

## 基因敲除技术发展史简概

Nanfang Peng

基因敲除在研究某个基因功能以及永久改变细胞表型的过程中是一个关键技术。其在生命科学研究以及临床上特别是遗传病的治疗方面都具有令人激动的广阔前景，因此科学家对其的研究探索一直保持高度热情。目前为止，已有的各种基因敲除技术在敲除效率、完成时长以及脱靶效率方面存在差别。

由于缺乏快速地靶定和敲除特定基因的方法，早期基因敲除在基础研究、功能基因组学以及细胞系工程中没有广泛应用，科学家多应用非特异性靶定的方法，如电离辐射、化学物质来诱导体细胞全基因组的随机突变；随后，科学家采用特异性的同源重组来获得某基因突变，但是在这个方法中，哺乳动物细胞系发生随机整合的概率远远高于特异性的同源重组，故后续试验中需要采用数月来筛选数以千计的克隆以获得双等位目的基因敲除的克隆；此外，腺病毒运载系统也被应用于基因组学研究，但是其在获得双等位目的基因敲除克隆上的效率低、时间长，严重阻碍其被广泛应用。随后一系列特定靶基因敲除技术的发现及成熟，大大推动了生命科学和临床医学的发展。



### 一、锌指核酸内切酶 (zinc-finger nucleases, ZNF)

在哺乳动物细胞中，双链 DNA 断裂后通过同源介导的修复 (homology directed repair, HDR) 或者非同源末端连接 (nonhomologous end joining, NHEJ) 来修复损伤部位，两者发生的概率比约为 9:13。目前为止，科学家发现，在果蝇、植物、线虫以及哺乳动物细胞中，位点特异性的锌指核酸内切酶能够造成双链 DNA 断裂，并最终通过同源重组或者非同源末端连接获得目的基因突变体。

锌指最开始是在 *Xenopus* 中发现的在 *Mtazoa* 中最常见的 DNA 结合基序。每一个指状蛋白能够通过  $\alpha$  螺旋识别 3-4 对 DNA 碱基对，因此，多个指状结构可以随机连接起来从而高度特异性地识别大量不同的 DNA 序列。Chandrasegaran 实验室研究发现，由锌指蛋白 DNA 结合结构域（此区域特异性地结合在 DNA 目的区域）以及 Type II S 限制性内切酶的非特异性剪切结构域 Fok I 融合可以形成锌指核酸内切酶。锌指核酸内切酶 Fok I 结构域的二聚化为 DNA 结合结构域发挥功能所必须，因此在通过锌指核酸

内切酶构建稳定敲除的突变体的过程中，需要设计成对的锌指核酸内切酶用于精准地靶定在特异的目的基因序列实现锌指核酸内切酶的二聚化以及 DNA 剪切。其具体过程为：通过在细胞内瞬时表达锌指核酸内切酶，其靶定在目的基因特异性位点并产生双链 DNA 断裂缺口，进而通过细胞内同源重组修复损伤部位进行基因敲除。

相比于非定点突变，锌指蛋白核酸内切酶具有显著优点：

1. 高度特异性：通过随机连接锌指蛋白决定着锌指蛋白核酸内切酶特异性识别的 DNA 序列，即：成对的锌指蛋白核酸内切酶根据 DNA 的方位和空间精确地定位在同一个 DNA 区域，从而产生一个双链 DNA 缺口。
2. Fyodor D. Urnov 等发现，锌指蛋白核酸内切酶诱导的在遗传病突变热点处发生双链 DNA 损伤并且通过同源重组获得永久精确改变的效率高达 20%。

高的同源重组效率以及锌指核酸内切酶特异性识别目的基因的特征使得锌指蛋白核酸内切酶介导的基因编辑成为了研究体细胞遗传学的一个有力工具，科学家也一度认为其可能是人类遗传病基因修正治疗的潜在手段：

1. 锌指蛋白核酸内切酶介导的基因组敲除在生命科学，生物技术以及医学领域将得到广泛应用。科学家可以利用其敲除目的基因从而研究其在生物体中的具体功能；

2. 锌指蛋白核酸内切酶介导的基因组敲除可以帮助科研工作者选择性地删除内含子等非编码的基因结构，从而更为精确地定位其生物学功能；
3. 锌指蛋白核酸内切酶介导的基因组敲除可以显著促进干细胞研究以及基因治疗。医学工作者提出假说：基因突变引发的遗传病可以通过选择性删除干细胞或者体细胞中突变基因来实现基因治疗的效果。

虽然通过编码锌指蛋白核酸内切酶介导的技术，科学家获得了多种目的基因敲除的生物体和细胞系，但是其也存在许多缺陷：

1. 为了删除目的基因，科学家必须在目的基因的编码区而不是非编码区寻找潜在的锌指蛋白核酸内切酶靶定位点，即：并非所有的潜在区域都是锌指蛋白核酸内切酶理想的靶定位点；
2. 获得识别特异性区域的“锌指蛋白装配模块”效率低于 6%，即：并非所有的锌指结构都可以有效的、可靠的组成锌指蛋白核酸内切酶，并且某些装配的锌指蛋白核酸内切酶具有细胞毒性，从而限制了对某些特定基因的敲除效果；
3. 脱靶效应。由于在细胞内表达的锌指蛋白核酸内切酶单体可能出现同源或者异源二聚化，因此很容易造成成对的锌指蛋白核酸内切酶并没有特异性靶定到目的基因序列；
4. 编码锌指蛋白核酸内切酶以

及其 DNA 分子导入细胞中的方式有限；

5. 锌指蛋白核酸内切酶自身的免疫原性使得其在临床治疗中需要考虑其会引起的机体免疫反应；
6. 目的基因删除的单克隆筛选繁重。

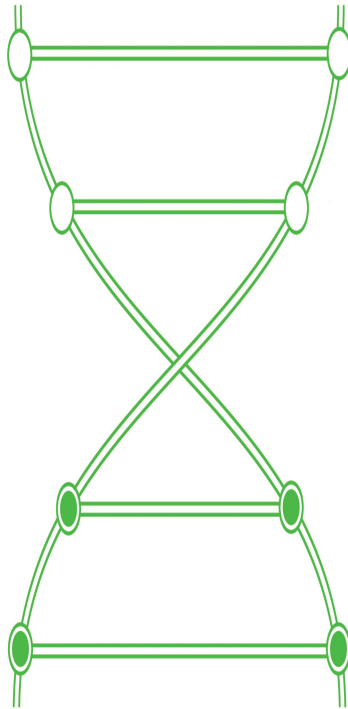
## 二、TALEN (transcription activator-like effector nucleases)

转录激活子样效应因子 (transcription activator-like effector, TALE) 是由植物病原体 *Xanthomonas spp* 衍生出的一类病毒因子家族，当这种病原体被注入到宿主植物时，其将结合至宿主启动子的特定区域从而调控植物疾病发生的进程。TALEs 由 17-18 个重复序列紧密排列组成，每个重复序列含有 34 个氨基酸，目前研究表明，每个重复序列中第 12,13 位氨基酸与其 DNA 结合特异性有关，与锌指蛋白单个模块识别 3-4 个碱基不同的是，TALEs 的每一个氨基酸只对应地识别单个碱基，这一方面意味着在识别相同基因序列的前提下，TALEs 需要 3 倍锌指蛋白模块数，但在另一方面也意味着，TALEs 在特异性识别目的基因序列的过程中更加灵活。因为每个锌指蛋白单个模块特异性识别 3-4 个碱基，也就意味着需要从 100 多种模块挑选组装特异性的锌指蛋白核酸内切酶，此外每个模块自身以及模块之间的连接顺序都会显著影响锌指蛋白核酸内切酶结

合及切割目的基因的效率。正如 Martin Luther University 的 Jens Boch 所说的那样：“TALEs 很神奇，当我们知道不同 TALEs 重复序列的氨基酸组成，就可以预测它会结合至某个特定的 DNA 序列，也就是说敲除的目的基因序列决定我们设计不同的 TALEs，而锌指蛋白在某种程度上决定我们对目的基因区域的选择。”

目前为止，酵母、植物、石斑鱼以及人类细胞系中 DNA 都能通过设计相应的 TALEs 来靶定，研究人员通过结合 TALEs 和限制性核酸酶来编辑基因。如前面我们提到的，与锌指核酸内切酶介导的基因敲除相比，TALEN 介导的基因敲除在针对不同目的基因设计转录激活子样效应因子模块时更加灵活，从而大大缩短了目的基因敲除的时间和花费。此外，我们提到锌指核酸内切酶存在脱靶效应以及细胞毒性，其脱靶效应主要与锌指蛋白 DNA 结合结构域非特异性结合以及 Fok I 核酸酶二聚化有关，而 TALEN 也使用 Fok I 核酸酶，从而也具有脱靶效应。但是，研究表明其脱靶效应以及细胞毒性明显要低于锌指核酸内切酶。虽然如此，TALEN 介导的基因敲除也存在一些明显缺陷：

1. 目前为止，虽然利用 TALEN 技术，很多公司和研究机构成功构建了多个目的基因敲除的小鼠等动物模型，但科学家对其背景缺乏基本了解，这就造成当构建的 TALEN 没有按研究者计划敲除目的基因时，研究者往往不知道



从哪处着手解决问题。

2. 与在自然界广泛存在的锌指蛋白相比，TALEs 仅在植物病原体中发现。因此，理论上，相对于前者来说，后者会是一种更强的免疫原。

### 三、CRISPR/Cas9 系统

虽然锌指蛋白核酸内切酶、TALEN 介导的基因编辑在生物学领域被深入研究和应用几十年，但是鉴于其存在的缺点，一种更易评估、负担及获得的新的基因编辑技术仍是科研以及临床应用的需求。基于 II 型原核 CRISPR 适应性免疫系统来源的 RNA 引导的 Cas9 核酸内切酶对基因组的精确编辑使得科学家在基因编辑领域取得了惊人的进步。

CRISPR-Cas 最初被发现是因为它作为细菌的适应性免疫

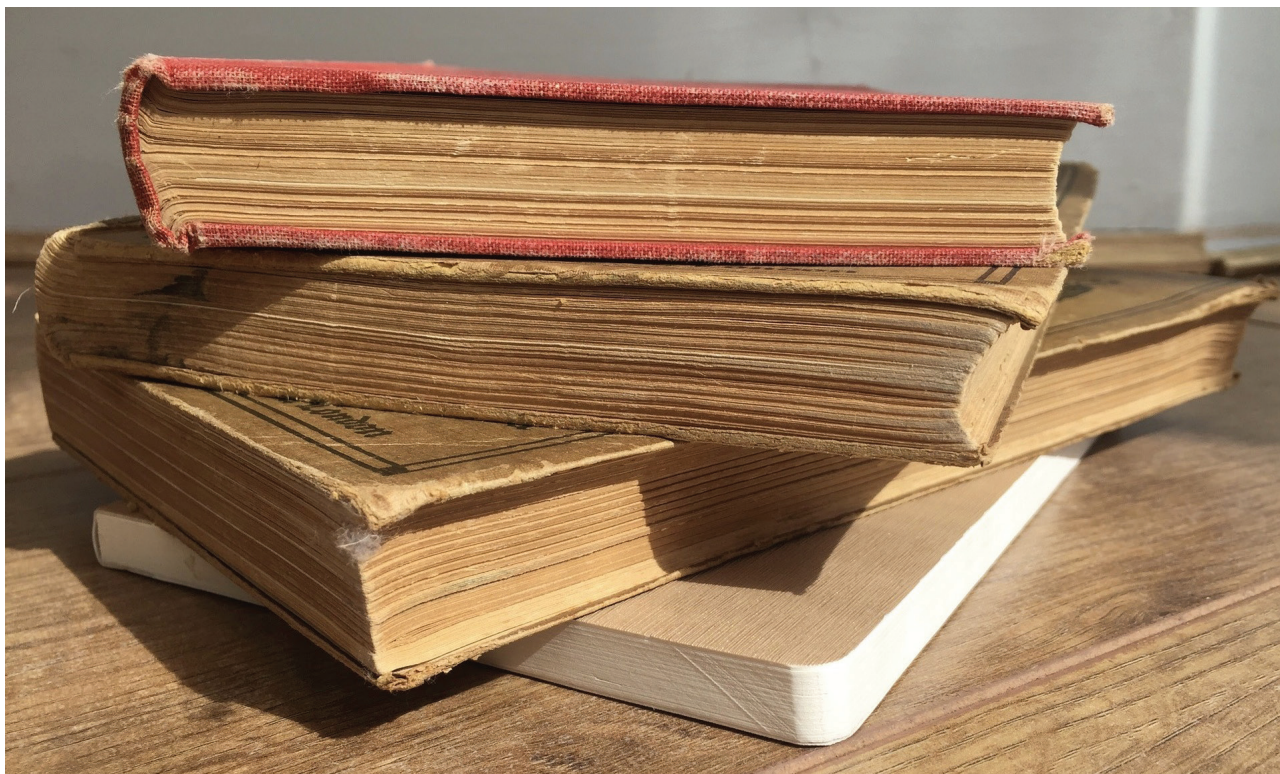
系统，利用 RNA 引导的核酸酶来剪切外来的基因元件。目前为止，在细菌和古生物中一共发现了三类 CRISPR 系统，它们都由 Cas 基因、非编码 RNA 和一个特殊的重复元件构成的 CRISPR RNA (crRNA) 阵列组成。例如，II 型 CRISPR 系统包含核酸酶 Cas9、编码引导 RNA 的 crRNA、以及激活 tracrRNA，后者可以将 crRNA 加工成独立的含有 20 个核苷酸的小单元，即引导 RNA (guided RNA)，引导 RNA 通过 Watson-Crick 碱基配对法则引导核酸酶 Cas9 结合至目的基因序列从而造成目的基因双链 DNA 断裂，被剪切的目的基因通过同源介导的修复或者非同源末端连接来修复损伤部位。

在利用 CRISPR/Cas9 系统进行基因编辑时，crRNA 和 tracrRNA 可以融合成单段引导 RNA (single-guide RNA, sgRNA)，通过改变 sgRNA 20 个核苷酸的序列，Cas9 可以被引导至不同目的基因的序列从而编辑不同的目的基因。此外，利用 CRISPR/Cas9 系统已经在不同的生物体和细胞系中成功进行了基因编辑，在这个过程中，为了保证基因编辑的最大成功率，通常要根据物种的不同，对 Cas9 要进行优化。

同锌指蛋白核酸内切酶以及 TALEN 相比，CRISPR/Cas9 系统因其更简单的操作，被科学家认为具有更加广阔的应用前景：

1. CRISPR/Cas9 系统通过特异性设计针对目的基因序列 20 个核苷酸大小的引物，可以将 Cas9 特异性地引导至目的





基因序列，在操作时间上大大缩短。设计针对靶基因的引物大约就花费 1-2 周时间，而目的基因敲除的细胞系在 2-3 周内即可获得；

2. 位点切割更具有特异性。Cas9 稳定地在靶序列第 17,18 个碱基处切割目的基因；
3. 编辑效率更高。Cas9 和 TALEN 都可以在不同的细胞系和生物体中有效编辑基因组，但是由于 Cas9 相比于 TALEN 更容易靶定至目的序列，这意味着利用 CRISPR/Cas9 系统可以通过设计针对多个目的基因的 sgRNA 在一个细胞系中同时进行多个目的基因的编辑。

但是，同锌指蛋白核酸内切酶以及 TALEN 的缺陷一样，CRISPR/Cas9 系统也存在脱靶效

应。目前为止，科研工作者对其主要缺点的解决办法是，通过生物信息学预测出多个容易脱靶的潜在位点，从而在 sgRNA 设计过程中避开这些潜在脱靶位点。

#### 四、NgAgo-gDNA 基因编辑技术

2016 年 5 月 2 日，河北科技大学韩春雨课题组在顶尖杂志《Nature biotechnology》发表了自己具有极大原始创新性的生物技术领域研究成果，称其发现了一种由 NgAgo (*Natronobacterium gregoryi* Argonaute) 介导的新的基因编辑技术，与时兴的 CRISPR/Cas9 系统利用 gRNA 引导 Cas9 至靶定区域不同的是，其利用引导 DNA (gDNA) 引导核酸内切酶 NgAgo 至目的序列。虽

然该技术后来遭到许多科学家对于其可重复性的质疑，韩春雨于 2017 年 1 月又向《自然·生物技术》杂志提交了与 NgAgo 系统可重复性相关的最新数据，目前《自然·生物技术》仍在对这些数据进行调查研究。这里我们不论事件的背景隐情及最终结论如何，仅针对这个技术进行描述和比较。

Argonaute 家族具有核酸内切酶活性，以前对此家族的研究发现其能有效利用 5' 磷酸化的 ssDNA 作为介质，精确切割靶基因，然而从 *Thermus thermophilus* 以及 *Pyrococcus furiosus* 中分离的 Argonaute 在 65°C 反应温度时才能发挥最大活性，因此限制了其作为工具在哺乳动物中进行基因编辑的应用。韩春雨课题组利用从 *Natronobacterium gregoryi* 分离出的 Argonaute 发

现, NgAgo 能够在 37°C 同 5' 磷酸化的 ssDNA 结合, 对目的基因进行剪切。

NgAgo-gDNA 基因编辑技术的应用方法同 CRISPR/Cas9 系统类似, 即: 设计合成一对针对目的基因的、互补的、长度为 24 个碱基的 5' 磷酸化的 ssDNA 作为引导 DNA, 引导核酸内切酶 NgAgo 靶定到目的基因部位并制造双链 DNA 缺口, 然后通过细胞内同源重组进行 DNA 修复。

虽然该文作者发现, NgAgo-gDNA 介导的基因敲除同 CRISPR/Cas9 系统的效率不相上下, 但是目前的研究结果表明, 前者在预防脱靶效应以及应用方面可能优于后者:

1. 碱基错配对 NgAgo 效率影响巨大。当 gDNA 同目的基因有一个碱基错配时, 内切酶 NgAgo 效率降低 73%-100%; 而当出现三个碱基错配时, NgAgo 效率几乎完全消失; 而在 CRISPR/Cas9 系统中, 其可以容忍 5 个碱基的错配;
2. 自然条件下, Argonaute 在

多个生物体中都存在, 其引导 DNA 即 5' 磷酸化的 ssDNA 在哺乳动物中含量少, 并且 NgAgo 同 5' 磷酸化的 ssDNA 结合是快速专一的;

3. 在 CRISPR/Cas9 系统中, 穿插在 crRNA 重复序列的前间区序列周围保守的前间区序列临近基序 (protospacer adjacent motif, PAM) 对 gRNA 在靶基因的定位有较大影响, 而在 NgAgo-gDNA 中并没有这个限制。

在本文, 我们对基因敲除技术的发展做了一个简单的梳理。总之, 相较于非定点突变, 锌指蛋白核酸内切酶因为首次能够特异性地敲除目的基因, 在发现的几十年的时间里, 科学家投入了大量的时间和金钱进行研究优化其条件; 而 TALEN 技术的出现, 极大改善了系统中 DNA 结合模块的灵活性, 因此逐渐取代了前者, 而在 TALEN 技术还没有被完全了解其背景时, CRISPR/Cas9 系统介导的基因敲除技术几乎成为了基因编辑的主流。因为其更加高效、耗时短、易操作, 目前为止,

已经有科研团队利用其在人类胚胎中进行了基因改造, 故其在科研、临床疾病治疗的方面一度被看好; 而 NgAgo-gDNA 系统, 目前为止, 其呈现出同 CRISPR/Cas9 系统相似的高效率, 并且在应用方面更加简单, 脱靶效率更低, 但是其后续是否能够取代 CRISPR/Cas9 成为一种能应用于科研及临床上的最成熟的基因编辑工具还有待科学家们进一步研究和证实。

#### 参考文献

- [1] Ran FA, Hsu PD, Wright J, et al. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* 2013;8(11):2281-2308.
- [2] Feng Z, Le C, Simona L et al. Programmable Sequence-Specific Transcriptional Regulation of Mammalian Genome Using Designer TAL Effectors. *Nature Biotechnology* 2011;29(2):149.
- [3] Listed N. Move over ZFNs. *Nature Biotechnology* 2011;29(8):681.
- [4] Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 2005;435(435):646-651.

