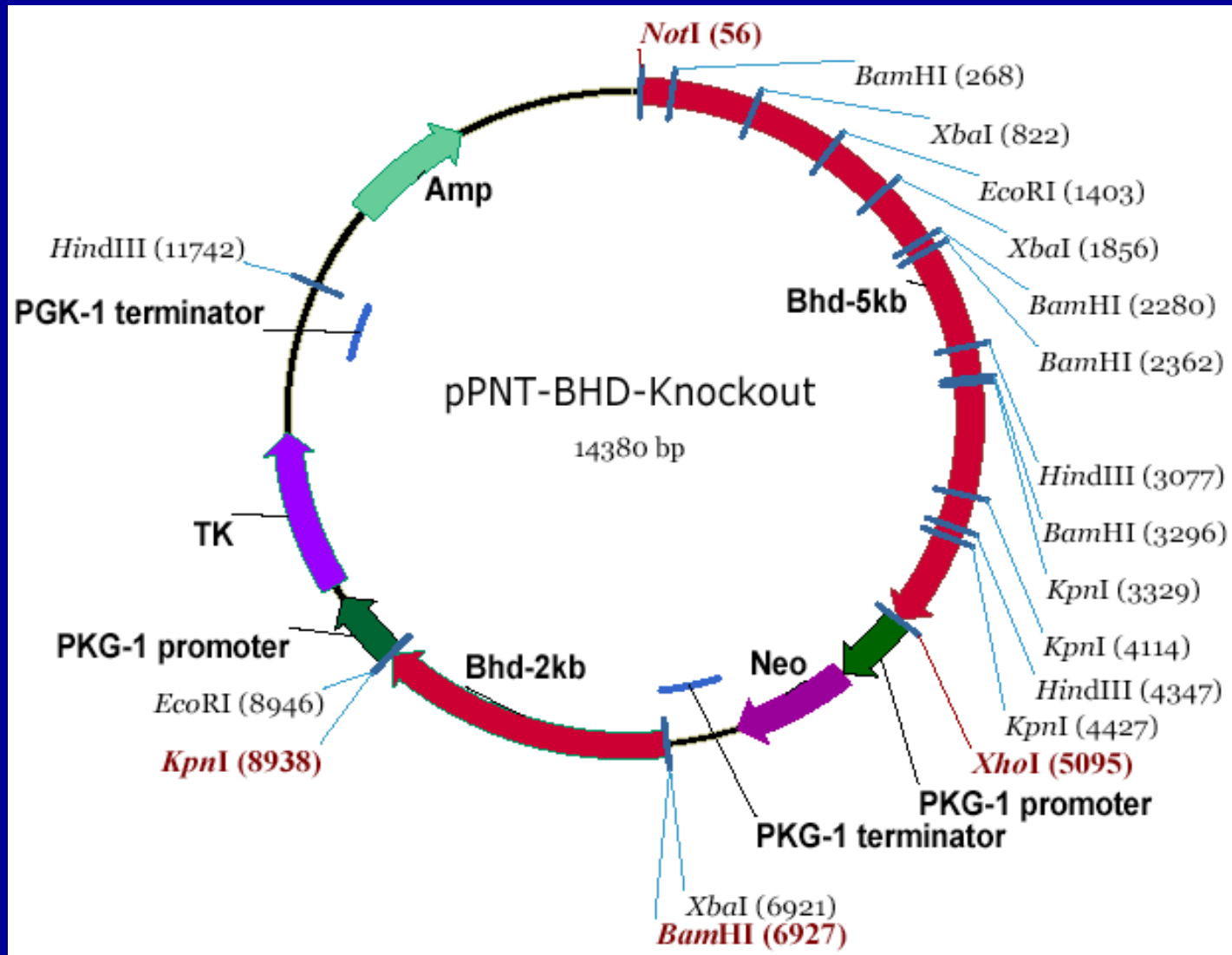
基因编辑技术的产生背景

- 自从证明DNA是遗传物质以后，人类就开始了通过改变DNA来改造生物体生理机理和功能的尝试，尤其是人类基因组计划的完成，有必要了解每个基因的功能。
 - 1973年，斯坦利·N·科恩(Stanley N. Cohen)和赫伯特·W·博耶(Herbert W. Boyer) 成功将蛙的DNA插入到细菌中。
 - 20世纪70年代，基因泰克(Genentech) 公司对大肠杆菌进行基因改造，使其带有一个入源基因(这个基因是人工合成的)，最后生产出治疗糖尿病的胰岛素。
 - 1974年，，加利福尼亚州索克生物研究所(Salk Institute for Biological Studies)的Rudolf Jaenisch通过将SV40 病毒的DNA注射到小鼠的囊胚中，培育出了第一只转基因小鼠。

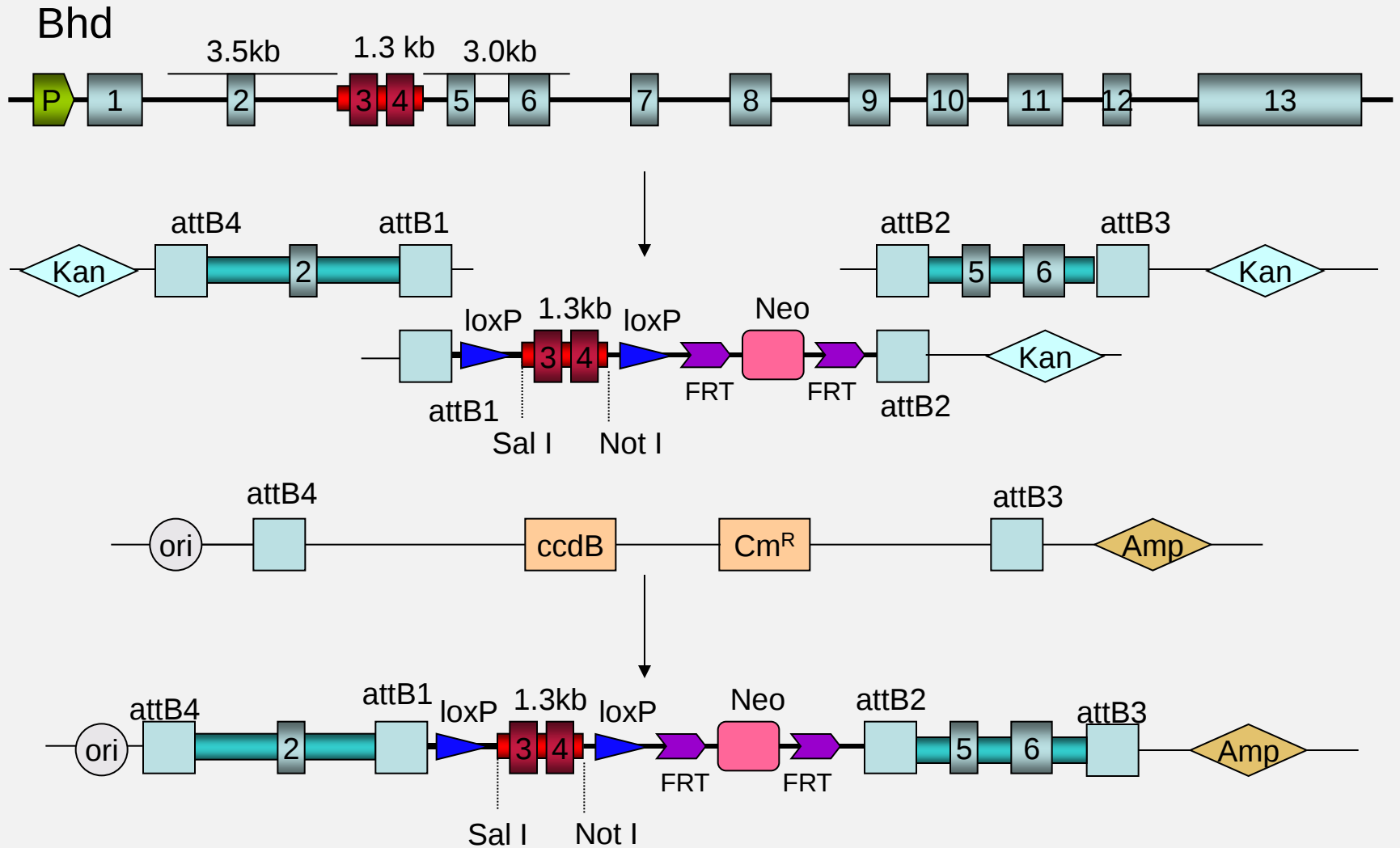
目前已有的基因改造技术

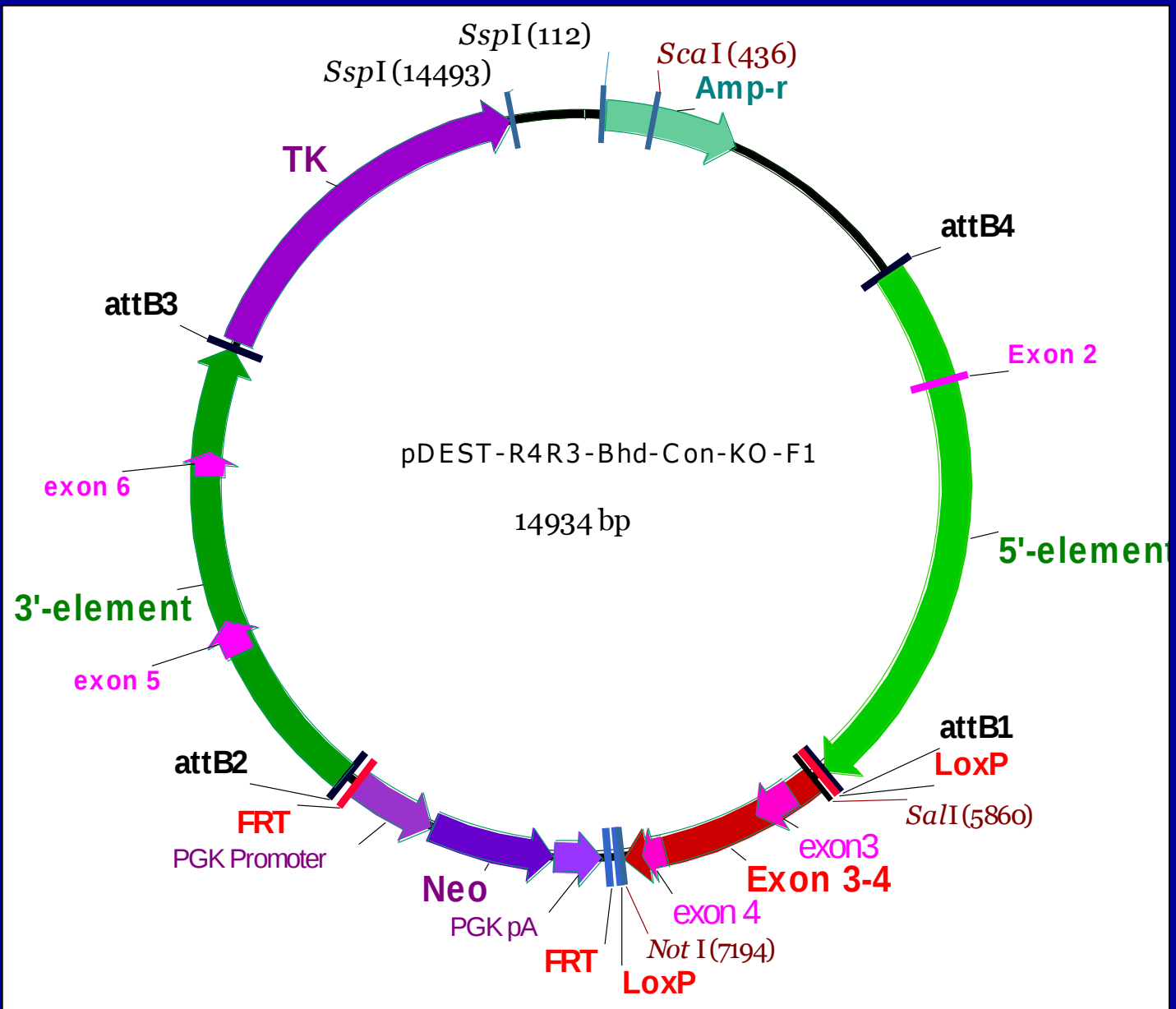
1. 基因突变技术 (物理诱变、化学诱变…)： 无法控制突变位点
2. 转基因技术 (T-DNA插入)： 不能精确控制外源基因插入位点
3. RNA干扰技术 (dsRNA、Artificial miRNA)： 对靶基因抑制不彻底、不特异
4. 核外遗传技术 (质粒、人工染色体)： 难以稳定的遗传
5. 同源重组技术 (e. g. 传统的基因敲除小鼠)： 费时费力费钱，难以大量应用并引起系列生物伦理的问题…
6. 基因编辑技术 (近十年出现)： 可以帮助科研人员对各种细胞和各种生物体内的几乎任意基因进行人工操作。

同源重组技术敲除基因



同源重组技术十分耗时耗力





基因组编辑技术

- **基因组编辑技术 (genome editing)** 是一种可以在基因组水平上对DNA序列进行改造的遗传操作技术。也称为基因打靶 (Gene targeting)。
- 技术的原理是通过精确识别靶细胞DNA片段中靶点的核苷酸序列，利用核酸内切酶，对预定的基因组某位置的靶点DNA序列进行切割，切断的DNA在被细胞内的DNA修复系统修复过程中会产生变异，从而实现了对靶细胞DNA片段的精确编辑或定点改造基因组的目的。

基因组编辑技术

历史

- 1993年前ZFN就已开始出现，迄今已发展了20多年才有今天的成就；
- 2009年底出现的TALEN似乎一夜之间呈现出取代ZFN的架势；
- 2012年底CRISPR/Cas已显现出取代TALEN的巨大潜力，在2013年成为现实；
- 2016年，韩春雨等又建立了NgAgo-gDNA基因编辑技术。

成就

1. 2011年：ZFN，TALEN和Meganuclease—2011年度方法
2. 2012年：TALEN — 2012年十大突破
3. 2013年：CRISPR/Cas被认为是诺贝尔奖级别的成果，华人科学家张峰已因此获得生物学大奖。

应用

- 基因组巡航导弹，分子剪刀，DNA外科医生
- 定点突变，定点修饰（包括得到转基因），转录调控
- 基因治疗（如遗传病）

基因编辑的三个主要方法

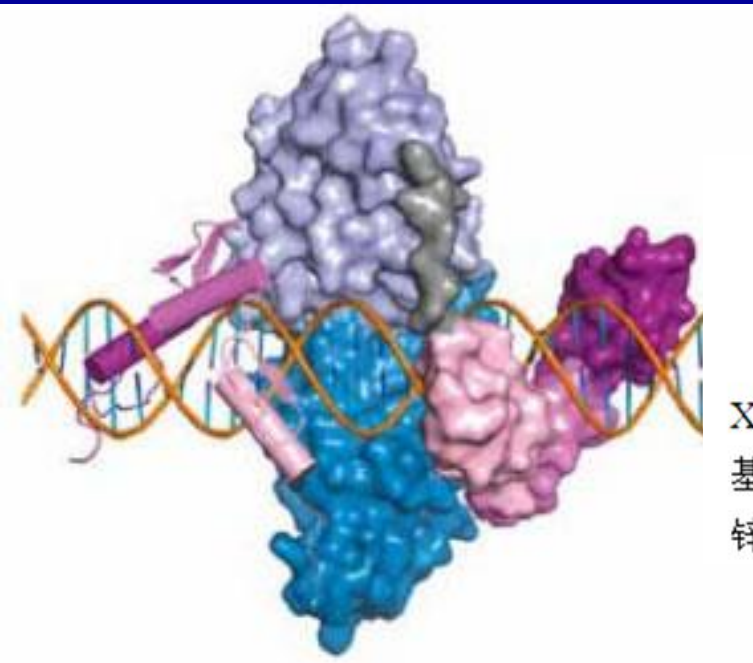
- **ZFN** (Zinc-finger nucleases, ZFN)
- **TALEN** (transcription activator-like effector nucleases)
- **CRISPR/Cas** (clustered regulatory interspaced short palindromic repeat /Cas-based RNA-guided DNA endonucleases)

特异性
DNA识别域

+

非特异性
核酸内切酶

ZFN原理



Linker -1123456
TGEKPYKCPECCKSFSXXXXXXXXHQRTH

锌指结构域的氨基酸序列

X 位置为可变的氨基酸，改变这些氨基酸可以识别不同的三联体碱基，其上部表示氨基酸在 α -螺旋上的位置。TGEK 序列用于连接两个锌指结构域。

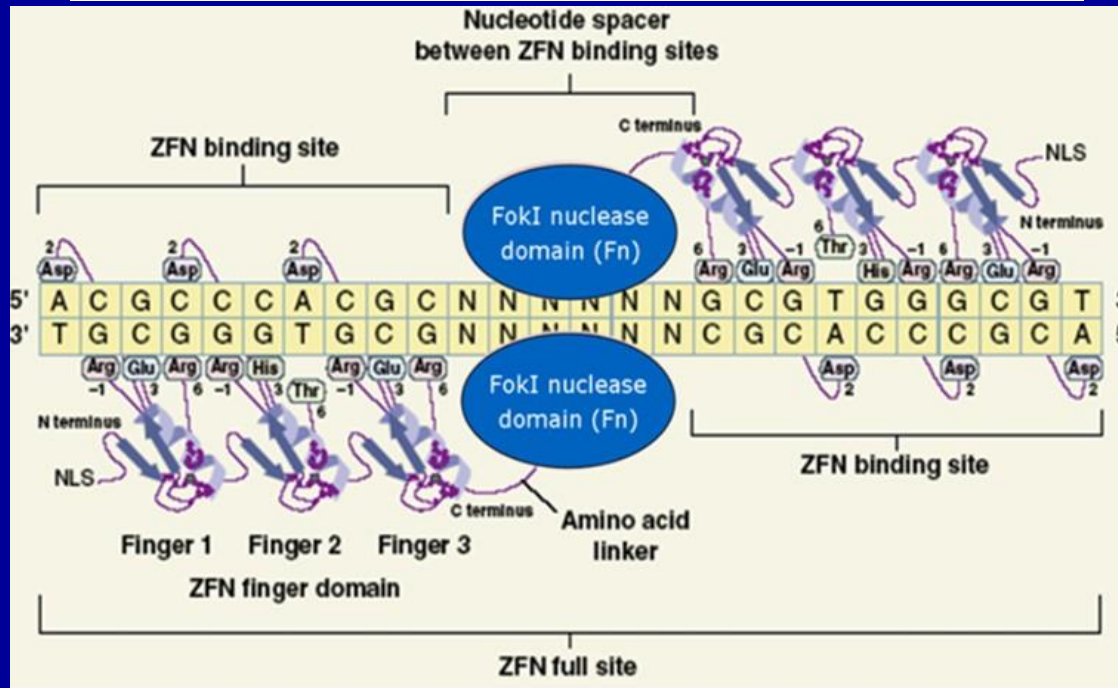
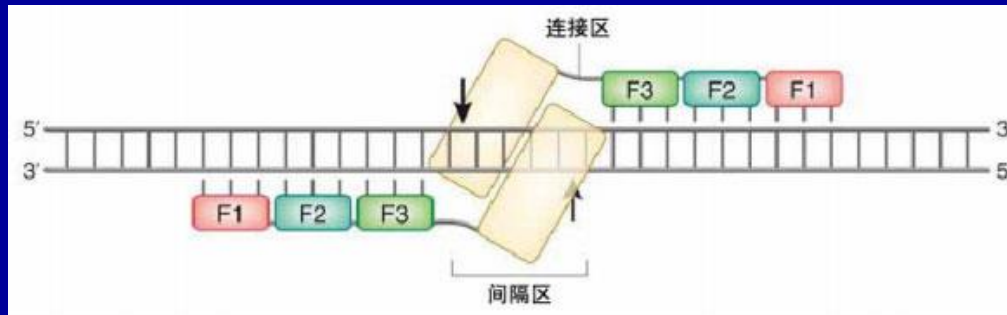
ZFN=蛋白的DNA识别域+核酸内切酶

DNA识别域：

由一系列 Cys2-His2锌指蛋白 (zinc-fingers) 串联组成 (一般3~4个)，每个锌指蛋白识别并结合一个特异的三联体碱基。

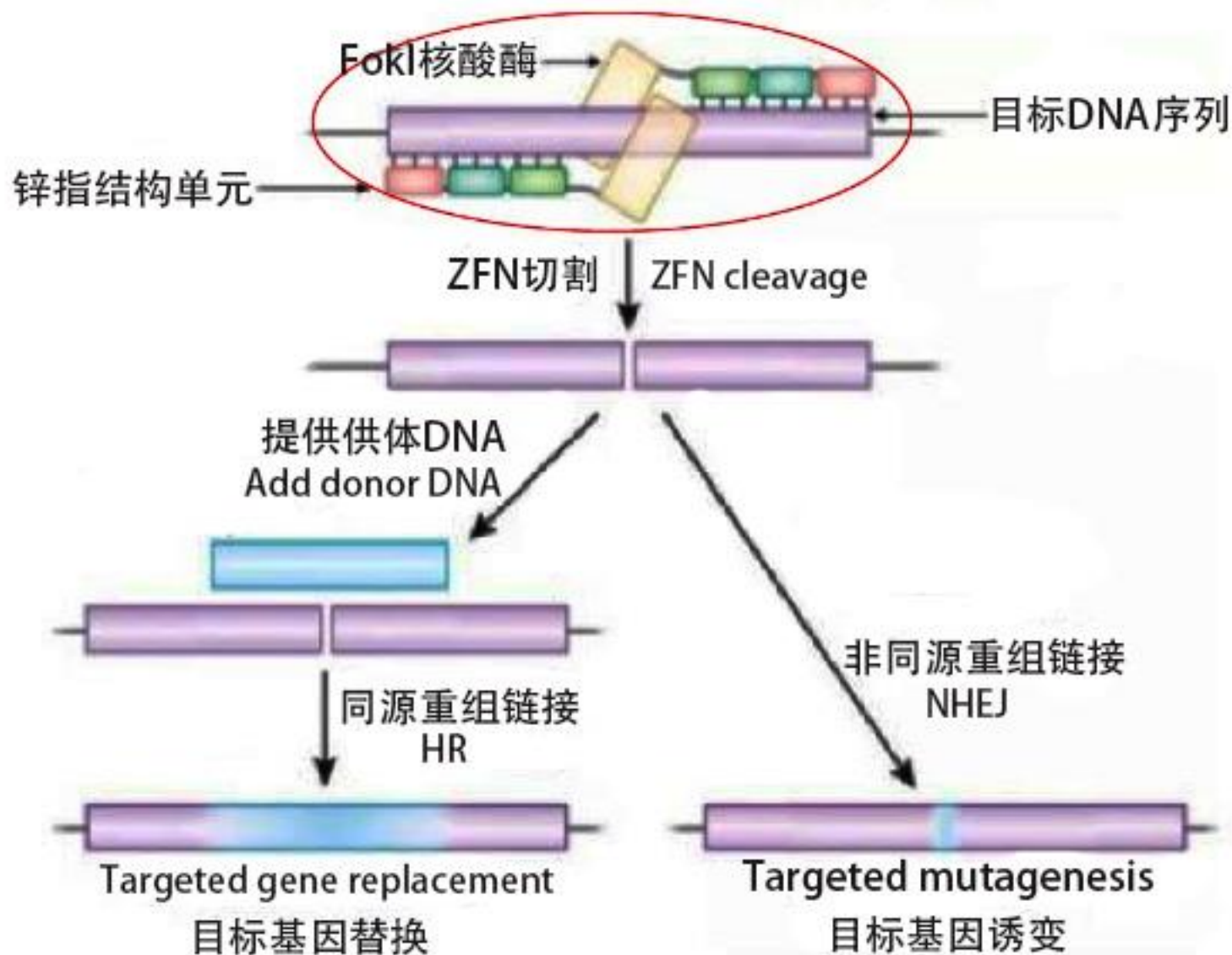
核酸内切酶：

非特异性核酸内切酶Fok I，形成二聚体时切割双链DNA



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₁₁工作空间。

ZFN 的工作原理



ZFN技术的演化

1983年，首次在非洲爪哇转录因子TFIIA中发现，锌指结构是由多个半胱氨酸和/或组氨酸与锌离子螯合组成四面体结构；

1993年，Desjais和Berg等在序列分析的基础上，通过天然锌指的组合与匹配，得到能够识别特异DNA位点的全新锌指蛋白，在锌指蛋白设计方面，做出了开创性的工作；

1996年，Kim和Cha等人发现，将已知的识别特异DNA序列的锌指结构与限制性内切酶FokI中无特异性识别作用的切割结构域通过一个连接子串连表达，可以组合成新的限制性内切酶，即锌指嵌合核酸酶；

2007年，Miller等提供改变FokI的结构，一个FokI上的 E 突变为 K (490 位) 和 I 突变为 K (538位)，另一个FokI的 Q 突变为E (486 位) 及I突变为 L (499 位)，使得靶位点两侧的两个Fok I 不同，它们的特异性结合大大增强，避免了单个同一个锌指核酸酶的结合与切割，大大减少了对基因组非靶位点的破坏性；

2010年，Guo等用体内进化的方法产生了一个新的ZFN-Sharkey，体外分析表明比野生型的活性提高了15倍，从而有利于ZFN在医学和生物学领域的应用；

2016年，基于ZFN技术的爱之病治疗药物（剥离T细胞表面的CCR5基因）已进入三期临床试验阶段。

ZFN技术的缺陷

- DNA识别域虽然具有较强的特异性识别能力，但是其识别的序列长度比较有限。
- 因为ZFN剪切的过程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；
- 如果出现脱靶，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，该手段在细胞内部操作的精确程度不足，则可能会引起相关基因突变，引发癌症等。
- 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。而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。
- 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。

TALEN

TALEN (Transcription activator -like effectors) :一种源于植物致病菌的靶向基因操作技术。

TALE (Transcription Activator-like Effector) 是由植物致病细菌 Xanthomonas sp 。 (黄单胞菌) 分泌的一类具有转录激活功能的蛋白 , 该种蛋白通过其内部保守的重复氨基酸序列 (即DNA结合区域) 与植物宿主基因启动子区 (靶位点) 的相应核苷酸序列发生高度且简单恒定的特异性结合 (一个模块单元识别一个碱基) , 并激活基因表达。

TALEN (Transcription Activator-like Effector Nuclease)是一种人工改造的限制性内切酶 , 是将TALE的DNA结合区域与限制性内切酶 (FokI)dDNA切割域融合而得到。

通过组合各类模块 , 可对任意核苷酸序列进行靶向特异性敲除或内源基因的表达调控。

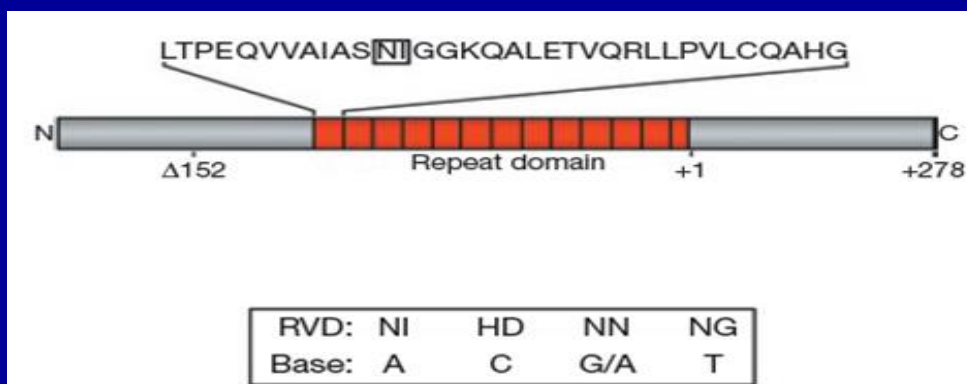
目前已在人、大鼠、小鼠、猪、羊、斑马鱼、拟南芥及酵母等多类物种中得到成功应用。

TALEN原理

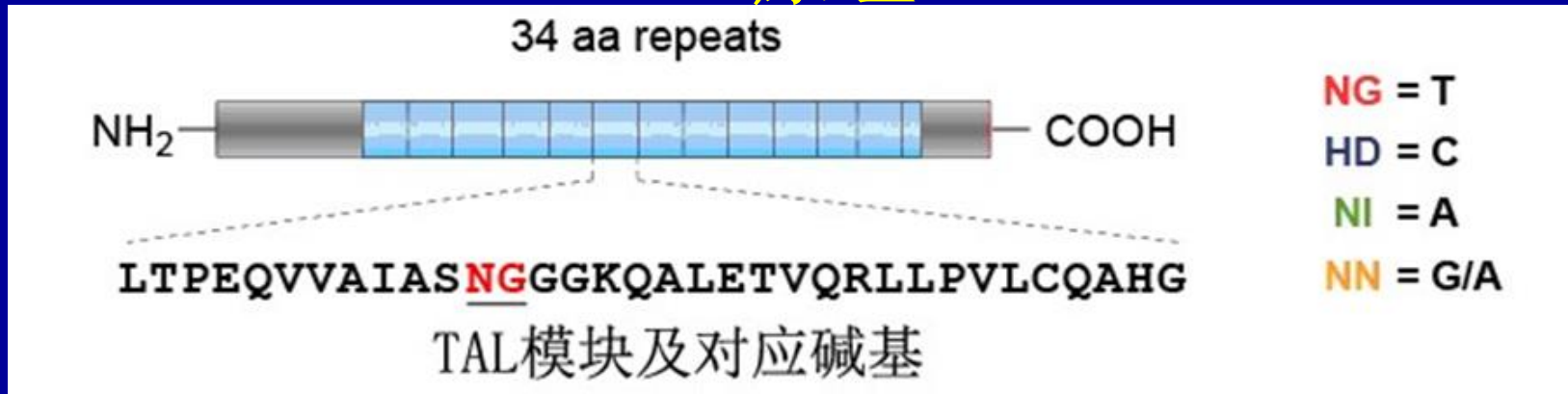
典型的 TALEN 由一个包含核定位信号（NLS）的 N 端结构域、一个包含可识别特定 DNA 序列的典型串联 TALE 重复序列的中央结构域，以及一个具有 FokI 核酸内切酶功能的 C 端结构域组成。

DNA 识别域：由一系列 TAL 蛋白串联组成（一般 20 个左右），每个 TAL 蛋白识别并结合一个对应的碱基。

核酸内切酶：非特异性核酸内切酶 FokI，形成二聚体时切割双链 DNA



TALEN原理



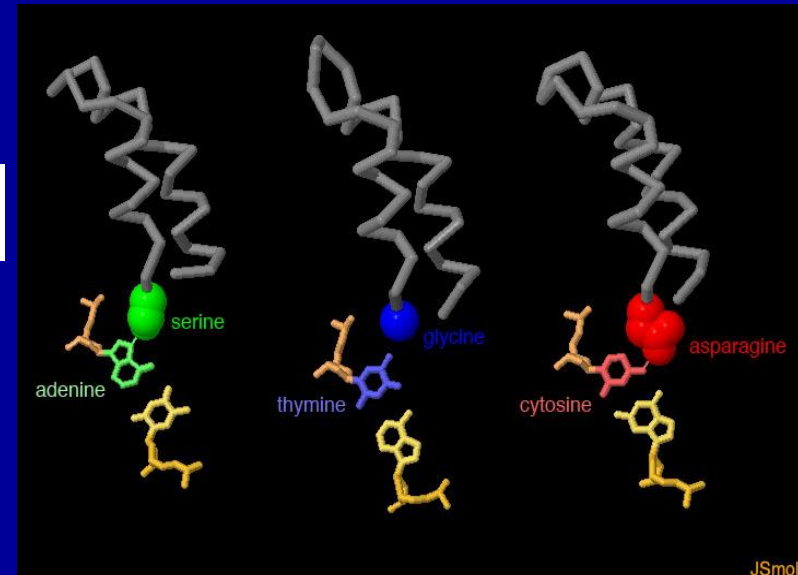
全长为34aa的Tal 蛋白的第12、13位氨基残基可以特异性识别核苷酸碱基。

NG可以识别T，HD可以识别C，NI可以识别A，NN可以识别G或A。多个Tal蛋白串联在一起，就可以构成可以识别目的片段的Tale蛋白。

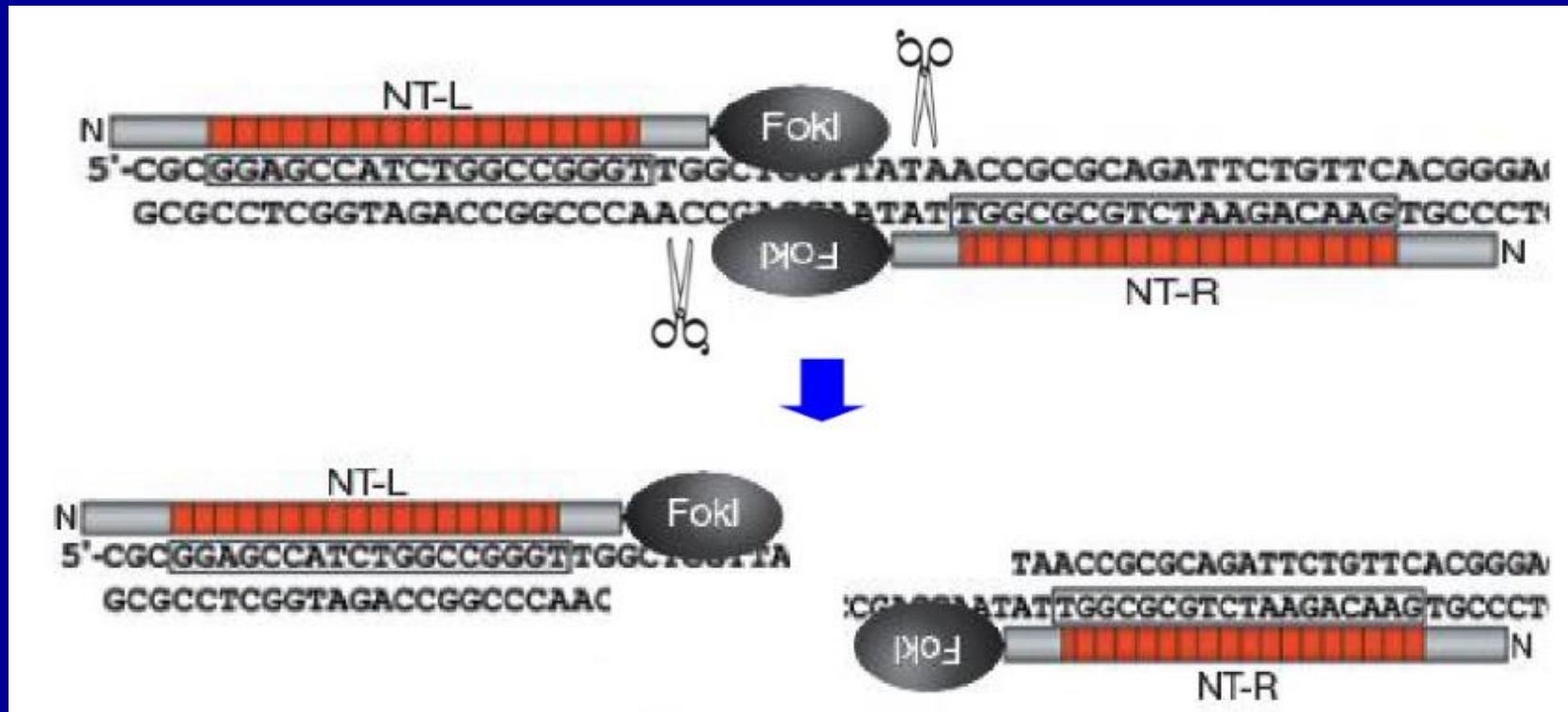
TALEN = DNA识别区 + 核酸内切酶



形成二聚体时，即可切割目的DNA。



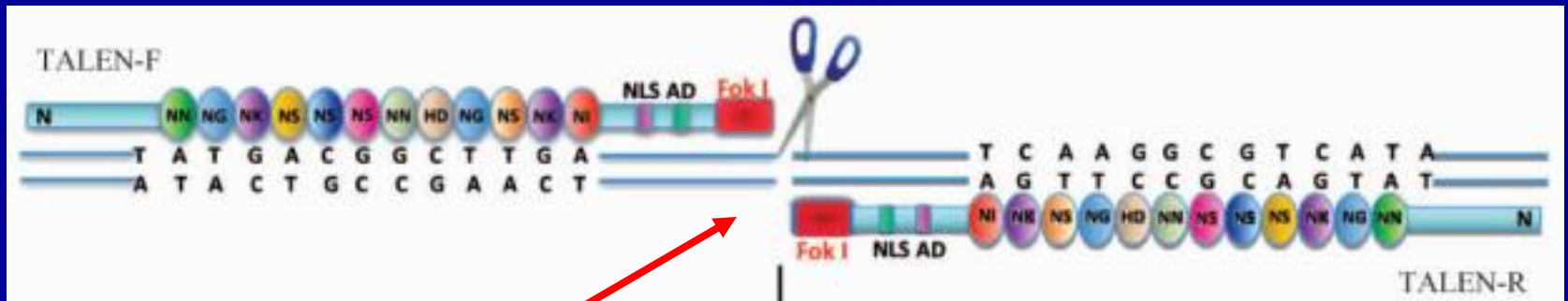
TALEN原理



TALEN的特异性打靶的原理：

一对可以特异性识别靶基因的TALE，定位到需要编辑的基因组区域；
然后非特异性核酸内切酶FokI切断双链DNA从而造成DNA双链断裂（double-strand break，DSB）；
再通过DNA的自我修复，引起碱基的缺失或突变，从而引起基因突变。
FokI只有在形成二聚体的时候才有内切酶活性。

TALEN介导的基因编辑



TALEN质粒在共转入细胞后，表达两个融合蛋白，分别与靶位点特异结合，由于两个TALEN融合蛋白中的Fok I临近，形成二聚体，发挥非特异性内切酶活性，在两个靶位点之间剪切DNA。

形成DSB时，同源重组可将需要敲入的序列组合入基因组。

TALEN的技术特点

- 任意敲除，无基因序列、细胞、物种限制，永久失活
- 成功率高，在保证转染效率的前提下，有效性可达95%以上
- 打靶效率高，可完全替代RNAi技术
- 脱靶率低，尚未发现明显脱靶效应
- 技术简单，只需按照常规构建质粒方法构建Talen质粒

TALEN技术演化

1989年，植物病原体黄单胞菌属 (Xanthomonas spp.) avrBs3 基因被克隆；

2009年，TAL effector 氨基酸序列与核酸靶序列的密码被破译：Boch等和Moscou等关于黄单胞菌效应子蛋白TALE (transcription activator-like effectors)与寄主靶基因DNA特异性识别分子密码的破解，使植物基因组定点改造呈现出新的曙光；

2011年8月，Sangamo BioSciences公司和哈弗大学的两个研究小组分别利用TALEN技术进行了基因组靶向修饰相关研究，并成功敲除了人类细胞靶向基因和调控内源性基因的转录。

CRISPR/Cas 系统

CRISPR是生命进化历史上，细菌和病毒进行斗争产生的免疫武器，简单说就是病毒能把自己的基因整合到细菌内，利用细菌的细胞工具为自己的基因复制服务，细菌为了将病毒的外来入侵基因清除，进化出很多成簇的、规律间隔的短回文重复序列(既CRISPR序列)和CRISPR相关基因，这一序列首先由日本学者于1987年首次发现⁽¹⁾，于2002年被Jansen等将正式命名。

由这些序列和基因组成的系统我们称之为CRISPR/Cas系统，而在这个系统中常用到的核酸内切酶为Cas9,所以通常称该系统CRISPR/Cas9 系统。利用这个系统，细菌可以不动声色地把病毒基因从自己的染色体上切除，这是细菌特有的免疫系统。

CRISPR-Cas系统的发展历史

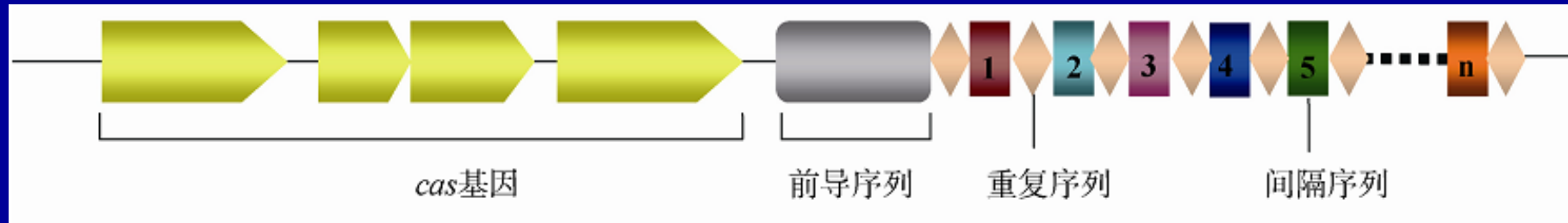
CRISPR-Cas系统的研究历史

- ❑ 1987 年，日本课题组在K12 大肠杆菌的碱性磷酸酶基因附近发现串联间隔重复序列，随后发现其广泛存在于细菌（40%）和古细菌（90%）的基因组中。2002 年，正式将其命名为成簇的规律间隔的短回文重复序列
- ❑ 2005 年发现CRISPR 的间隔序列(spacer)与宿主菌的染色体外的遗传物质高度同源，推测细菌可能通过CRISPR 系统可能以类似于真核生物的RNAi 方式抵抗外源遗传物质的入侵。
- ❑ 2007 年，Barrangou 等首次发现细菌可能利用CRISPR 系统抵抗噬菌体入侵；2008 年，Marraffini 等发现细菌CRISPR 系统能阻止外源质粒的转移，首次利用实验验证了CRISPR 系统的功能
- ❑ 2013 年初，MIT 的研究组首次利用CRISPR/Cas9 系统对人293T 细胞*EMX1* 和*PVALB* 基因以及小鼠Nero2A 细胞 *Th* 基因实现了定点突变。同年Mali 利用CRISPR/ Cas9 在人293T 细胞和K652 细胞基因的靶位点形成双链或单链的切口，从而激活细胞的DNA 修复机制高效介导外源基因定点插入。

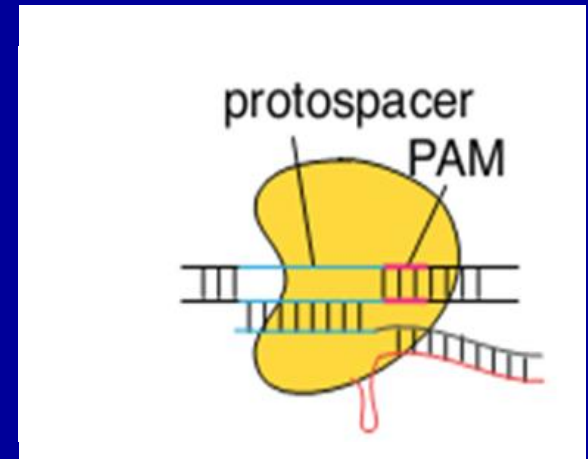
德国学者Helmholtz、Doudna、Charpentier分别对不同的CRISPR及相关系统进行了研究，阐明了DNA空格序列在细菌免疫防御中的机制。目前已经发现了3种CRISPR/Cas 系统，其中依赖Cas9蛋白的CRISPR系统较简单。

CRISPR-Cas系统的结构

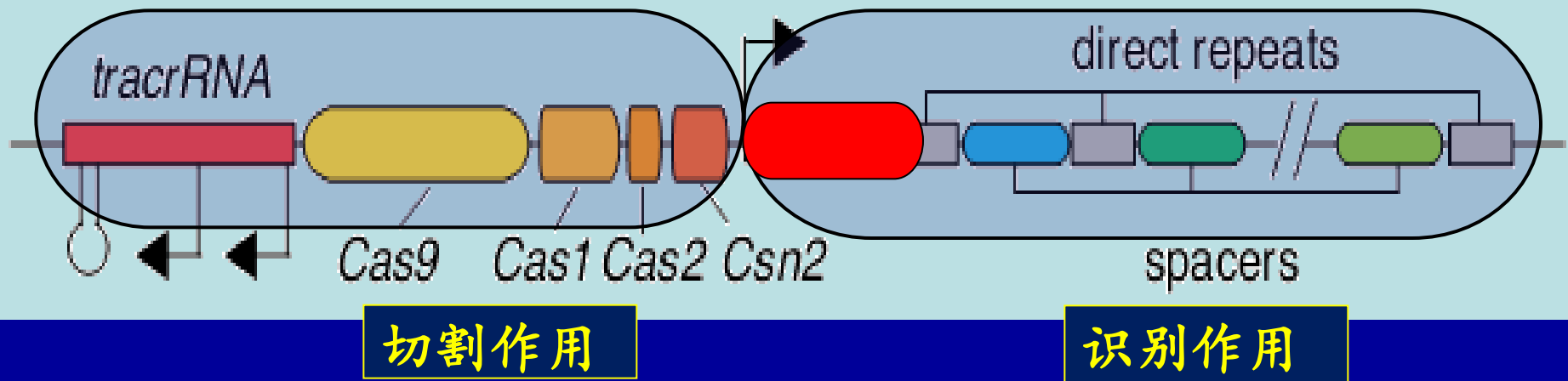
- CRISPR-CAS 系统的组成主要包括: 由不连续的**重复序列R**(repeat) 与长度相似的**间区序列S**(spacers) 间隔排列而成的CRISPR 簇, **前导序列L**(leader) 以及一系列**CRISPR相关蛋白基因cas**。



Cas蛋白是一种双链DNA核酸酶，能在guide RNA引导下对靶位点进行切割。它与folk酶功能类似，但是它并不需要形成二聚体才能发挥作用。



CRISPR/Cas系统的基本结构

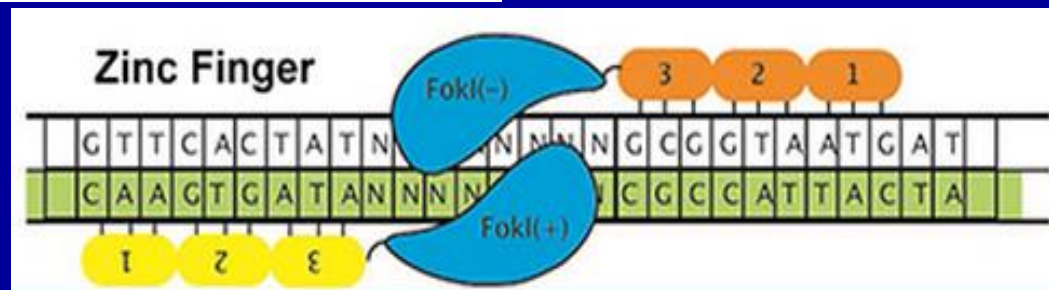
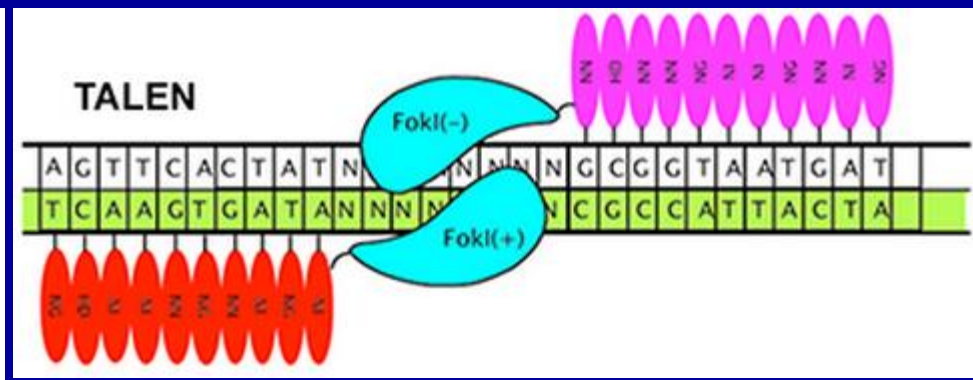
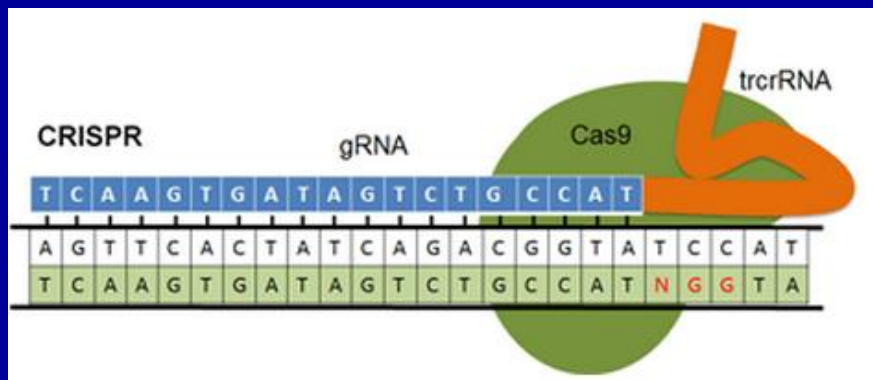


CRISPR系统通常包括：由不连续的重复序列（ repeats , R ）与长度相似的间区序列（ spacers , S ）间隔排列而成的CRISPR簇，前导序列(leader , L)以及一系列的CRISPR相关蛋白基因(cas)。

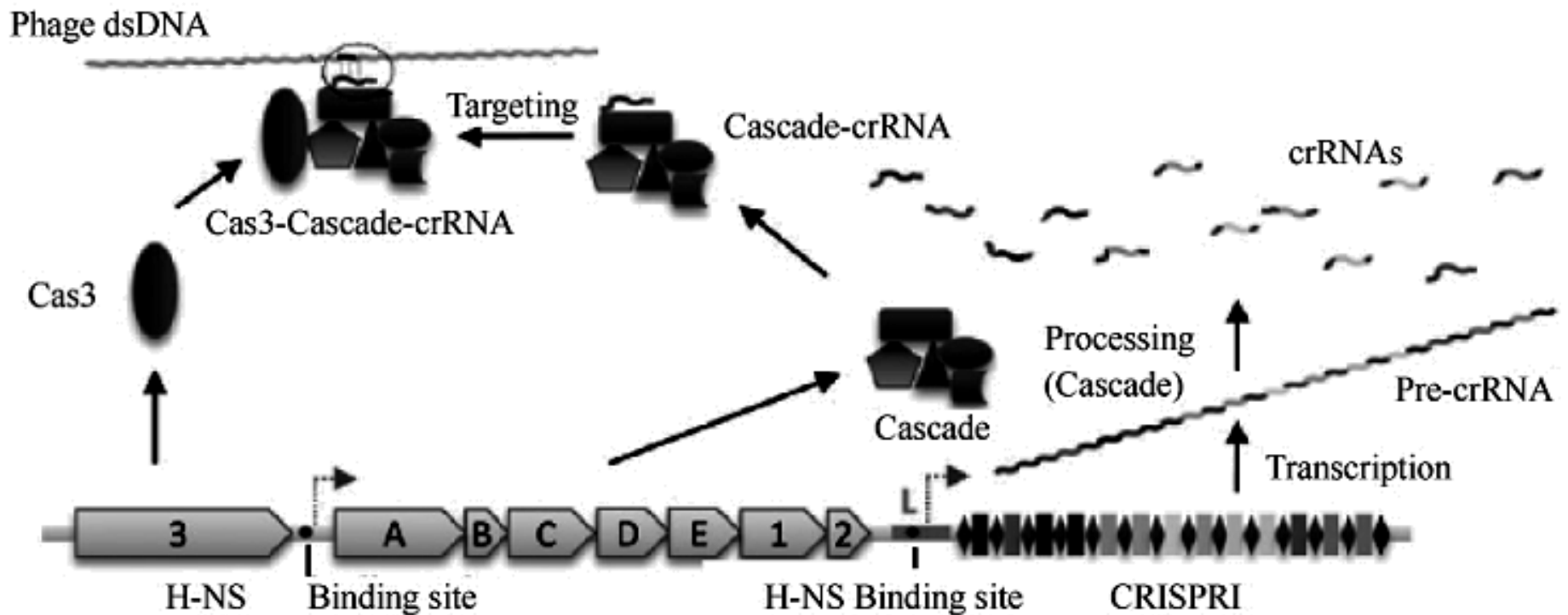
Cas (CRISPR associated) ：存在于CRISPR位点附近，是一种双链DNA核酸酶，能在guide RNA引导下对靶位点进行切割。它不需要形成二聚体才能发挥作用。

CRISPR/Cas系统与ZFN及TALEN的区别

Cas(CRISPR associated)存在于CRISPR位点附近，是一种双链DNA核酸酶，能在guide RNA 引导下对靶位点进行切割。它与fokI酶功能类似，但与ZFN和TALEN不同的是它并不需要形成二聚体才能发挥作用。



CRISPR的转录与加工



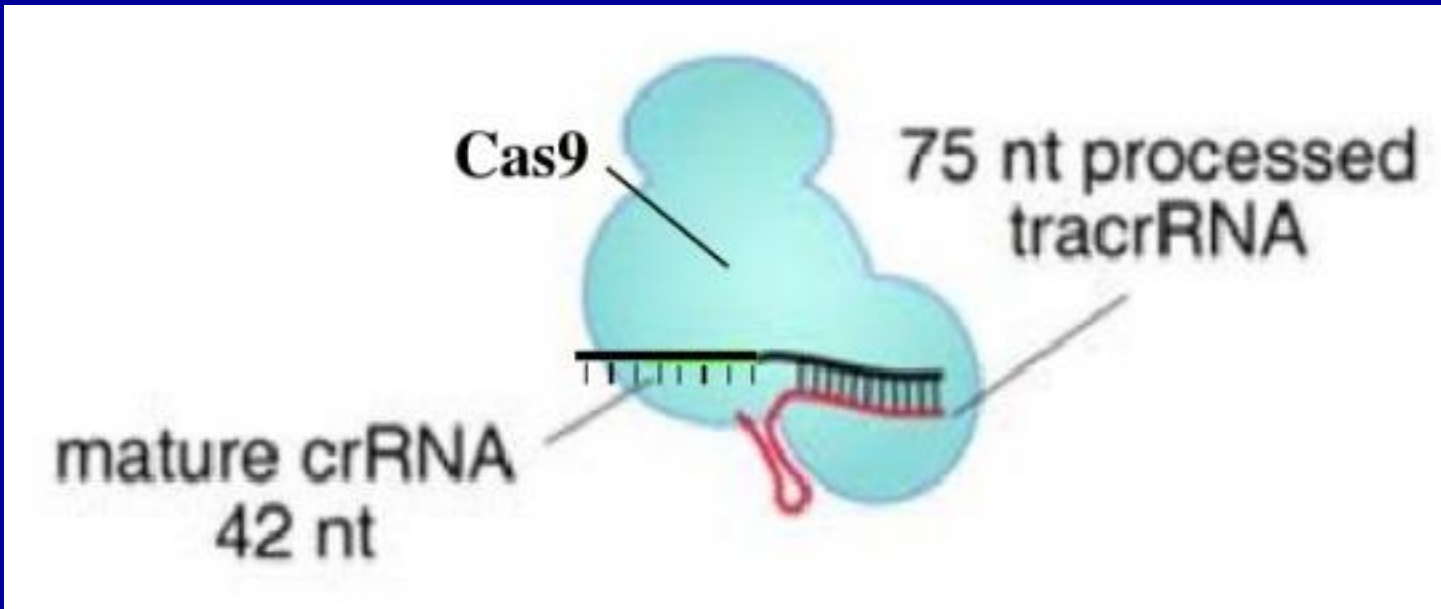
CRISPR簇首先转录成长的转录体，即（crRNA），然后逐步被加工成小的 crRNA。

CRISPR-Cas9编辑系统

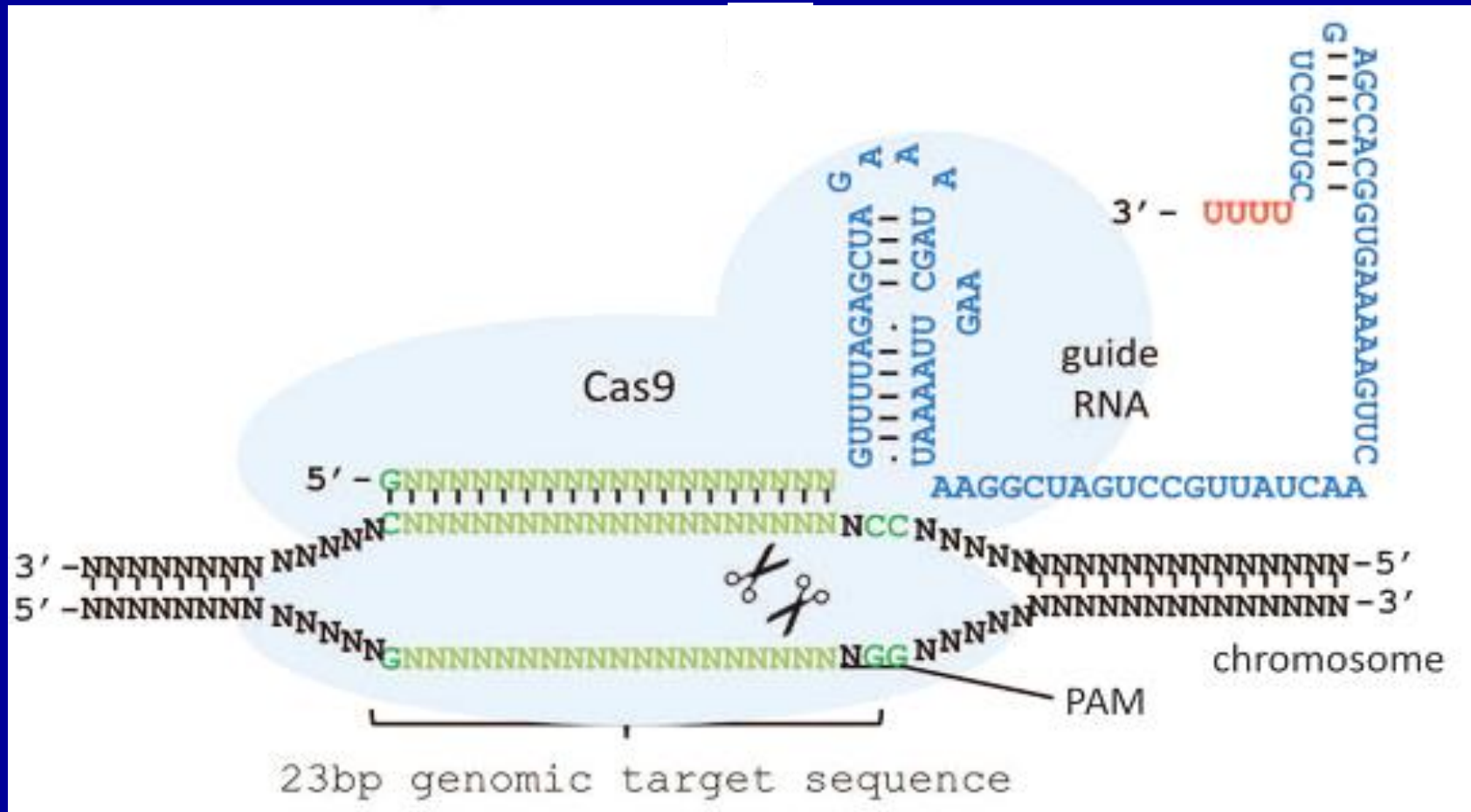
CRISPR-Cas9 = DNA识别区 + 核酸内切酶

DNA识别区：向导RNA，由两部分组成，即CRISPR RNA (crRNA)和反式作用CRISPR RNA (transacting CRISPR RNA , tracrRNA)。 crRNA 一端与靶标核苷酸互补结合，另一端与tracrRNA互补结合。

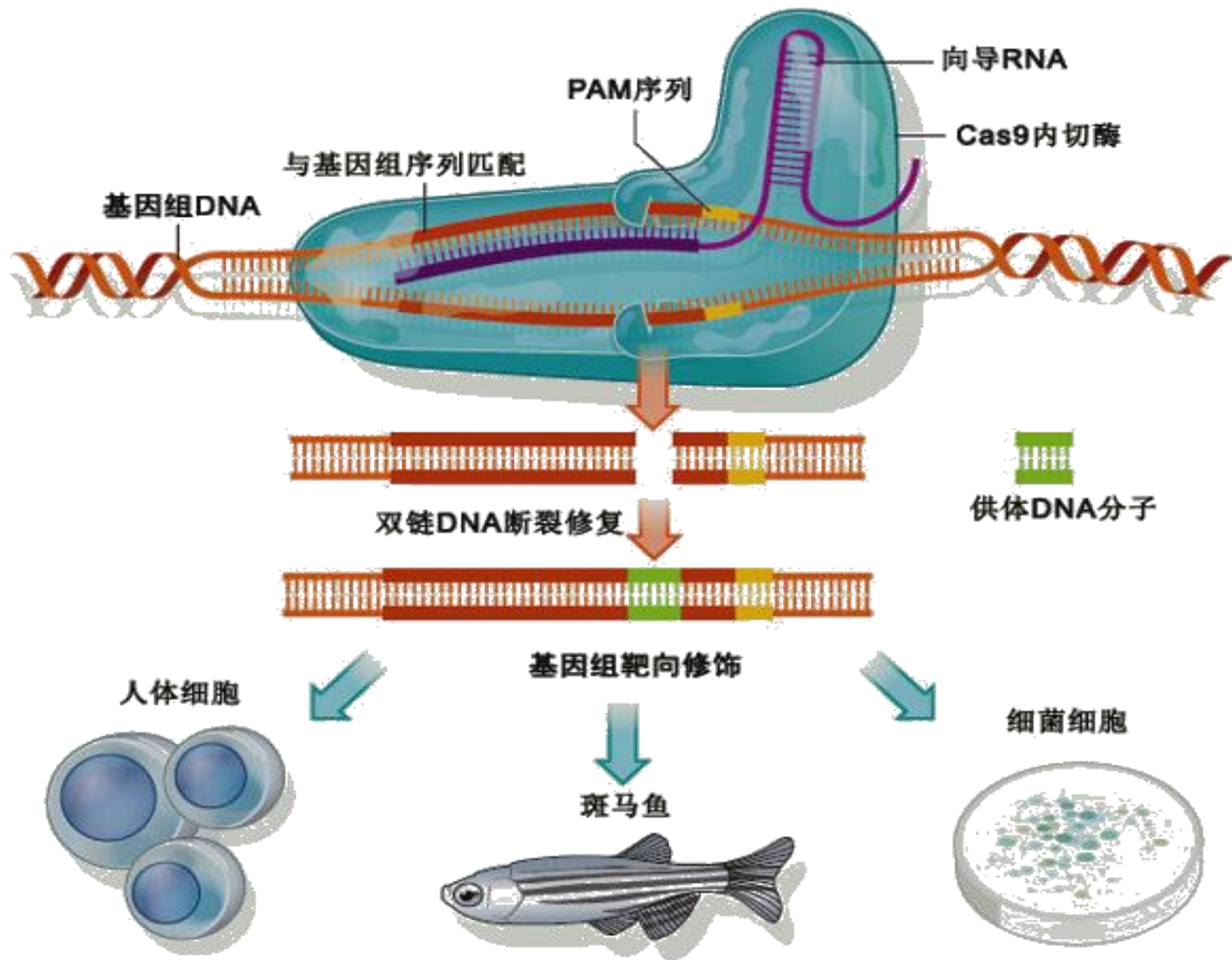
核酸内切酶：Cas9 核酸内切酶。



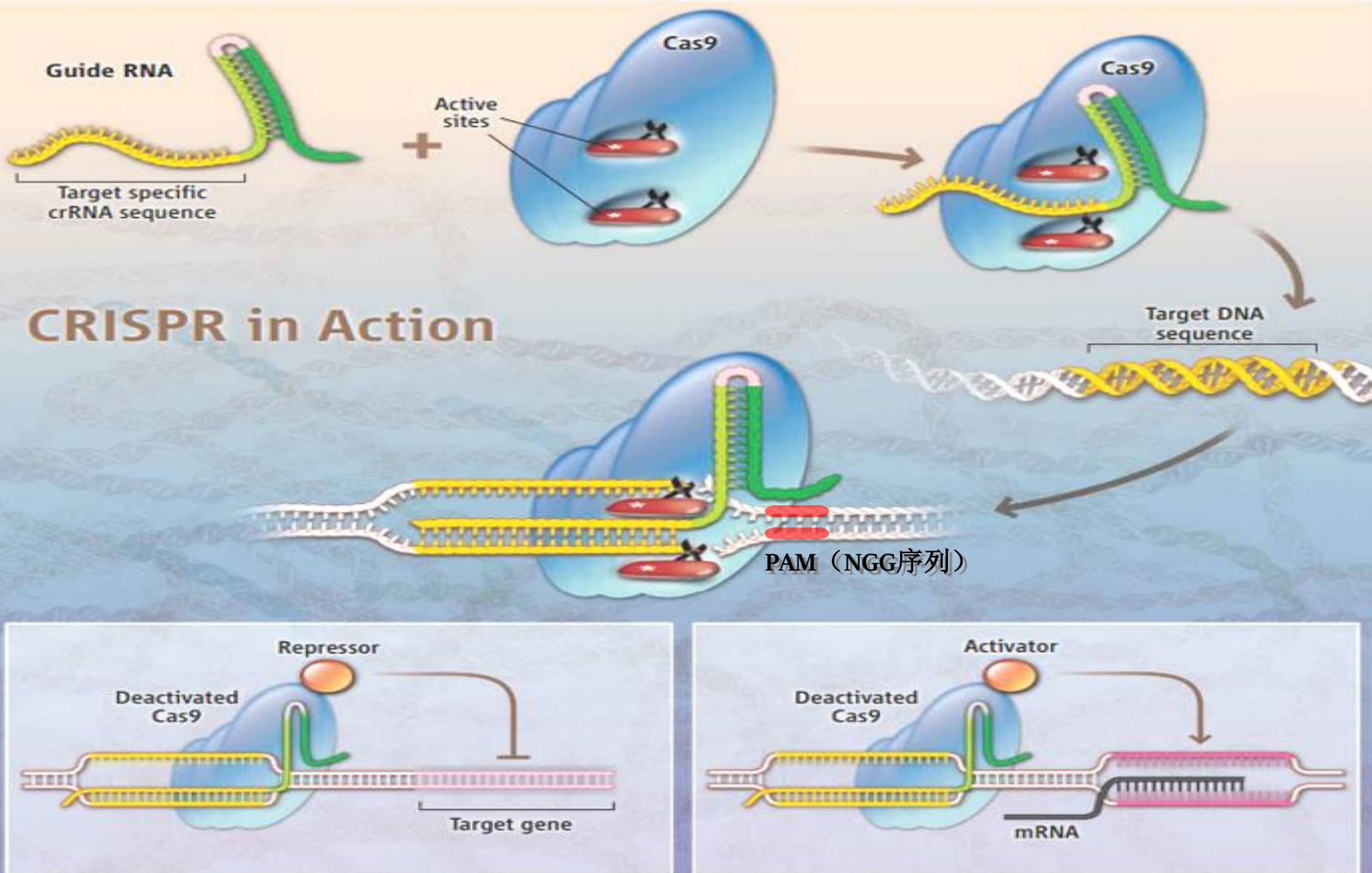
CRISPR-Cas9的作用原理



CRISPR/Cas 9的应用

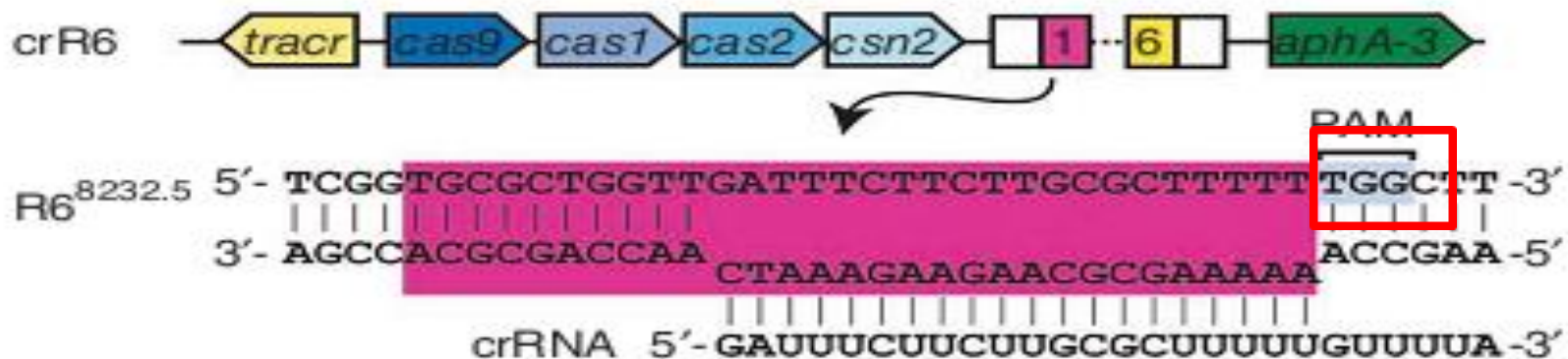
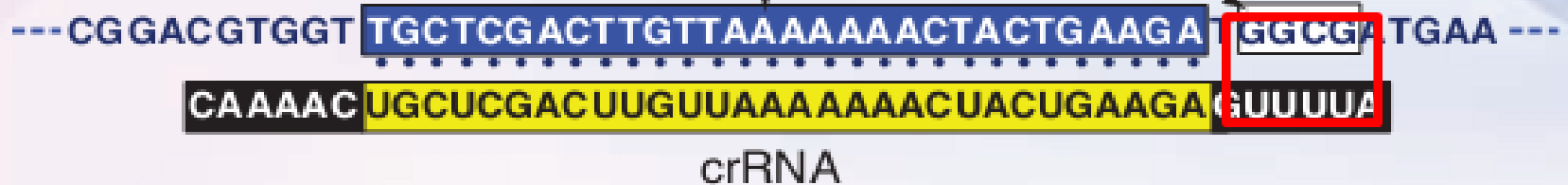
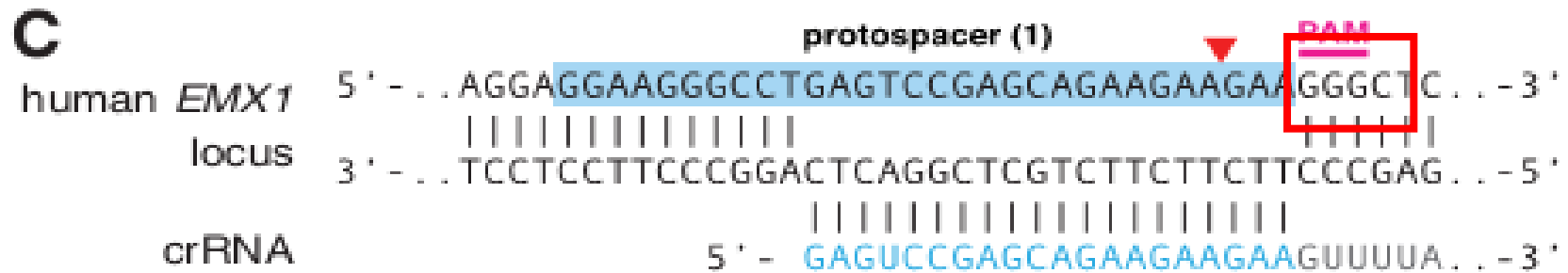


CRISPR/Cas 9作用机理



CRISPR/Cas 9系统靶向要求

最主要的要求：PAM (protospacer-adjacent motif) 为NGG。
在人类基因组中，平均每8bp就出现一个NGG PAM。



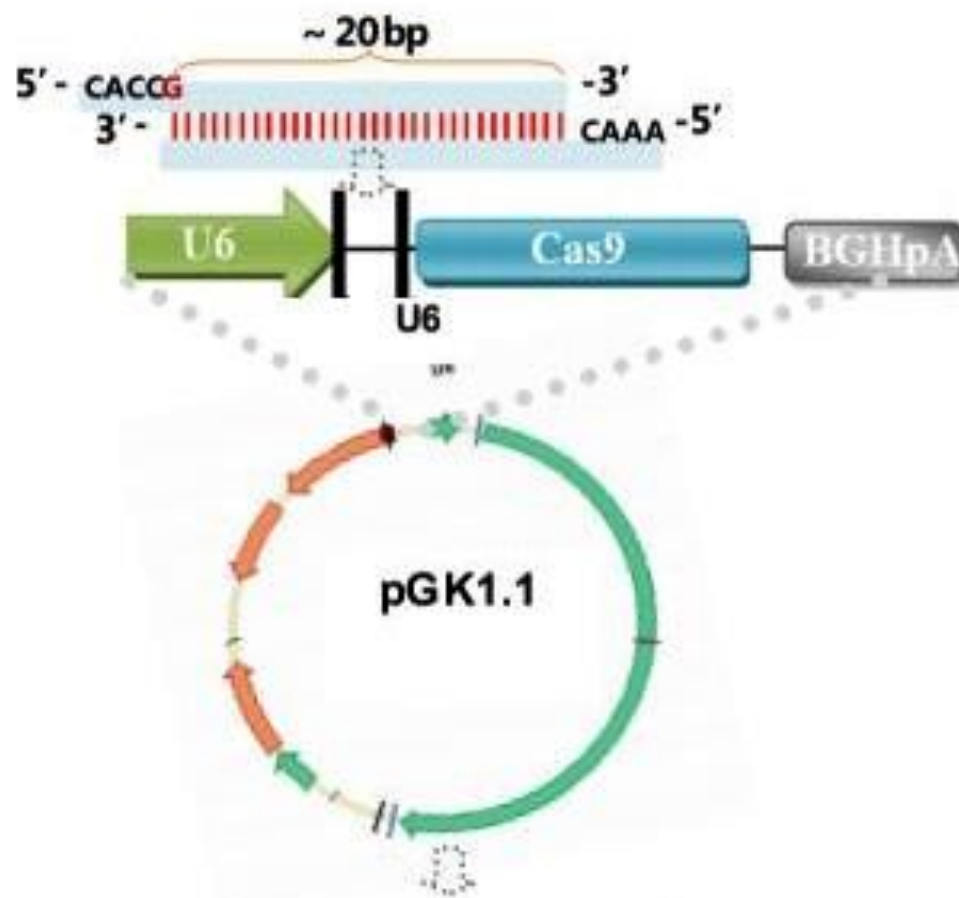
CRISPR/Cas 9基因编辑实验流程图

1, CRISPR-Cas9 基因编辑实验流程图如下:

1

寡核苷酸序列

设计 2 条单链 oligo 序列; 退火形成双链 DNA。



2

将双链 DNA 连接到载体中。

3

G10感受态细胞

转化 G10 Competent Cell; 筛选阳性克隆; 测序验证序列;



CRISPR/Cas 9的技术应用

- ◆ 应用于基因的敲除、敲入以及基因沉默、激活等。
- ◆ 能够在真核细胞中发挥作用，使真核基因组中的特异性位点发生双链断裂。



中国科学院遗传
所高彩霞团队：
成功地使三种水
稻基因失活。

在模式生物中的应用：成功地对线虫（左）、斑马鱼胚胎（中）和果蝇进行遗传学改造，获得更粗短的线虫，腹部组织体积更大的斑马鱼胚胎，和眼睛颜色更深的果蝇，表明CRISPR技术在动物基因改造方面有巨大应用潜力。



CRISPR/Cas 9的技术应用



近日，中国科学家利用基因编辑技术——CRISPR/Cas9，对抑制狗骨骼肌生长的基因（MSTN）进行了敲除，培育出两只肌肉发达的“大力神”狗，成功构建了世界首个基因敲除狗模型。

三大基因修饰技术比较

	TALEN	ZFN	CRISPR/Cas9
靶点DNA序列的识别区域	重复可变双残基 (RVD) 的重复	锌指 (ZF) 结构域	CRISPR RNA (crRNA) 或向导RNA (gRNA)
DNA的剪切	FokI核酸酶结构域	FokI核酸酶结构域	Cas9蛋白
典型核酸酶的构建当时	8-31个重复可变双残基的拼接	通过搜索各类ZF组合 数据库，拼接3-4个ZF 结构	gRNA的寡核苷酸合成 和分子克隆 (或RNA合成)
所识别的靶位点大小	(8-31 bp) * 2	(9或12 bp) * 2	20bp + "NGG" * 1
最小模块识别碱基数	1	3	1
优点	设计较ZFN简单、特异性高	平台成熟、效率高于 被动同源重组	靶向精确、脱靶率 低、细胞毒性低、廉 价简便
缺点	细胞毒性、模块组装 过程繁琐、需要大量 测序工作、一般大型 公司才有能力开展、成本 高	设计依赖上下游序 列、脱靶率高、具有 细胞毒性	靶区前无PAM则不能 切割、特异性不高、 NHEJ依然会产生随机 毒性

不同基因组编辑技术比较

	ZFN	TALEN	CRISPR
识别模式	蛋白质-DNA	蛋白质-DNA	NA-DNA
识别长度	(3-6) × 3 × 2 bp	(12-20) × 2 bp	20 bp
识别序列特点	3 bp 为单位	以5'前一位为T	3'序列为NGG
特异性	较高	一般	一般
构建难易	难度较大	容易	容易
细胞毒性	大	较小	较小

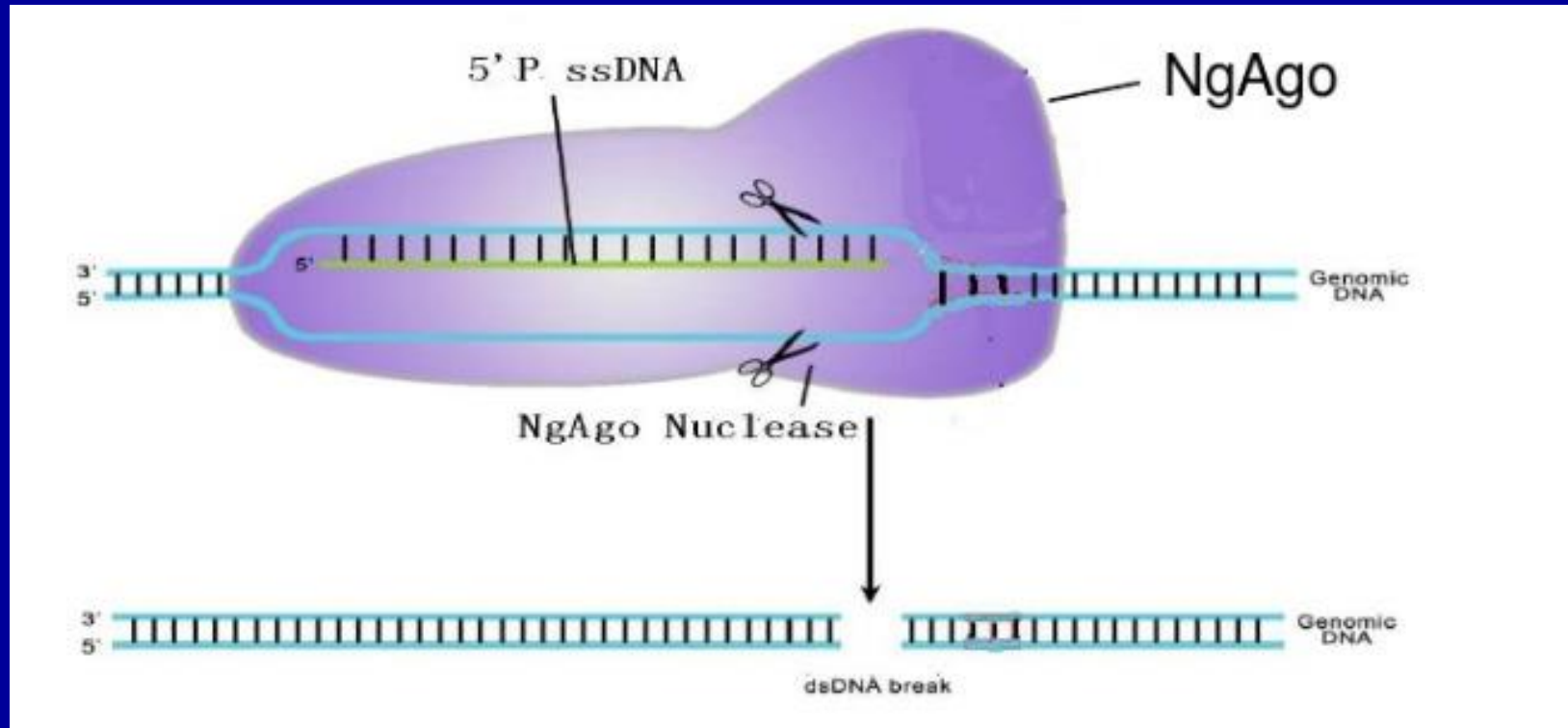
CRISPR/Cas 9系统的优势

- 操作简单，靶向精确性更高。sgRNA靶向序列和基因组序列必须完全匹配，Cas9才会对DNA进行剪切。编码sgRNA的序列不超过100bp，因此比构建TALENs和ZFNs更简单方便，用于CRISPR的sgRNA识别序列仅需20个核苷酸。
- CRISPR/Cas9系统是由RNA调控的对DNA的修饰，其基因修饰可遗传。
- 基因修饰率高，CRISPRs基因敲入的效率为51%-79%，TALENs的效率为0%-34%。
- 基因调控方式多样，例如敲除、插入、抑制、激活等
- 无物种限制，可实现对靶基因多个位点同时敲除。
- 实验周期短，最快仅需2个月，节省大量时间和成本。(ZFNs一个靶点成本6000 \$)

NgAgo-gDNA基因编辑技术

NgAgo-gDNA: 一种通过5'端磷酸化的单链DNA介导NgAgo蛋白对靶向基因进行修饰的技术;

NgAgo (*Natronobacterium gregoryi* Argonaute) 是在格式嗜烟碱杆菌中发现的一种DNA介导的双链核酸内切酶。



NgAgo-gDNA编辑技术的特点

- 1、向导设计制作简便：可以像合成PCR引物一样合成短单链DNA向导；向导可直接转染细胞和组织而无需构建向导表达载体。由于向导核酸是DNA而非RNA，因此避免了RNA易于形成复杂的二级结构而带来的失效或者脱靶效应。
- 2、可编辑基因组内任何位置：Cas9基因组的靶点选择受到PAM区和富含GC区的限制。而NgAgo对靶点选择没有限制。对基因组任何位置都能有效引入双链断裂。
- 3、“Cas9”的取材范围有限，NgAgo的优势之一，就是大大扩充了基因编辑的取材范围。
- 4、精确性增高，脱靶率降低。NgAgo要结合24个碱基，比Cas9结合的19个碱基要长5个碱基，理论上精确性要提高1024倍。
- 5、*Natronobacterium gregoryi*（格式嗜盐碱杆菌）的Argonaute蛋白实现了DNA引导的基因组编辑，并发现NgAgo作为一种DNA介导的核酸内切酶，适合人体细胞（37C）中进行基因编辑。

NgAgo-gDNA与CRISPR-Cas9比较

NgAgo-gDNA	CRISPR-Cas9
5' -P-ssDNA 介导	sgRNA 介导
ssDNA无特殊二级结构要求，除靶序列一致的碱基无需其它碱基参与	sgRNA由tracrRNA和crRNA组成，需求固定的二级结构
结合靶序列23-25bp	结合靶序列19-21bp
无需PAM区	需要PAM区
NgAgo蛋白，由887 个氨基酸组成	Cas9蛋白由1368个氨基酸组成
可在哺乳动物细胞水平进行基因组编辑	可在哺乳动物细胞水平进行基因组编辑

Thank you for your attention!

